The Latest Analytical Electron Microscope and its Application to Ceramics

Yoshiyasu Harada* and Yuichi Ikuhara†
*JEOL Ltd., 3-1-2 Musashino, Akishima, Tokyo 196-8558, Japan, †Institute of Engineering Innovation, School of Engineering, University of Tokyo, 2-11-16 Yayoi, Bunkyo, Tokyo 113-8656, Japan

Chapter Outline
1. Introduction 3
2. General Overview of Analytical Electron Microscope 3
3. Transmission Electron Microscopy 7
4. Scanning Transmission Electron Microscopy 8
5. Analysis Method 10
6. Application of Analytical Electron Microscopy to Ceramics 12
7. Conclusion 20
References 20

1. INTRODUCTION

The transmission electron microscope (TEM) has been utilized for observation of fine structures and analysis of a crystalline structure of material, since its invention in 1932 [1]. After that, in the 1970s, the functions of scanning transmission electron microscope (STEM), energy dispersive spectroscopy (EDS), electron energy loss spectroscopy (EELS), etc. were built in TEM, and its application fields have been drastically expanded to include compositional analysis and analysis of an electronic structure as well as image observation by TEM/STEM up to the present date. As such, the TEM/STEM, which can make both the image observation and the analysis, is called an analytical electron microscope (AEM) [2,3].

Then, the electron gun and objective lens, which are important elemental technologies of the electron microscope, have been greatly improved, and the performance of AEM has improved drastically. That is, as for the electron gun, in the 1990s, instead of the conventional thermionic emission electron gun (TEG), field emission electron gun (FEG) of high brightness has been put into practical use, and the current of the nano-probe focused on a specimen has been greatly increased. In addition, as for the objective lens, by commercialization of TEM equipped with a condenser objective lens (C/O lens) in the 1980s, observation of TEM image and STEM image became possible by using the identical objective lens excitation, and then research of material science field has been developed. Moreover, in the 2000s, the corrector of spherical aberration, which used to restrict the resolution of TEM/STEM, has been put into practical use; therefore, the image observation and the analysis of atomic levels have become possible, and AEM has been utilized as a useful evaluation instrument in the most advanced area of material development in the material science fields such as metal, semiconductor, and ceramics, etc.

In this article, we will describe the general overview of the most advanced AEM first, and will especially explain the electron gun and objective lens as well as the spherical aberration corrector, which relates to the AEM performance. Next, we will describe the image formation in TEM and STEM and analysis method (EDS and EELS), which are indispensable for applying AEM to material science field. Finally, we will introduce how AEM is utilized in research of ceramics material concretely based on several application examples.

2. GENERAL OVERVIEW OF ANALYTICAL ELECTRON MICROSCOPE

2.1. Various Information Obtained by Interaction between Electron Beam and Specimen

When electron beams are irradiated to a specimen, as shown in Figure 1, as a result of an interaction between the
electron beam and specimen, information such as the secondary electron, Auger electron, backscattered electron, absorbed electron, characteristic X-ray, cathode luminescence, etc. can be obtained. In addition, in case the specimen is a thin film, by interaction between the electron beam and the specimen, information of the transmitted electron, which suffered elastic scattering and inelastic scattering, can be obtained. By utilizing this information, AEM can obtain a great deal of knowledge about a specimen.

In this chapter, we cover major image-forming methods by TEM and STEM which obtain magnified images of a specimen by using transmitted electrons; in addition, we describe the compositional analysis and chemical state analysis of a specimen using EDS and EELS.

2.2. Composition of Analytical Electron Microscope

Figure 2 shows the appearance of the most advanced AEM (JEM-ARM200F) with spherical aberration corrector (Cs-corrector), which enables the observation of both TEM image and STEM image [4]. Electron beams are emitted from the electron gun at the top, and the emitted electron beams are accelerated and magnified (in the case of TEM) or demagnified (in the case of STEM) by the illumination lens system, and then irradiated to the specimen. In the case of TEM, electrons passing through a specimen are formed into an image by image-forming lens system, and the image are observed on a fluorescent screen in an image observation chamber. In addition, they can be observed through a CCD camera in an image observation chamber or, at the lower part of it, on a liquid crystal monitor.

In the case of STEM, by scanning a demagnified electron probe, a synchronized signal is detected by various detectors attached at the lower part of the camera chamber; thus, compositional analysis can be possible.

Among many elements composing of AEM, the factors to determine the performance are the electron gun and objective lens (OL) including aberration correction technology. With the instrument shown in Figure 2, FEG is provided as an electron gun; however, there are instruments which have TEG where tungsten (W) hairpin filament and lanthanum hexaboride (LaB₆), etc. are used as an electron source. We describe the characteristics comparison of these electron guns in 2.3. In addition, as for the OL and aberration correction technology, we describe in 2.4 and 2.5 respectively.

2.3. Performance Comparison of Various Electron Guns

Table 1 shows the features of various electron guns. Since the invention of the electron microscope, TEG with W hairpin filament has been utilized as a standard electron gun until the 1970s, but in the 1980s, the use of TEG with LaB₆ [5] of high brightness became popular. Later on, in the 1990s, FEG of higher brightness was put into practical use, and began being used as the electron gun of AEM. There are principally two types of FEGs. One is the cold-type FEG(C-FEG) which uses W(310) as electron source which is kept at room temperature [6]. The other is the thermal-type FEG (T-FEG) with heated ZrO/W(100) [7] as an electron source. The feature of C-FEG is that the energy spread of an emission current is small, while the features of T-FEG are that large emission currents can be taken, the current stability is high, and that the ease of operation is also good. With the most advanced AEM, though it has both advantages and disadvantages, either type of FEG has been used as a standard. Recently, Schottky-type electron gun whose feature is about same as the T-FEG is commonly used in place of T-FEG.

2.4. Objective Lens

The resolution (d) of the TEM is expressed as 
\[ d = 0.65C_s^{1/4}λ^{3/4} \]
Here, Cs is the spherical aberration coefficient and λ is the wavelength of electron. Since the invention of the electron microscope, in order to improve the resolution, efforts have been made whether making spherical aberration coefficient of OL smaller or raising the accelerating voltage, that is, making λ shorter. At present, the standard accelerating voltage of AEM is 200—300 kV, that is, λ = 0.00251—0.00197 nm.

On the other hand, the efforts to reduce Cs of the OL have been done by many researchers since the invention of TEM. There are principally two trends. In other words, the shape of the OL pole pieces differs greatly, whether inserting a specimen from the top through the upper bore of the pole piece (top entry method) or inserting a specimen from the side through the gap (side entry method) to enter a specimen in the
gap between pole pieces. To correspond to the top entry method, it is necessary to select asymmetric pole pieces where the bore diameter of the upper pole ($b_1$) is larger than that of the lower pole ($b_2$). Also, to correspond to the side entry method, it is necessary to select symmetric pole pieces of $b_1 = b_2$, and to make the gap length(s) of the pole pieces relatively greater. As for the asymmetric pole pieces, since the top entry specimen stage is more stable in terms of antivibration properties, it has been improved along the trend that pursues the resolution of the TEM image. However, in order to use it in the case of AEM, it is necessary to change the excitation of the OL in order to switch from TEM image to STEM image, which is difficult in terms of ease of operation, so that its use in standard AEM was stopped in the mid-1980s. On the other hand, as for the symmetric pole pieces, in consideration that the improvement of antivibration properties of the side entry specimen stage has advanced, as the TEM image and STEM image can be observed by using an identical excitation of the OL, it is excellent in terms of ease of operation; therefore, it developed rapidly in the 1980s, and it has been used as a standard OL of AEM up to the present date. Table 2 shows the shape of pole pieces in these two types of
OL to obtain 0.5 mm Cs at an accelerating voltage of 100 kV. Here, Cc is the chromatic aberration coefficient, f0 is the focal length, and NI/√Vr is the excitation parameter (NI: ampere turn, Vr: relativistic accelerating voltage). In the case of an asymmetric lens [8], it is necessary to make b1 great, therefore, b₁ = 5 mm. In addition, in the case of a symmetric lens which is used as C/O lens [9], it is necessary to make s relatively large, therefore, s = 2.5 mm. If we assume that Cs of approximately 0.5 mm can be realized at the accelerating voltage in the range of 100–300 kV with either OL, the resolution at that time will be 0.26 nm, 0.19 nm, and 0.16 nm, respectively, at 100 kV, 200 kV, and 300 kV. When the Cs-corrector is not used, these resolutions are the best at each accelerating voltage.

2.5. Spherical Aberration Corrector

In the mid-1990s, Cs-corrector was put into practical use, and, in the 2000s, AEM with Cs-corrector of TEM/STEM has rapidly become widely used, until today. Right after the electron microscope was invented, a study on the aberration of the electron lens was executed actively. Scherzer demonstrated that Cs and chromatic aberration coefficient (Cc) of rotational symmetric electron lens cannot take a negative value [10], and proposed a method to correct these aberrations by using a rotational asymmetric electron lens which can take a negative value of these aberration coefficients [11].

Later on, many researchers have done theoretical and experimental studies of aberration correction by using rotational asymmetric electron lens such as the quadrupole lens, octupole lens, hexapole lens, and dodecapole lens; however, these were hardly put into practical use. The reasons are that (1) the simulation of rotational asymmetric electron lens has never been accurately done, (2) the technology to process the rotational asymmetric electron lens with accuracy was not available, (3) the electric stability of rotational asymmetric electron lens has never been improved, and (4) the electric stability of the electron microscope basic unit has not been sufficient. However, in 1995, about 50 years after the proposal by Scherzer, the above-described problems were resolved by Rose and Haider who are the successors of Scherzer’s research laboratory, and Cs-corrector using a 2-stage hexapole lens was put into practical use [12,13]. Lens structure of the corrector, which can correct spherical aberration of the third order, is shown in Figure 3.

Figure 3(a) is the case of TEM which forms a magnified image of the specimen, and Cs-corrector is placed at the lower side of the OL. Figure 3(b) shows the case of STEM which forms the nano-probe on the specimen, and the Cs-corrector is placed at the upper side of OL. With this Cs-corrector, along with the large three-fold astigmatism, negative spherical aberration appearing in terms of higher order from the three-fold astigmatism is utilized for the spherical aberration correction of the OL with the 1st-stage and 2nd-stage hexapole lens. The three-fold astigmatism arising from the hexapole lens of the 1st stage and the other of 2nd stage is set in reverse directions; therefore, it became

<table>
<thead>
<tr>
<th>TABLE 1 Comparison of Characteristics of Various Electron Guns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermionic emission</strong></td>
</tr>
<tr>
<td>W-hairpin</td>
</tr>
<tr>
<td>Brightness (A/cm² · sr) at 200 kV</td>
</tr>
<tr>
<td>-5 × 10⁵</td>
</tr>
<tr>
<td>Optical source size (μm)</td>
</tr>
<tr>
<td>Energy spread (eV)</td>
</tr>
<tr>
<td>Operating temperature (K)</td>
</tr>
<tr>
<td>Typical vacuum (Pa)</td>
</tr>
<tr>
<td>Emission current (μA)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2 Comparison of Optical Properties of Symmetrical and Asymmetrical Objective Lenses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symmetrical b₁ = b₂</strong></td>
</tr>
<tr>
<td>b₁ (mm)</td>
</tr>
<tr>
<td>s (mm)</td>
</tr>
<tr>
<td>b₂ (mm)</td>
</tr>
<tr>
<td>Cs (nm)</td>
</tr>
<tr>
<td>Cc (nm)</td>
</tr>
<tr>
<td>f₀ (mm)</td>
</tr>
<tr>
<td>NI/Vr</td>
</tr>
</tbody>
</table>
the optical system that makes three-fold astigmatism zero as a whole.

With the electron microscope which has the Cs-corrector, if the accelerating voltage is 300 kV, the resolution of 0.05 nm is obtained in TEM/STEM, and the minimum probe diameter on the specimen is approximately 0.1 nm[14]. In addition, in case the accelerating voltage is 200 kV, the resolution is 0.1 nm in TEM/STEM, and approximately 0.1 nm as the minimum probe diameter on the specimen is obtained[4].

3. TRANSMISSION ELECTRON MICROSCOPY


In the case of an observation of a magnified image and observation of an electron diffraction pattern, which are the basic observation methods of TEM, the ray diagram of the image-forming lens system is shown in Figure 4. Observation of a wide range of magnified images is possible from low magnification of 50 times to high magnification of 1,000,000 times.

Figure 4(a) shows a ray diagram to obtain low magnification of 50–1000 times, and in order to observe an image of a wide field of view with small aberration, excitation of OL is cut, objective mini (OM) lens is used instead of the OL, and the image is magnified by the intermediate lens (IL) and projector lens (PL), and can be observed on the fluorescent screen. Figure 4(b) is the ray diagram to obtain a magnified image at high magnification of 1000–1000,000 times, and the image magnified by the OL is magnified by several intermediate lens (IL1–IL3) and PL further, and observed on the fluorescent screen. The insertion of OL aperture on the back focal plane of the OL is possible, and by selecting the size of the OL aperture, observation of magnified image, which has various contrasts, is possible. Figure 4(c) is the ray diagram to obtain an electron diffraction pattern. The electron diffraction pattern formed at the back focal plane of the OL, is magnified by IL and PL, and observed on the fluorescent screen. By selecting the size of the selected-area (SA) aperture placed on the object plane of the IL1, selection of an area on the specimen, which corresponds to the electron diffraction pattern, is possible.

Electron microscope image and electron diffraction pattern can be switched on at the flip of a switch and observed. In case the specimen is crystal, a spot-shaped electron diffraction pattern is obtained by Bragg reflection on the crystal plane which is parallel to the incident electron beam, and by selecting only the transmitted spot without being diffracted by the OL aperture, a bright field image can be observed, and by selecting the diffraction spot only, a dark field image can be observed. With a bright field image and dark field image, the bright and dark contrasts are observed, and these contrasts differ depending on the specimen thickness and the local distortion accompanied by the curvature of the crystal plane, dislocation, stacking fault, etc. As for the reason why these contrasts are formed, explanation by the electron diffraction theory[15], which consists of the kinematical theory and dynamical theory, is possible, but we do not cover it here.

3.2. High-Resolution Electron Microscopy

There are principally two methods of high-resolution electron microscopy (HREM) to study crystal material[16].
One is the observation of crystal lattice images. When the electron beam is illuminated in parallel to crystal zone axis, a two-dimensional electron diffraction pattern is obtained. If only the transmitted wave (spot) and the diffracted wave, which is the closest to the transmitted wave, are taken in the OL aperture for image formation, a two-dimensional crystal lattice image, which has information of unit cell of crystal by interference of transmitted wave and diffracted wave, can be observed. With this method, information on defects such as twins, stacking fault, dislocation, grain boundary, etc. can be obtained. The other is the observation of crystal structure images. If the transmitted wave and many diffracted waves are taken in the OL aperture, and an image is formed under the condition to be explained later, crystal structure image corresponding to its projected potential can be observed. With the most advanced TEM with Cs-corrector, a crystal lattice image and crystal structure image can be easily observed at a resolution of about 0.1 nm. In the crystal structure image, the atomic column in thin crystal (several atomic rows lined up in the direction of specimen thickness) can be observed directly.

As for the thin specimen like this, amplitude of the electron wave passing through specimen hardly changes, and only phase changes; therefore, such a thin specimen is called a phase object. The image formation theory of phase object was reported by Scherzer in 1949 [17]. We do not cover the details here, but when the contrast transfer function of the OL, which is determined by the Cs of OL and defocus values (df), fell under a certain value without vibration in a wide spatial frequency area, the projected potential of the crystal, that is, the crystal structure, will be well reflected to an electron microscopic image. In other words, the resolution is the highest, and the optimal df is given as $df = 1.2(Cs\lambda)^{1/2}$, which is called Scherzer focus. The resolution ($d$) of TEM is, as explained in Section 2.4, expressed as $d = 0.65Cs^{1/4}\lambda^{3/4}$.

In addition, in case the actual HREM image is observed, it is necessary to consider the deterioration of the resolution, caused by chromatic aberration attributed to energy spread ($dE$) of the incident electron beam and $Cc$ of OL, and illumination angle ($\alpha$) of the electron beam on the specimen. These influences are described as an envelope function showing the attenuation of high-frequency component of a scattering wave [18,19]. Figure 5 shows the calculation example of contrast transfer function in consideration of the envelope function when an electron microscope of 300 kV accelerating voltage with Cs-corrector is provided with T-FEG ($dE = 0.8$ eV), or C-FEG ($dE = 0.3$ eV) In the case of T-FEG, as energy spread is large, observation is only possible at a resolution of about 0.07 nm, but, in the case of C-FEG, it is estimated that there is a possibility that a resolution of 0.05 nm can be obtained.

![FIGURE 5 Contrast transfer function in the case of C-FEG ($dE = 0.3$ eV) and T-FEG($dE = 0.8$ eV).](image)

What to pay attention most in interpreting the image observed by this image-forming method is that the contrast of the electron microscopic image changes greatly by df, that is, by vibration of contrast transfer function. If one wrong step is taken, there is a possibility that an incorrect image interpretation will be made. In this sense, it is inevitable to conduct image observation under the Scherzer focus condition. Also, it is essential to interpret the observed image carefully, by comparing it with the computer simulation image by using the multi-slice method which was offered by Cowley & Moodie [20].

4. SCANNING TRANSMISSION ELECTRON MICROSCOPY

In the 1980s, as the C/O lens was adopted as a standard OL for TEM, switching between TEM and STEM images became easy, and thus the observation method by STEM rapidly became common. With the C/O lens, a specimen is placed almost in the center of the gap of OL pole pieces, and the prefield above the specimen works as a CL and the postfield below the specimen works as an OL.

Figure 6 shows a typical ray diagram of the illumination lens system in TEM mode and in STEM mode. Figure 6(a) shows the ray diagram of the illumination lens system in the TEM mode. By focusing electron beams at the front focal plane of the prefield of OL by using a CM lens, the parallel beam is irradiated to the specimen and an electron microscopic image of high coherency can be observed. Figure 6(b) and (c) show the ray diagram of the illumination lens system in STEM mode. In STEM mode, electron beams are demagnified to form a nano-probe on the specimen and the electron probe diameter is a great factor in determining the resolution of the STEM image. Figure 6(b) shows that, if excitation of the CM lens is disconnected and if the size of the CL aperture is constant, the specimen illumination angle becomes the maximum angle $\alpha_1$, which is determined by the size of the CL aperture. Also, by changing the excitation of the CM lens which is exited, the...
illumination angle \( (\alpha_2) \) can be controlled. As such, to change the illumination angle, it is not only important in changing the electron probe current on the specimen, but also important in determining what kind of transmitted electrons to be taken in STEM detectors, which will be mentioned later.

In addition, under the condition of image-forming lens system to obtain an electron diffraction pattern as described in Section 3.1, if a large illumination angle (convergence angle) is set in a similar way as the STEM mode in Figure 6(b) and (c), a convergent beam electron diffraction (CBED) pattern can be obtained from the nanometer area. If CBED is used, determination of the point group and space group of the crystal is possible [21]; moreover, measurement of the lattice distortion around the grain boundary in the crystal is possible.

4.1. Various Detection Methods of Transmitted Electron

Electrons passing through a thin film specimen contain transmitted electrons without being suffered by scattering, electrons that suffered elastic scattering, and electrons that suffered inelastic scattering. In general, it is known that electrons that suffered elastic scattering will distribute in a relatively wide scattering angle from 50 to 100 mrad or so, and electrons that suffered inelastic scattering which excites an electron in an atom will distribute in a relatively small scattering angle of about 1 mrad. In the case of TEM, as the parallel electron beam illuminates the specimen, the electron diffraction pattern of the crystalline specimen becomes spot shaped, but in the case of STEM, as mentioned earlier, as incident electron beams are focused on the specimen, in case the Cs-corrector is not installed, the convergent angle (full angle) is about 10–20 mrad or so, and in case Cs-corrector is installed, the convergent angle (full angle) becomes larger to be 50–100 mrad or so; then, in the case of the crystalline specimen, the electron diffraction pattern made at the back focal plane of the OL will be disk shaped.

The image formation of STEM is determined by taking which part of the disk-shaped electron diffraction pattern in the detector [22]. If the detector is set at the transmission disk corresponding to the convergent angle of the incident electron beam, a bright field image can be obtained. If the detector is set at the diffraction disk, a dark field image is obtained. In addition, if the detector is set where the transmission disk and the diffraction disk overlap, as in TEM’s case, a crystal lattice image which is the typical example of phase contrast can be observed. In other words, it has been reported that the reciprocity theorem can apply to the image formation by the TEM and STEM [23,24].

Here, we describe two detection methods which are extremely useful in the research of material science field among several STEM image detection methods, namely, HAADF (high-angle annular dark-field) method and ABF (annular bright field) method. The schematic diagram of those methods is shown in Figure 7. Figure 7(a) shows the schematic diagram of the HAADF method and Figure 7(b) shows the schematic diagram of the ABF method.

4.2. HAADF-STEM

HAADF-STEM is a method to form an STEM image by detecting only high-angle scattered electrons by using an annular detector. The annular detector was first used by Crewe [25], in order to observe the image of a single atom which depends on the atomic number, by forming STEM image in proportion with the elastically scattered electron of 10 mrad or more and inelastically scattered electron of 1 mrad or less. Later, since the result was presented in the field of crystal material by Pennycook et al. by using HAADF-STEM with an annular detector of 50–70 mrad or so [26], many researchers have made use of HAADF-STEM. The elastically scattered electrons which were detected by the annular detector of 50–70 mrad or so are the Rutherford scattering, and as its intensity \( (I) \) is in proportion to the square of the atomic number \( (Z) \) of the atom in the specimen, the image contrast observed by using HAADF-STEM is called a \( Z \) square contrast image. As the scattering angle becomes large, the atomic scattering factor of elastic scattering reduces rapidly and the intensity of thermal diffuse scattering (TDS) by phonon excitation is dominant. As a result, TDS forms the major contrast of HAADF-STEM image [27]. As this image is basically an incoherent image, the image interpretation is easy which is
a great feature and different from the HREM image in the TEM and the STEM phase contrast images that were mentioned earlier. That is, if the specimen thickness is constant, the atomic column of the large $Z$ shows a brighter contrast than that of a small $Z$; therefore, it has a feature that the difference between the atomic columns can be clearly distinguished.

4.3. ABF-STEM

ABF-STEM [28,29], which was recently proposed by JEOL, University of Tokyo, and Japan Fine Ceramics Center, is a method to form an STEM image by using an annular detector set at the transmission disk area corresponding to the convergent angle of the incident electron beam. It was verified that contrast of the atomic column of light element in ABF-STEM is higher than in HAADF-STEM by systematic experiments and simulation of STEM image [30].

They used several kinds of oxide as a specimen and compared the images observed by using ABF-STEM and HAADF-STEM, and showed that the atomic column of oxygen was not observed by HAADF-STEM but by ABF-STEM, in which the atomic column of oxygen can be clearly observed as a dark contrast [29]. Also, they made a simulation of an ABF-STEM image by systematically changing the specimen thickness and defocus values, made a defocus-thickness map, and made it clear that the STEM images of both the light atom column and heavy atom column can be observed with a wide range of defocus values and specimen thickness [30]. ABF-STEM is a novel observation method of the STEM image, which has been proposed recently, and it is a useful method which enables a simultaneous observation of both the column of light atom and heavy atom. It is expected that it will be utilized by many researchers in the field of ceramics.

5. ANALYSIS METHOD

Here, we explain the EDS and EELS analyses which have been widely used since the 1970s as the most basic compositional analysis of AEM. EDS is a method to analyze the composition by a pulse-height analysis of the energy of the characteristic X-ray coming from the specimen by using an energy dispersive X-ray detector. EELS is a method of compositional analysis by measuring energy loss of transmitted electrons which suffered inelastic scattering in the specimen.

5.1. EDS Analysis

Figure 8 shows the principle of the EDS analysis. By using a semiconductor detector (SSD: solid state detector) of a high-purity silicon single crystal doped with a trace of lithium, it can make an elemental analysis by immediately converting all of the energy of a characteristic X-ray generated from the specimen into a pulse voltage according to the amplitude of the field-effect transistor (FET), and counting the pulse number by a multichannel pulse-height analyzer. The spectrum is indicated where X-ray energy is in horizontal axis and the number of photons is in the
vertical axis. In the EDS system, it is always necessary to cool SSD and FET to liquid nitrogen temperature in order to suppress their thermal noise.

There are two kinds of EDS detectors which are used with AEM. One is a normal detector (Be-window type) which detects an X-ray through Be-window of about 7-μm thickness and the other is UTW (ultrathin window)-type detector which detects an X-ray through the window where a thin plastic film is coated with about 0.1-μm-thick aluminum. The elements that are heavier than Na (Z = 11) are detectable in case of Be-window type, because Be film absorbs X-ray, while with the UTW type, the detectable elements are the ones that are heavier than C (Z = 6). In other words, with the UTW type, the analysis capability of light elements such as C, N, O, etc. is the biggest feature. The feature of Be-window type is the high detection efficiency because it can make an effective area of the detector large.

EDS can provide not only a qualitative analysis but also a quantitative analysis and it has the feature that elemental analysis can be done easily. However, the utmost disadvantage is that the energy resolution is as low as about 150 eV. Therefore, spectrum overlapping occurs and attention will be needed in interpretation.

There are principally two methods in EDS analysis. One is the point analysis method in which the electron probe is stopped at one point on the specimen by using TEM/STEM, and an X-ray spectrum is acquired. The other is the elemental mapping method, which scans two-dimensionally the electron probe on the specimen by using STEM, modulates the brightness corresponding to the intensity of a certain characteristic X-ray, synchronizes with a scanning signal, and displays the two-dimensional image of the characteristic X-ray intensity on the liquid crystal monitor. The most advanced AEM with FEG along with a Cs-corrector can focus the minimum electron probe diameter on the specimen to be 0.1 nm or so. Therefore, the atomic level point analysis and two-dimension elemental mapping are possible and is widely used for research in the material science field.

5.2. EELS Analysis

Among the electrons passing through the thin film specimen, transmitted electrons, elastically scattered electrons, and inelastically scattered electrons which lose energy as a result of interaction with an atom in the specimen exist. Inelastically scattered electrons arise from lattice vibration (phonon excitation), collective excitation of valence electron (plasmon excitation), interband transition, inner-shell electron excitation (core electron excitation), etc., and by measuring the energy of the inelastically scattered electrons, information such as the electronic structure and composition of the specimen can be obtained. Focusing on the energy loss peak accompanied by the inner-shell electron excitation, identification of the element in the specimen, composition analysis, and chemical state analysis is possible. In addition, focusing on the fine structure of the inner-shell electron excitation spectrum, that is, the energy loss near-edge structure (ELNES), information on the electronic structure can be obtained.

EELS spectrometer which measures the energy of the transmitted electron has several types. Here the most commonly used parallel detection system [31] is introduced as shown in Figure 9. The parallel detection system consists of a parallel detector of semiconductor which is composed of a fluorescent material (YAG), a fiberoptic plate, and photo diode array. As the parallel detector can read signals of each channel at the same time, it has the characteristics of high detection efficiency. Table 3 shows a comparison of the UTW/EDS and the EELS analyses. In the case of the EELS analysis, as transmitted electron energy loss is to be measured, the specimen thickness needs to be thin enough. The thickness of the specimen is said to be appropriate to be about the same degree of thickness as the mean free path(λp) of plasmon loss. If the specimen becomes thick, the signal of plasmon loss increases and it will become the background in detecting a signal of the inner-shell electron excitation. The detection element is possible from Li(3) to U(92) and its feature is the high-energy resolution compared with the EDS analysis.

| TABLE 3 Comparison of Characteristics of UTW/EDS and EELS |
|--------------------------------|----------------|---------------|
| Conditions of specimen         | optional       | thin film (−λp) |
| Detectable elements            | C(6)           | Li(3)–U(92)   |
| −U(92)                         |                |
| Energy resolution (eV)         | −150           | −1            |
| S/B ratio                      | −80            | −0.3          |
| Element mapping                | easy           | necessary data processing |
Figure 10 shows the spectra of EELS and UTW/EDS by using barium titanate (BaTiO$_3$) as a specimen with AEM of 200 kV accelerating voltage. In the UTW/EDS analysis, peaks of Ti–Kz (4.51 keV) and Ba–Lz (4.47 keV), Ti–Kβ (4.93 keV) and Ba–Lβ (4.83 keV), and O–Kz (0.53 keV) and Ti–Lα (0.45 keV) overlapped, and identification of Ti and Ba was not possible. On the other hand, with the EELS analysis, L-absorption edge of Ti (0.46 keV) and M-absorption edge (0.80 keV) of Ba have been clearly separated; thus, it can be understood that the resolution of EELS is superior to the one in EDS. In addition, though oxygen is detected in either spectrum, it is demonstrated that while in the EDS analysis, the ratio of signal ($S$) and background ($B$), $S/B$ ratio is high, while in the EELS analysis, the $S/B$ ratio is low.

6. APPLICATION OF ANALYTICAL ELECTRON MICROSCOPY TO CERAMICS

The mechanical and electromagnetic properties of a ceramic material have a close relation to microstructures. The
microstructure contains grain boundary, hetero-interface, dislocation, planar defect, precipitate, composition distribution, etc., and each factor determines various properties of material. Recently, as the material synthetic process becomes more accurate, the microstructure of material began being controlled in atomic levels. From such a viewpoint, characterization of the microstructure, which has a higher resolution and accuracy, is needed. As a method to observe the inner microstructures of a material, the transmission electron microscopy (TEM) is the most useful method. Moreover, it can also handle a structure analysis of the atomic level, local area compositional and chemical bonding state analysis in nanometer order, and the information obtained by using this method is quite a lot. In this chapter, in relation with the above-mentioned clause, we will explain the most advanced application examples by using high-resolution electron microscopy (HREM), CBED, scanning transmission electron microscopy (STEM), EDS, and EELS.

6.1. Application of High-Resolution Electron Microscopy

When grain boundary or interface of a material is observed at a high resolution, the interface has to be edge on (meaning that the incident direction of the electron beam has to be parallel to the interface), and the Bragg condition of two crystal grains which sandwiches the interface has to be satisfied. Figure 11 shows the observation procedure for the grain boundary. This example shows the case where a thin amorphous phase exists in the ceramic grain boundary [32]. When the electron beam is irradiated at the grain boundary in the edge-on state, the electron diffraction pattern from two crystal grains and the diffuse scattering electrons from the amorphous phase are formed on the back focal plane of the OL. By taking in these waves in the OL aperture and form image, an HREM image of the grain boundary can be obtained. The size of the OL aperture is determined by considering the resolution of the electron microscope and image contrast. In the case of grain-boundary observation of a polycrystal, as the grain boundary which satisfies the above-mentioned observation conditions are limited, it is necessary to search for the grain boundary which is observable through a trial-and-error process. In cases like when the grain boundary is curved to the thickness direction, the observation will be difficult as lattice images from two crystals that overlap in the grain boundary. In addition, if only the diffuse scattering electrons are taken in the objective aperture and observed by dark field imaging, only the amorphous phase can be observed as a white contrast, and it is convenient when the distribution is desired to be known. Figure 12 shows an HREM image of grain boundary in silicon nitride and it is recognized that the amorphous phase with a thickness of about 1 nm, which is not observable at low magnification, has been clearly observed. This photograph is observed under almost the same conditions as in Figure 11. The high-temperature strength of silicon nitride is considered to be dependent on the characteristics of the amorphous phase which exists at the grain boundary. Therefore, the information on the thickness and distribution of the phase becomes important; thus, HREM can be a useful approach. In addition, as for the evaluation of the composition and chemical bonding state in the grain boundary of amorphous phase, it can be possible to use EDS together with EELS by using a nano-probe as explained later.

In order for further quantitative analysis, such as the identification of the location of an atom in the grain boundary and interface, comparison with computer simulation will be needed. The structure model of grain boundary (super cell) needs to be considered for the simulation. In constructing a grain-boundary model, various methods can be considered and the simplest way is the rigid model. This is a model to simply connect two
crystals without consideration of the atomic structure relaxation in the grain boundary. This is a simple and easy method that enables the construction of a structure which can be assumed from the high-resolution image and compares it with a simulation image. However, in general, with the rigid model, the structure cannot be uniquely determined and several assumed structure models need to be constructed. In addition, as it does not take in the atomic structure relaxation of the grain boundary, it cannot provide a quantitative analysis of grain-boundary atomic structure. For more quantitative analyses, the general method is to construct an optimal structure of the grain boundary or interface including structure relaxation by using the first principles calculation, static lattice calculation, and molecular dynamics calculation, and to determine the structure by comparing the simulation image with the observed image [33–35].

Figure 13 shows the analysis example of Σ9 grain boundary of YSZ (yttrium-stabilized zirconia) [36]. This specimen is a twin crystal made by adjusting two single crystals in a certain orientation, and by using thermal diffusion bonding at a high temperature. This twin crystal is made so that the common rotation axis is [110] and the grain-boundary plane is {221} with Σ9 symmetry. Figure 13(a) shows the HREM image of the grain boundary observed by an ultra-high voltage electron microscope (accelerating voltage 1250 kV), and the black contrast corresponds to the atomic position. As shown in the figure, at the grain boundary, two kinds of asymmetric structure units are formed periodically. HREM image simulation for the most stable structure of the same grain boundary is shown in Figure 13(b), and the most stable structure model obtained by theoretical calculation is shown in Figure 13(c), respectively. It can be seen that, also in the theoretical model, an asymmetrical unit periodically appears and this result is consistent with the experimental image. At the grain boundary, cation sites of a different coordination number are formed as shown with an arrow in Figure 13(c), and these sites would be an important influence in determining the characteristic of this grain boundary.

6.2. Application of Convergent Beam Electron Diffraction

Electron diffraction has been also used to identify various crystal phases. The diffraction method from a narrow area using the nano-probe is used for various materials in general. As a method of the electron diffraction, CBED has been well known, and it is effectively used in determining the space group of the crystal phase, but it is also used in measuring lattice distortion of a narrow area. In other words, as CBED can measure the lattice constant in high accuracy from the diffraction line in higher-order Laue zone (HOLZ), it is useful in measuring small distortion of a local area such as the grain boundary and interface, etc. As the functional property of material is greatly influenced by not only the atomic structure of grain boundary, but also the lattice distortion due to the atomic structure of the grain boundary, CBED with a high spatial resolution attracts attention. Figure 14 shows the analysis results of Σ21 grain boundary of alumina by using CBED. In the figure, (a) shows an HREM image of the grain boundary, (b) shows a HOLZ line which is formed in the CBED pattern, and (c) shows the strain distribution in the vicinity of the grain boundary which was measured by this technique [37]. (b) shows the relation of the HOLZ line which is used in measuring the lattice constant of a-plane(2110) plane, and the incident direction of the electron beam is [0001] direction. As shown in the figure, by measuring the distance $d$ and $e$, which connects the intersecting point of the HOLZ line, an estimation of the lattice constant of a-plane is possible. In this case, a probe diameter of about 0.7 nm is used, but calibration of the accelerating voltage of the electron microscope is performed by using a bulk sapphire single crystal, and its lattice constant is used for calibration in advance as the standard value. Also, the measurement of
the distance between intersecting points is performed on
the image which was magnified by 20 times of the negative
photographed with a 1.2-m camera length. The relation
between the lattice constant of a-plane and the distance
from the grain boundary, which were measured as such, is
shown in (c). It is revealed that the lattice constant greatly
changes in the vicinity of the grain boundary and that it
reaches about 0.07% lattice distortion at the maximum.

6.3. Application of HAADF-STEM

6.3.1. The Grain Boundary of Rare Earth
Element-Doped Alumina Ceramics

It is known that the high-temperature strength of alumina
ceramics improves drastically by doping a trace of rare
earth element such as Y and Lu, etc. [38]. These doped trace
elements are assumed to segregate in the grain boundary of
alumina polycrystal, and prevent the grain boundary sliding
at high temperatures, but there are many unknown issues
regarding the intrinsic mechanism. Here, the grain-
boundary structure of the Y-doped alumina was observed
using the STEM and the result of clarification of the grain-
boundary strengthening mechanism by Y is shown.

Figure 15 shows the HAADF-STEM image of the Y-doped
alumina grain boundary [39]. The grain boundary shown
here is so called Σ31 which can be classified in the general
grain boundary [40]. As shown from this figure, in the
Y-doped grain boundary, it is recognized that a very strong
contrast has been observed at the particular atomic site
along the grain boundary. This corresponds to the Y atom
position segregated in the grain-boundary cores, and it
shows that the Y atoms segregate selectively and regularly
only at the center of the seven-membered ring structure in
the grain-boundary cores. Figure 16 shows an electron
density map of the Y-doped grain boundary which was
obtained by the first principles calculation based on these
observation results. The bonds without segregation have

FIGURE 14 Analysis result of Σ21 grain boundary of alumina by
CBED. (a): HREM image of the grain boundary, (b): HOLZ line which is
formed in the CBED pattern, and (c): Distortion distribution in the vicinity
of the grain boundary. For color version of this figure, the reader is referred
to the online version of this book [37].

FIGURE 15 HAADF-STEM image of Y-doped alumina grain boundary.
For color version of this figure, the reader is referred to the online version
of this book [39].

FIGURE 16 Charge density map for Y-doped alumina grain boundary
obtained by the first principles calculation. For color version of this figure,
the reader is referred to the online version of this book.
mainly ionic bonding. However, it is recognized that the electron density distribution around Y atoms which segregated in the center of the seven-membered ring structure caused an orientation and covalent-like bonds have been formed [39]. It has been clear that rare earth elements have the effect to greatly change the surrounding chemical bonds and improve the grain-boundary strength by segregating in the grain boundary. That is, we can bring to a close that the intrinsic effect of rare earth elements doped in alumina ceramics is attributed to the local change of a chemical bonding state in the grain boundary.

6.3.2. STEM-EELS Observation of Superlattice Thin Film

It is known that superlattice thin film provides a new function which is not available with bulk material. For example, the thermoelectric characteristic (performance index ZT) of strontium titanate (SrTiO₃) is 0.08 or so at room temperature and it is far from ZT > 1, which is supposed to be the index for practical application. However, with the superlattice thin film where 1-unit cell of Nb-doped SrTiO₃ layer is inserted (made) in the SrTiO₃ without dopant, ZT goes up to 2.4, the thermoelectric characteristic which is equivalent to the ones of Bi–Te system and Pb–Te system which are already in practical use, can be obtained [41].

In order to efficiently develop such a superlattice thin film, it is necessary to clarify the mechanism where this function appears, by combining the analysis method of a high spatial resolution with the theoretical calculation. Here, we will introduce the atomic and electronic state analysis of SrTiO₃/Nb-doped SrTiO₃/SrTiO₃ superlattice, by combining STEM with EELS and first principles calculations.

Figure 17(a) shows the HAADF-STEM image of SrTiO₃/Nb-doped SrTiO₃/SrTiO₃ superlattice thin film which was fabricated by the pulsed laser deposition (PLD) method [41]. As the intensity of the HAADF-STEM image is proportional to about the square of Z, compared with Ti (Z = 22) column, Sr (Z = 38) column can be observed more brightly. It is known that Nb-doped SrTiO₃ layer inserted by every 24-unit cell appears as a stripe-shaped contrast.

Figure 17(b) and (c) show the high-resolution image around the Nb-doped SrTiO₃ layer and the line profile of image intensity of the Sr atomic row and Ti atomic row in the same area. It is recognized that the image intensity does not change in Sr atomic row: however, the image intensity in Ti atomic row became stronger at the Nb-doped SrTiO₃ layer. Taking into consideration that the atomic number of Nb is 41, it is considered that Nb exists in the Ti site. On the other hand, since the atomic numbers of Nb and Sr are close, whether Nb exists in the Sr site or not cannot be judged only from the HAADF-STEM image. In order to clarify that, the solution energy of Nb was estimated by using the first principles PAW (projector augmented-wave) method. It has been proven that the solution energy of Nb to Sr sites is higher by 7.6 eV compared with the one to Ti sites. This shows that Nb does not dissolve in the Sr sites and it exists in the Ti sites.

Figure 18 shows the spectra of Ti-L₂,₃ ELNES obtained from the SrTiO₃ area and Nb-doped SrTiO₃ layer in SrTiO₃/Nb-doped SrTiO₃/SrTiO₃ superlattice. It is recognized that in the SrTiO₃ layer, four peaks clearly appear; on the other hand, in the Nb-doped SrTiO₃ layer, the peaks become broad. In order to interpret this change, the first principles relativistic multi-electron calculation was performed. As a result, it became clear that the change of the experimental spectrum is due to the transition from Ti⁴⁺ to Ti³⁺, which accompanies Nb doping. This shows that in the Nb-doped SrTiO₃ layer, the surplus electrons were introduced and the two-dimensional electron gas is being formed [41]. So far, it has been clear that the high thermoelectric characteristics of SrTiO₃/Nb-doped SrTiO₃/SrTiO₃ superlattice are attributed to this two-dimensional electron gas.

6.4. Application of ABF-STEM

The clarification of a lithium-ion battery mechanism and improvement of the reliability are related to the behavior of the lithium ion, and it is considered that if direct visualization of a lithium atom is possible, it will be a great breakthrough in the research and development. In the past,
an attempt to observe the lithium atom by using a high-resolution electron microscopy has been made. However, only indirect observations such as the method to reconstruct HREM image [42], and the other method to modify the lithium site with other element [43], etc., have been reported, and it has been considered that the direct observation of a lithium atom which is a light element is very difficult in terms of technique.

Here, it is demonstrated that the observation of a lithium atom with the atomic number 3 became possible by using ABF-STEM [29,30]. A series of observations of lithium ion have been carried out for materials such as LiMn₂O₄, LiCoO₂, and LiFePO₄, etc. so far, and direct observations of the lithium atomic row have been done successfully [44,45].

Here, the observation example of the lithium-ion row in LiCoO₂ crystal with a layer structure is introduced. Figure 19(a) shows the crystal structure model of LiCoO₂. It is understood from the figure that this crystal has a layer structure, and that along \( \frac{1}{2}11 \frac{2}{20} / C138 \) direction, each atom of Li, O, Co is allocated side by side. Figure 19(b) is the ABF-STEM image of LiCoO₂ crystal observed from \( \frac{1}{2}11 \frac{2}{20} / C138 \) direction, and it is recognized that in addition to the Co atomic row, which has a strong black contrast, the Li atomic row can be clearly observed between oxygen atomic layers. This is the image obtained at the 8–25-mrad detection angle, and the observation of atomic row has been made for a light element and a heavy element at the same time. This observation is only possible by using STEM with a resolution of 0.1 nm or lower by the spherical aberration correction and by setting the detection angle where the lithium atom can be observed by using theoretical calculation. By further applying this method to practical materials, the diffusion mechanism of the lithium ion and the behavior of the lithium ion during charging and discharging have become clear and it is expected that future research and development of the lithium-ion battery will be achieved. It has been recently reported that hydrogen atoms can be observed by ABF-STEM imaging technique [46].

6.5. Application of EDS

In a nano-probe analysis of the local area such as the grain boundary and interface, etc., it is important to fix the probe at the interface position for sure. In that case, as measurement

![Figure 18](image1.png)  
**Figure 18** (a): Ti-L₂,₃ ELNES spectrum obtained from SrTiO₃ layer and Nb-doped SrTiO₃ layer in SrTiO₃/Nb-doped SrTiO₃ superlattice and (b): The spectrum obtained by the first principles relativistic multi-electron calculation.

![Figure 19](image2.png)  
**Figure 19** (a): Crystal structure model of LiCoO₂ and (b): ABF-STEM image of LiCoO₂ observed from \( [1120] \) direction. For color version of this figure, the reader is referred to the online version of this book.
by the probe of sub-nanometer diameter or lower is required, FEG with high brightness is normally used. Figure 20(a)–(c) show schematically the position relation of the nano-probe and the grain boundary or interface during analysis. The nano-probe is illuminated to the grain boundary or interface as the conical beam which has a certain illumination angle. Therefore, unless a probe is precisely focused within the area to be analyzed (position of disk of least confusion), it is not a correct analysis including the quantitative analysis. In case the focus position is correct, the focusing will be almost in the center of the specimen thickness (Figure 20(a)), but if the focus is misaligned (Figure 20(b)), electron beams are transmitting the specimen while spreading depends on the illumination angle; thus, the actual spatial resolution turned out to be deteriorated. On the one hand, as shown in Figure 20(c), when the interface is tilted against the probe, the interface area that occupies in the probe reduces, and then the detection sensitivity of the segregation element, etc. declines as well as its quantification. Figure 20(d) shows the position relation of the interface and nano-probe during an EDS analysis.

In the case of an EDS analysis, as that the EDS detector is positioned on the extension of the interface will be an advantage in terms of detection efficiency (plane hkl in Figure 20(d)), it is necessary to set the tilt direction of the specimen to be an EDS detector side at least. This not only improves the quantification of the EDS analysis, but also increases the detection efficiency of X-ray; therefore, it also leads to the reduction of the specimen drift and specimen damage, etc. In addition, the position accuracy of the probe and the influence of the stray X-ray need to be investigated in advance. For example, the influence of the spurious X-ray and actual probe diameter can be confirmed from the EDS spectrum when the probe is fixed in a space of about 1 nm away from the specimen edge. Moreover, with that state, when it is changed to the diffraction mode, if we check whether a diffraction pattern is observed or not, then, we will be able to know the probe position accuracy due to the stray magnetic field of the electromagnetic lens system. Also, an attention is needed, that in order to perform a precise quantitative analysis, the analysis area also needs to be a thin area where a thin film approximation can be applied.

Figure 21 is an analysis example of the WC/Co interface of 0.5 wt% VC-doped WC–Co super hard alloy. WC–Co super hard alloy is widely used in cutting tool materials, etc., and it has a characteristic composition where the hard WC particle is bonded by Co phase, which is a metal phase. Figure 21(a) is the HREM image of WC/Co interface. Figure 21(b) is the EDS spectra obtained at the position of A to E in the HREM image, and these are the results of an analysis by using a nano-probe of about

---

**FIGURE 20** (a) to (c): Schematic diagram showing position relation between nano-probe and the grain boundary or interface during analysis and (d): Position relation between nano-probe and the grain boundary or interface during an EDS analysis.

**FIGURE 21** Analysis example of the WC/Co interface of 0.5 wt% VC-doped WC–Co super hard alloy. (a): HREM image of WC/Co interface and (b): EDS spectra obtained at the position of A to E in (a) by using a nano-probe of about 0.5 nm in diameter.
0.5 nm in diameter. Judging from a comparison of B to D, clear V-K peak is detected at the WC/Co interface (C). However, in the WC and Co grains (B and D), which are away by only 2 nm from the interface, V is not detected; therefore, it is shown that the doped V segregates only near WC/Co interface. This segregation volume differs by a facet of the WC grain, that is, more segregation of V on the (0001) facet (C) is recognized compared with the other facet (E). On the other hand, at the intersection points of these two facets, micro-precipitates as shown by arrows in the HREM image are formed, and from the EDS analysis as well as the HREM observation, it is clear that these precipitates are (W,V)\(_2\)C. From a series of these analysis results, the grain growth inhibitor mechanism of the WC/Co super hard alloy by VC has been clarified and it can be understood that EDS with a nano-probe is useful for practical materials.

On the other hand, by scanning the nano-probe in STEM, elemental mapping by EDS is possible. Figure 22 shows (a) bright field-STEM image of a trace amount of Lu\(_2\)O\(_3\)-doped Al\(_2\)O\(_3\) polycrystal, and (b) Lu-K\(_\alpha\)-STEM mapping image (probe diameter of about 0.5 nm). It is clearly known that the doped Lu ions segregate along the grain boundary. It has been reported that the creep properties of the Al\(_2\)O\(_3\) polycrystal greatly improve by doping the trace amount of Lu\(_2\)O\(_3\) [38]. However, it has been clarified by such a series of analysis that the segregations of Lu in the grain boundary is the reason for that. The STEM-EDS analysis has the feature that elemental distribution can be obtained with a wide field of view and at a high resolution, different from the above-mentioned point analysis, and its application is expected a lot in the future.

6.6. Application of EELS

EELS is useful for the chemical bonding state analysis of a local area. In other words, with the spectrum in EELS, ELNES contains information such as a local atomic arrangement and chemical bonding. So far, for interpretation of ELNES, comparison with a spectrum from reference substance, the so-called fingerprint matching has been done, but along with the complication of the substance, a more quantitative method is needed. In the past, calculation of ELNES has been done by various methods such as the cluster method, etc. At present, calculation by using OLCAO (orthogonalized linear combinations of atomic orbitals) method, a kind of the first principles band calculation methods, has been attracting attention as it can reproduce an experimental spectrum of ELNES well [48]. ELNES is equivalent of transition energy of an electron from the inner-shell orbit to the unoccupied orbit, and during that transition process an inner-shell hole is formed. The calculation is made by taking in the effect of inner-shell hole and, by using the OLCAO method, an effective calculation is possible.

Figure 23 shows the experimentally obtained O-K ELNES in the vicinity of the ZnO grain boundary, and the calculated spectrum by the OLCAO method [49]. It is widely known that the grain boundary of ZnO shows a nonlinear current—voltage characteristic and it is assumed that the origin of this characteristic is due to the point defects segregated in the grain boundary. Figure 23 (upper) is the O-K ELNES measured in the vicinity of the high angle grain boundary of ZnO (area about 20 nm away from the grain boundary), and this spectrum is characterized by three main peaks as shown by A, B, and C. On the other hand, the lower spectrum is the theoretical O-K
ELNES from the O atom in a perfect crystal of ZnO. It is recognized that this theoretical O-K ELNES is very similar to the experimental spectrum in terms of both shape and position of peaks. This result suggests that remarkable segregation of point defects, which has an influence on the current—voltage characteristic in the vicinity of ZnO grain boundary, does not occur. As such, for the quantitative evaluation of ELNES, interpretation by a theoretical calculation is useful.

7. CONCLUSION

As FEG and Cs-corrector have been installed with the most advanced AEM, not only image observation at a resolution of about 0.1 nm but also formation of the probe diameter less than 0.1 nm have been possible. As a result, not only direct observation of the atomic column from a light element to a heavy element in the thin crystal by TEM and direct observation of the atomic column from a light element to a heavy element in the thin crystal by TEM and STEM, but also the atomic-level elemental mapping by EDS and EELS, and ELNES, have become possible.

In this paper, we have explained the general overview of the most advanced AEM, as well as shown its application research of ceramic materials concretely. It would be wonderful if this paper is able to serve as some kind of help in conducting research and development of ceramic materials.

REFERENCES


1. CARBON MATERIALS

Carbon materials have played important roles for human beings: charcoals as a heat source and as an adsorbent since prehistorical age, flaky natural graphite powder as pencil lead, soot in black ink for the development of communication techniques, graphite electrodes for steel production, carbon blacks for reinforcing tires in order to develop the motorization, electric conductive carbon rods and carbon blacks for supporting the development of primary batteries, a compound of graphite with fluorine (graphite fluoride) for improving the performance of primary batteries, thin graphite flakes in membrane switches for making computers and control panels thinner and lighter, etc. In relation to global-warming problems, nuclear reactors are recognized to be important, which are constructed from high-density isotropic graphite blocks, having different roles as reflectors, moderators, etc. For the storage and efficient usage of sustainable energies, which are usually unstable, lithium-ion rechargeable batteries and electric double-layer capacitors are essential devices, in which carbon materials are used as electrodes and govern their performance: in the former, lithium intercalation/deintercalation into the galleries of the anode graphite is a fundamental electrochemical reaction and, in the latter, physical adsorption of electrolyte ions onto the surfaces of the electrodes of porous carbon is a fundamental electrochemical reaction. Nanocarbons developed recently, which are represented by carbon nanotubes, fullerenes, and graphene, are promoting the development of nanotechnology in various fields of science and engineering.

Carbon materials are predominantly composed of carbon atoms, only one kind of element, but they have largely diverse structures and properties. Diamond has a three-dimensional structure and graphite has a two-dimensional nature, whereas carbon nanotubes are one-dimensional and buckminsterfullerene C\(_{60}\) is quasi-zero-dimensional. Fullerenes behave as molecules, although other carbon materials do not. Graphite is an electric conductor and its conductivity is strongly enhanced by AsF\(_5\) intercalation, higher than metallic copper, whereas diamond is completely insulating. Diamond which is the hardest material is used for cutting tools, and graphite is so soft that it can be used as a lubricant.

This chapter provides fundamental information on carbon materials, before going into the details on various advanced carbon materials, nanocarbons including carbon nanotubes, high-density isotropic carbons, carbon fibers and their composites, and carbon materials for energy devices.

2. CHEMICAL BONDING AND CARBON FAMILIES

The carbon atom has four electrons in the outermost 2\(s\) and 2\(p\) orbitals, namely 2\(s^2\), 2\(p^2\). When it forms chemical bonds with other atoms, it can have three different hybridizations, 2\(sp\), 2\(sp^2\), and 2\(sp^3\). The hybridized 2\(sp\) orbital is composed of two \(\sigma\) electrons, associated with two \(\pi\) electrons, the 2\(sp^2\) of three \(\sigma\) electrons with one \(\pi\) electron, and the 2\(sp^3\) of four \(\sigma\) electrons. It is well known that various combinations of these hybrid orbitals produce a variety of organic hydrocarbons, from aliphatic to aromatic ones. This situation is illustrated in Figure 1, by showing some hydrocarbons [54,59]. The carbon–carbon bonds (C–C bonds) using 2\(sp^3\) hybrid orbitals give a series of aromatic hydrocarbons, such as benzene, anthracene, phenanthrene, etc. The C–C bonds using 2\(sp^2\) hybrid orbitals give various aliphatic hydrocarbons, such as methane, ethane, propane, adamantane, etc. Mixing the
different types of hybridizations, $sp$, $sp^2$, and $sp^3$, in a molecule expands the variety of hydrocarbons.

Inorganic carbon materials, such as diamond, graphite, fullerenes and carbynes, are considered as an extension of these organic molecules to giant ones, as shown in Figure 1. A simple repetition of $C\equiv C$ bonds with $sp^2$ hybrid orbital gives planar hexagons of carbon atoms, as benzene, anthracene, ovalene, etc. The extension of this series of aromatic hydrocarbons reaches graphite, in which the $\sigma$ bonds using the $sp^2$ hybrid orbitals are reinforced by the delocalized $\pi$-electrons. The giant planes tend to stack on each other due to the interactions between the $\pi$-electron clouds of the stacked layers, giving rise to weak van der Waals-like interaction, i.e., graphite. The
somewhat curved corannulene molecule is formed by associating a pentagon of $sp^2$-hybridized carbon atoms with 5 hexagons. By the polymerization of these coronulene molecules to form a closed shell, various sizes of carbon clusters are formed and the resultant carbon materials are called fullerenes, where carbon hexagons have to be located between pentagons. The smallest closed shell is the buckminsterfullerene C$_{60}$, consisting of 12 pentagons and 20 hexagons of carbon atoms. As shown schematically in Figure 1, the infinite repetition of C–C bonds using $sp^3$ hybrid orbitals results in a three-dimensional network of carbon atoms, which is known to be diamond. The carbon atoms in the organic compound named as adamantane are exactly in the same atomic positions as those in diamond crystal. The infinite repetition of C–C bonds with $sp$ hybrid orbital gives the carbon materials called carbynes, in which the carbon atoms form linear chains either with double bonds (one $\sigma$-bond with one $\pi$-bond) or with the alternative repetition of single (one $\sigma$-bond with two $\pi$-bonds), the former being called cumulene type and the latter polyyne type, respectively.

Based on the extension by the repetition of three kinds of C–C bonds to infinite molecules, we may define carbon families, represented by the perfect structure in each family, that is, diamond, graphite, fullerene, and carbyne [54]. However, each family is known to have characteristic diversity in structure, as will be explained briefly in the following sections.

3. STRUCTURE
3.1. Graphite Family
The carbon family having a planar $sp^2$ bonding is represented by graphite, where the layers of carbon hexagons are stacked in parallel; the stacking regularity of ABAB⋯ sequence belongs to the hexagonal crystallographic system (hexagonal graphite) and that of ABCABC⋯ to the rhombohedral system (rhombohedral graphite). The unit cells of these two allotropes are shown in Figure 2, together with their crystallographic data: space group, equivalent points, and lattice parameters. For the rhombohedral graphite, both rhombohedral and hexagonal cells are shown together in the figure.

The hexagonal graphite is a thermodynamically stable phase. The rhombohedral structure of graphite was shown to form by applying shearing forces, for example, by grinding, which seems to be formed mainly due to the introduction of stacking faults in the crystallites having the hexagonal stacking regularity.

In addition to these regular stacking modes, the parallel stacking of the layers without any regularity occurs mostly in the carbon materials prepared at low temperatures, below 1300 °C. In this case, the layers of hexagons are usually small

![Hexagonal graphite](image1)

![Rhombohedral graphite](image2)

**FIGURE 2** Crystal structures and lattice parameters of two graphite allotropes.
in size and only few layers are stacked in parallel. This random stacking of layers is called turbostratic structure.

Figure 3a shows the X-ray powder diffraction pattern (powder pattern) of natural graphite with a high crystallinity. The diffraction lines of graphite have to be classified into three groups, the lines with 00l, hk0, and hkl indices, mainly because of the strong anisotropy in structure. The lines with 00l indices are due to the reflection from basal planes (hexagonal carbon layers, graphite basal planes), where only the even l indices are allowed because of the extinction rule due to the parallel stacking of the layers. The lines with hk0 indices are due to the reflection from the crystallographic planes perpendicular to the basal planes and the lines with hkl indices come from the planes declining to the basal planes, where the three-dimensional structure with regular stacking of layers has to be established. Therefore, the hkl lines are called three-dimensional lines.

On the other hand, the powder pattern which is given by low-temperature-treated carbons is quite different from well-crystallized graphite, because it consists of random stacking of small layers, as shown on a petroleum coke heat-treated up to 1000 °C in Figure 3b. The diffraction pattern is characterized by very broad 00l lines due to the small number of stacked layers, and by unsymmetrical hk lines and no hkl lines because of the random turbostratic stacking of layers. The diffraction lines due to the planes perpendicular to the hexagonal carbon layers are missing l index because of the absence of regularity in the direction along the normal to the layers (along the c-axis); therefore, they are expressed as 10 and 11 in Figure 3b.

When graphite is observed by scanning tunneling microscopy (STM), the hexagonal carbon layer looks like a triangular arrangement of spots (Figure 4a). Because of the interaction with the lower-lying atoms in the B layer, only a half of the carbon atoms are detected under STM. In the turbostratic structure, however, the interactions with the lower-lying atoms is much weaker, because of the
irregular stacking, and consequently the spots form a hexagonal arrangement (Figure 4b) [21].

Since the turbostratic structure can be partly transformed into a regular stacking of layers with a heat treatment at high temperatures, a wide range of structural diversity may result in the graphite family. The graphitization degree is defined as the probability of occurrence of the regular graphitic AB stacking in turbostratic stacking.

Carbon materials that belong to the graphite family can have various textures mainly because the basic structural units (BSUs), i.e., the stacked hexagonal carbon layers, are highly anisotropic. The texture at the nanometric scale (nanotexture) will be separately discussed in the Section 4.

In 2010, Nobel Prize in physics was given to Profs A. K. Geim and K. S. Novoselov for their work on graphene. Graphene is defined by an isolated single layer of carbon hexagon, two-dimensional crystal with just one atom thickness, and is expected to have high mechanical strength, high electrical and thermal conductivities, etc. On graphene, many reviews have been published [29,62,129,136].

### 3.2. Fullerene Family

Although based on $sp^2$-type hybridization of carbon atoms, the bonding nature in fullerenes is slightly different from graphite in the fact that the layer is curved by the introduction of pentagons, as seen in corannulene (Figure 1). The repetition of these curved layers can result in various closed shells with various sizes, from the smallest closed shell of $C_{60}$ to giant fullerenes. The introduction of additional hexagons among pentagons in $C_{60}$ leads to giant fullerenes by keeping closed shell morphology, as shown in the left-hand side of Figure 5. In this fullerene family, the variety of structure is mainly due to the number of carbon atoms building the closed shell.

Most of fullerenes behave as molecules [18], i.e., they can be dissolved into an organic solvent, giving a characteristic color, which is exceptional in the carbon materials because all other carbon materials are not dissolved into any organic solvent.

As shown in the center of Figure 5, there is another way to increase the number of hexagons between two groups of 6 pentagons, which results in a single-wall carbon nanotube. The carbon nanotube seems to locate in between the graphite and fullerene families, because it has the characteristics of graphite together with those of fullerenes. The closed ends of a nanotube consist of carbon pentagons with hexagons, which is the characteristic for the fullerene family, and the tube wall consists of carbon hexagons and can be considered to be rolled from an hexagonal carbon layer, which is the fundamental structural unit composed of the graphite family. A single-wall carbon nanotube is characterized by its diameter and chiral vector [18,132]. The illustration on the right-hand side of Figure 5 presents an example to roll up to form a single-wall carbon nanotube with a chirality of (5, 2). Fullerenes and nanotubes consisting of multi-layers are also formed.
3.3. Diamond Family

Diamond consists of $sp^3$-hybridized carbons, where the purely covalent chemical bonds extend three-dimensionally. Diamond is very hard and electrically insulating owing to no $\pi$ electrons. In order to construct a diamond crystal, a periodical and regular repetition of the C–C bond is required in a long range. For a couple of carbon atoms indicated as A and B in Figure 6a, the carbon atom A has to be connected with four carbon atoms, including B, to make a tetrahedron because of directional $sp^3$ bonds, and the atom B also has to be surrounded by four carbon atoms, including A. Looking down these two tetrahedra centered at the A and B atoms along their connecting line, there are two possibilities in mutual relation between two basement triangles consisting of three carbon atoms. If these two basement triangles are rotated with respect to each other by $60^\circ$, as shown in Figure 6b, the resultant diamond crystal belongs to the cubic system (cubic diamond, Figure 6b’). If there is no rotation between these two basement triangles (Figure 6c), the structure belongs to the hexagonal system (hexagonal diamond, Figure 6c’). Most of the diamond crystals, either occurring in nature or being synthesized, are cubic.

In the case where a long-range periodicity is not attained between interconnected tetrahedra (Figure 6a), in other words, in the case of a random rotation between the two basement triangles of tetrahedra, an amorphous structure is formed, which is called diamond-like carbon (DLC). Because of a random rotation between tetrahedra, some carbon atoms cannot form chemical bonds with neighboring carbon atoms, most of which are supposed to be connected with hydrogen atoms for stabilization. DLC is usually obtained as a thin film, mainly because a random repetition of tetrahedra is difficult to keep on a long distance. DLC is as hard as the diamond crystal, because it is also based on $sp^3$-hybridized atoms, and it contains a relatively large amount of hydrogen.

A metastable crystalline phase of carbon atoms, consisting of a face-centered cubic unit cell with a parameter of 0.3563 nm, was found in either thin film or nanoparticles [68].

3.4. Carbyne Family

Carbynes have been supposed to be formed by $sp$-hybridized carbon atoms bound linearly, where two $\pi$-electrons have to be involved, giving two possibilities, i.e., an alternative repetition of single and triple bonds (poly-yne) and a simple repetition of double bonds (cumulene) (Figure 1) [87]. The detailed structure of carbynes is not yet clarified, but some structural models have been proposed [76]. A structural model is illustrated in Figure 7, where some numbers of $sp$-hybridized carbon atoms form chains, which are associated together by van der Waals interaction between $\pi$-electron clouds to make a layer, and then the layers are stacked. In the carbyne family, the variety of structures seems to be mainly due to the number of carbon atoms forming a linear chain, in other words, due to the layer thickness and the density of chains in a layer.
3.5. Acceptance of Foreign Atoms in the Structure

Each carbon family also shows various characteristic possibilities for accepting foreign species. The substitution of either boron or nitrogen atom for carbon atom is possible for all carbon families, even though the amount substituted is very limited. Since this substitution is expected to modify the properties of carbon materials without noticeable structural changes, various modifications have been investigated on different types of carbon materials for various applications in order to reinforce their functions. Effects of boron doping on accelerating the development of graphitic stacking and on improving the electrode performance in lithium ion rechargeable batteries have been reported. Nitrogen-substituted diamond crystals occur rarely in nature. Recently, nitrogen-containing carbon materials were reported to have a high capacitance in electrochemical capacitors, though the state of the nitrogen atoms in the carbon matrix was not yet well understood. Nitrogen substitution in buckminsterfullerene, C_{60}N (heterofullerene), was also reported.

Intercalation between the layers of graphite and carbynes is effective and useful for the modification of their properties. The intercalation of either iron or potassium is reported to stabilize the structure of carbynes. Into the carbon materials belonging to the graphite family, a wide variety of foreign species (atoms, ions, and even molecules) have been intercalated to give interesting functionalities to the materials. Either donors (such as lithium and potassium) or acceptors (such as sulfuric acid, bromine, and ferric chloride) can be intercalated in the galleries between neighboring hexagonal carbon layers. The intercalation of sulfuric acid is widely used for the production of exfoliated graphite, which is converted to flexible graphite sheets for shielding and packing. Intercalation/deintercalation of lithium ions is the fundamental electrochemical reaction for the discharge/charge of lithium-ion rechargeable batteries. Graphite intercalation compounds with AsF_5 and SbF_5 were reported to have high electrical conductivity, higher than metallic copper, but they were never used in practice mainly because of instability of the compounds in air. The buckminsterfullerene C_{60} crystal can be doped by alkali metals (e.g., potassium), which occupy all the tetrahedral and octahedral interstices of the closest packing of fullerene clusters, giving rise to superconductive materials.

In the fullerene family, besides the substitution of carbon atoms in the shell and doping in the interstices between shells described above, there are two additional possibilities to accept foreign species: the insertion into the inner space of the closed shells and the covalent bonding of organic radicals to the surface of the shell. Rare earth elements have been inserted into fullerene cages (endohedral metallofullerenes).

In the carbyne family, doping by intercalation into the space between the carbon chains in a layer is theoretically possible, but to date experimental results have not been reported.

4. CARBONIZATION AND GRAPHITIZATION

4.1. Pyrolysis and Carbonization

Carbon materials are produced by the heat treatment of some organic polymers, i.e., carbon precursors, up to a temperature of 800–1500 °C in an inert atmosphere (carbonization). The general scheme of carbonization process is shown in Figure 8 for a carbon precursor, such as a pitch, by indicating the main out-gassed components and changes in residues, although its details depend strongly on the starting precursors.

At the beginning of precursor pyrolysis, aliphatic and then aromatic hydrocarbons with low molecular weights are released as gases, mainly because some of the C–C bonds are weaker than C–H bonds. Associated with this hydrocarbon release, cyclization and aromatization proceed in the residues and then polycondensation of aromatic molecules occurs. From around 600 °C, mainly foreign atoms such as oxygen and hydrogen are released as CO_2, CO, and CH_4. In this stage, the residues are either liquid or solid phases depending on the starting precursors. Above 800 °C, the main evolved gas is H_2 because of polycondensation of aromatics, and the residues are...
carbonaceous materials which still contain hydrogen and a small amount of foreign atoms, such as O, N, and/or S, even though they are often called carbon materials. In order to eliminate these foreign atoms, a heat treatment up to 2000 °C is required, strongly depending on the chemical state of these foreign atoms in the carbonaceous solid.

The principal reactions in the carbon precursor up to about 800 °C are pyrolysis and those above 800 °C are carbonization. However, these two processes occur continuously and often overlap with each other. Usually, the whole process from the precursor to the carbon material is called “carbonization”. This process depends strongly on the carbon precursors and also on the heat-treatment conditions, mainly temperature, heating rate, and pressure.

A simple carbonization process is exemplified on a polyimide film (commercially available Kapton film with 25 μm thickness); the structure of the repeating unit in the polyimide, the evolution of gases, and the changes in weight and size of the film up to 1000 °C are shown in Figure 9a–c, respectively [50]. The first step of carbonization occurs in rather narrow temperature range of 500–600 °C, showing an abrupt weight loss associated

---

**FIGURE 8** Scheme of out-gases, reactions, and state of residues during carbonization.

**FIGURE 9** Carbonization of a polyimide film.
with the evolution of a large amount of carbon monoxide and a pronounced shrinkage along the film (pyrolysis). The second step occurs in the temperature range from 600 °C to 1000 °C with a small weight loss, the evolution of small amounts of methane and hydrogen, and a little shrinkage. The third step above 1000 °C, which is not shown in Figure 9, is accompanied by the evolution of nitrogen, a negligibly small shrinkage, and a small weight loss. The release of nitrogen in the third step was found to continue up to temperatures higher than 2000 °C, leaving a large amount of pores in the film if it was heated continuously up to 2400 °C [36]. In the case of polyimides, the starting molecular structure of polyimide used is known to govern the structure and nanotexture of the resultant carbons. Carbonization process of pitches, which have been used in many carbon industries, various organic resins, such as phenol and furfuryl alcohol, and various biomasses, is much more complicated, because they have much more complicated molecular compositions.

In Table 1, the carbonization processes are classified on the basis of the intermediate phases (carbonaceous materials) to the final carbon materials and the representative carbon materials formed are listed with their characteristics.

As can be seen from Table 1, most of carbon materials belong to the graphite family, of which the fundamental structural unit is a stack of carbon layers with a strong anisotropy. The way how the structural units tend to agglomerate is determined during carbonization process by the schemes and degrees of their preferred orientation, and results in a variety of nanotextures in carbon materials, which have the determining effects on their properties and also on their structural changes at high temperatures. Therefore, carbonization is the most important process for producing carbon materials with various functions.

### 4.2. Nanotexture

A classification of nanotextures in carbon materials, which are constructed by fundamental structural units with nanosize, has been proposed as shown in Figure 10 [48,54]. Firstly, random and oriented nanotextures are differentiated and, secondly, the latter is divided accordingly to the scheme of orientation, in parallel to a reference plane (planar orientation), along a reference axis (axial orientation) and around a reference point (point orientation). In Figure 10, some representative carbon materials are listed for each nanotexture.

<table>
<thead>
<tr>
<th>TABLE 1 Carbonization Processes for the Production of Various Carbon Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonization process</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Gas-phase carbonization</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Solid-phase carbonization</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Liquid-phase carbonization</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
The extreme case of planar orientation is a graphite single crystal, i.e., a perfect orientation with a regular stacking of large-sized hexagonal carbon layers. Single crystals can be found in either naturally occurring graphite (natural graphite) or in precipitated graphite flakes from molten iron (kish graphite). However, finding large-size crystals (more than 10 mm) is extremely difficult.

Various intermediate nanotextures between the perfectly planar and random orientations are found in pyrolytic carbons and coke particles, depending on the preparation and heat-treatment conditions. In pitch-derived cokes, the planar orientation of the layers is improved with increasing the heat-treatment temperature (HTT), as demonstrated through high-resolution transmission electron microscopy (HR-TEM)[115,117].

The extreme case of co-axial orientation is carbon nanotubes, including single-wall and multi-walled nanotubes. An axial orientation of the layers is found in various fibrous carbon materials; in other words, a fibrous morphology is possible because of this axial orientation scheme. In carbon fibers, the co-axial and radial alignments of the layers along the reference axis (i.e., fiber axis) are possible.

In the point orientation, the concentric and radial alignments have also to be differentiated (Figure 10). The extreme case of concentric point orientation is found in the fullerene family. This orientation is also found in spherical particles of carbon blacks, of which the diameter is from few tens to few hundreds nanometers, fundamental structural units consisting of minute hexagonal carbon layers being preferentially oriented along the surface (Figure 11a). A radial alignment of the layers to form a sphere is found in the carbon spherules, which are formed by pressure carbonization of a mixture.
of polyethylene and poly(vinyl chloride) (Figure 11c) [47,159]. The radial point orientation scheme appears near the surface of most mesosphere spheres (mesocarbon microbeads, MCMBs), but in their center the orientation of the layers is not radial (Figure 11b) [2,38]. As a consequence of the radial orientation near the surface, tangential cracks are formed in most of MCMB particles after heat treatment at high temperatures [45,46].

A randomly oriented nanotexture occurs in glass-like carbons and also in carbon materials obtained by carbonization of polymer precursors (e.g., phenol resin and sugar). The fundamental structural units composing most glass-like carbons are so small that they are difficult to be observed by TEM. Therefore, their nanotexture was often discussed on the basis of TEM observations on high-temperature-treated materials, where the layers become somewhat larger. A 002 lattice fringe image of sugar coke heat-treated at a high temperature is shown in Figure 12a [117] and a model of nanotexture in Figure 12b [135]. In this model, closed pores are formed by the concentric alignment of small structural units. By taking the large amount of closed micropores in most glass-like carbons and their gas impermeability into consideration, this model is believed to be realistic.

4.3. Graphitization

Heat treatment at high temperatures, in most cases up to 3000 °C, in inert atmosphere is an essential process for the production of polycrystalline graphite blocks. The principal purpose of this treatment is the improvement in structure, i.e., the increase in the graphitization degree. The development of the graphite structure is evaluated by different parameters. It is exactly evaluated by the graphitization degree \( P_1 \), which is the probability to have the AB stacking sequence for neighboring hexagonal carbon layers. Conventionally, however, the average interlayer spacing between neighboring layers \( d_{002} \) and the crystallite sizes, thickness along the c-axis \( L_c \), and lateral size of the layers \( L_a \) which are determined by X-ray diffraction analysis are often used as parameters to estimate the graphitization degree. The \( d_{002} \) values of highly crystallized graphite and amorphous carbon are 0.3354 nm and more than 0.344 nm, respectively. Figure 13a–c show the variation of \( d_{002} \), \( L_c \), and \( L_a \) versus heat treatment temperature, HTT, for various carbon materials with different nanotextures [59]. The needle-like coke mainly with a planar orientation and vapor-grown carbon fibers with a coaxial orientation show a rapid decrease in \( d_{002} \) and a rapid increases in \( L_c \) and \( L_a \) with increasing HTT, \( d_{002} \) approaching 0.3354 nm and \( L_a \) being more than 100 nm. Carbon blacks, thermal and furnace blacks, which have a co-axial point orientation, give relatively high \( d_{002} \) values of 0.339–0.341 nm and small \( L_c \) and \( L_a \), even after the heat treatment up to 3000 °C, the former being able to reach the smaller \( d_{002} \) and the larger \( L_c \) and \( L_a \), mainly because of larger particle size than the latter with smaller size. Glass-like carbon with a random orientation gives much larger \( d_{002} \) and very small \( L_a \) even after 3000 °C treatment.

It had been proposed to classify the carbon materials into two classes, graphitizing (graphitizable and sometimes soft) and nongraphitizing (nongraphitizable and hard) carbons: the former can change to graphite by high temperature treatment under atmospheric pressure but the latter cannot. However, this classification is not recommended now, because intermediate behaviors are very often
5. VARIOUS CARBON MATERIALS

In Table 1, various carbon materials are listed up on the basis of carbonization processes: gas-, liquid-, and solid-phase carbonization. Here, various carbon materials, which have been widely used in the industries and listed in Table 1, are briefly described on their preparation process and fundamental properties by dividing into three carbonization processes. Some of carbon materials shown in Table 1, however, are described in more detail later because most of them can be called advanced carbon materials.

5.1. Gas-Phase Carbonization

When carbon materials are produced through the decomposition of hydrocarbon molecules in gaseous phase, the process is called gas-phase carbonization. A large variety of carbon materials over the whole carbon families were produced through this process (Table 1). For a high concentration of hydrocarbon gases, carbon blacks are obtained. When a solid substrate is placed in the carbonization system, pyrolytic carbons are produced under a low concentration of hydrocarbon gases. If some fine metallic particles, such as Fe or Ni,
are added into the carbonization system, various carbon materials are produced: carbon nanotubes, carbon fibers (vapor-grown carbon fibers), carbon nanofibers with various nanotextures and morphologies. From the carbon vapor produced by electric-arc discharge or laser ablation, carbon nanotubes and fullerenes are formed.

5.1.1. Carbon Blacks

Carbon blacks formed through incomplete combustion of either gaseous or mist-like hydrocarbons are very important industrial products [59]. They have been used as a colorant in inks since the third century AD and are now applied in a large amount for rubber reinforcement. They are characterized by spherical primary particles, which are in different sizes and more or less coalesced into aggregates, what is called “structure” in the field of rubber reinforcement. In Figure 14, transmission electron micrographs of some carbon blacks are shown. On the basis of the reaction process and raw materials, carbon blacks are classified into furnace black, channel black, lamp black, thermal black, and acetylene black; furnace black consists of the primary particles of few tens nanometer diameters with a marked “structure” as shown in Figure 14c and d, which are used for the reinforcement of tires, but lamp black and thermal black consist of larger primary particles of few hundreds nanometer diameters without noticeable “structure” (Figure 14a). Ketjenblack is produced as a by-product of heavy oil gasification and has a size of primary particles of 30–40 nm and a high surface area (ca. 1300 m²/g) [99]. Ketjenblack and also acetylene black are often used as a conductive additive in the electrodes of primary and secondary batteries, and also for electrochemical capacitors.

The concentric orientation of the carbon layers is more marked near the surface, whereas the layers are randomly oriented at the center, of which 002 lattice fringe image of TEM is shown in Figure 15a. When large-sized carbon black particles are heat-treated at a high temperature as 3000 °C, the particles are polygonized and a hollow core appears (Figure 15b), because of the crystallite growth.

5.1.2. Pyrolytic Carbons

Carbon deposition occurs on the surface of a substrate inserted into the carbonization system from hydrocarbon gases, such as methane and propane [59]. This process is a chemical vapor deposition (CVD) and the products are called pyrolytic carbons. The deposition conditions are

![Figure 14](image-url) Transmission electron micrographs of various carbon blacks.
known to govern strongly the nanotexture of the resultant pyrolytic carbons. The deposition can occur on either static or dynamic substrates. In the former, the substrate is placed in a furnace, which is heated by direct passing either of electric currents or from the surroundings. For a solid substrate, graphite plates and also the mats composed from carbon fibers are often used. In the latter, small substrate particles are fluidized in the furnace. The factors controlling the nanotexture and properties of pyrolytic carbons are (1) the kind of precursor hydrocarbon gas, (2) the gas concentration, (3) the deposition temperature, (4) the contact time between the gas molecule and the substrate heated at a high temperature, which is governed by the gas flow rate and the size of the furnace, and finally (5) the geometrical arrangement of the furnace, particularly the ratio of the substrate surface area and the volume occupied by the gas [8].

In Figure 16, the bulk density of pyrolytic carbons formed on the solid substrate was plotted against the deposition temperature as a function of the precursor gases and their partial pressures. Two representative cross-sectional optical textures are also shown in the figure. In order to prepare pyrolytic carbon with high density and singularly-nucleated texture at 1400–2000 °C, it is necessary to use a low methane pressure, as low as 2 Pa. High methane pressure as 400 Pa gives low-density pyrolytic carbon in this temperature range, and the resultant pyrolytic carbon has regenerative optical texture, which is...
known to be mainly due to the formation of carbon black particles incorporated into the pyrolytic carbon.

The plates of so-called highly-oriented pyrolytic graphite (HOPG) [7] are produced by the annealing under high-temperature and high-pressure conditions. HOPG has a very high degree of planar orientation of the hexagonal carbon layers, but the size of the layers is not so large. In other words, the structure is close to a perfect planar orientation in the cross-section of the HOPG plate, but it is polycrystalline in the plane parallel to the plate [164].

The SEM image on a fractured cross-section and a channeling contrast image on a surface of an HOPG plate are shown in Figure 17a and b, respectively. A well-ordered stacking of the hexagonal carbon layers is seen in Figure 17a, whereas Figure 17b demonstrates the polycrystalline nature, consisting of small grains of graphite basal planes in different a-axis orientations.

The so-called carbon/carbon composites were fabricated by the filling of the pyrolytic carbons in the spaces between carbon fibers, the process being often called chemical vapor infiltration (CVI).

5.1.3. Vapor-Grown Carbon Fibers

Vapor-grown carbon fibers (VGCFs) are formed on a substrate from benzene vapor in a flow of high-purity hydrogen using fine iron particles as catalyst [19,116]. The high-resolution TEM analysis using various modes (bright-field images, selected area electron diffraction, dark-field images with 002 and 10 diffractions, 002 and 10 lattice fringes) shows that the VGCFs are constituted of two parts: a hollow tube with few straight layers (carbon nanotube) (Figures 18a and c) and small layers deposited on the carbon nanotube and oriented preferentially and concentrically along the fiber axis (Figures 18a and d) [116]. The formation of similar VGCFs was reported by using different precursor hydrocarbon gases, carrier gases, catalyst metals and deposition conditions [64,65,75,151]. In order to get a high yield of VGCFs, the floating catalyst method was developed [20], in which nanometric catalyst metal particles are formed in situ by the decomposition of its precursor, instead of the method where the catalyst is seeded on the surface of a substrate plate (seeding catalyst method).

Since most VGCFs can be changed to well-graphitized carbon fibers and they can be prepared with a high purity, without the coexistence of other carbon forms like carbon blacks, several works were performed to prepare fibrous carbons through the decomposition of various gases, not only hydrocarbons but also CO, under different conditions [1,3,6,128,137].

By selecting the decomposition conditions, carbon helical microcoils with a controlled pitch were prepared [12,108,109]. Fibrous carbons with a stacking of cup-like carbon layers along the fiber axis, cup-stacked nanofibers, were also prepared in a similar process [24,25]. Carbon nanohorns, showing a high-storage capacity for methane, were also synthesized by a similar technique [44]. The addition of a small amount of VGCF into the electrode of lead-acid and lithium-ion rechargeable batteries was reported to improve their performance [23].

5.1.4. Carbon Nanofibers

CVD processes under the conditions similar to VGCF production was reported to give various carbon nanofibers. Some of the papers, particularly in the beginning, reported that they prepared carbon nanotubes, but most of their products were carbon nanofibers.

A concentric alignment of the layers is realized in multiwalled carbon nanotubes and various carbon nanofibers, most of which are grown through CVD using fine catalyst particles. However, different orientation schemes can be found along their axis, from parallel (tubular type) to perpendicular (platelet type) through herring-bone type, as illustrated in Figure 19. Examples of herring-bone- and tubular-type nanotextures are shown in Figure 20 for both as-prepared and high-temperature-treated carbon.
5.1.5. Carbon Nanotubes

Carbon nanotubes were found to be formed at the beginning of VGCF formation, as shown in Figure 18 [116], and later formed in the carbon deposits on the graphite anode after arc discharge in He atmosphere [42]. The carbon nanotubes consisting of a single carbon layer were also found later [5,43]. The temperature at the graphite electrode during arc discharging was estimated to reach up to 2500 °C. By selecting the arc discharge conditions between the graphite electrodes, relatively high yields of carbon nanotubes were reported. The hollow tubes thus obtained show a wide range of diameters from 1 to 50 nm and their walls consist of different numbers of carbon layers. Most of them are closed at the end with the help of pentagons, the smallest diameter being the same as the size of the smallest fullerene C₆₀.

5.1.6. Diamond and Diamond-Like Carbons

Diamond was firstly synthesized under a high static pressure of about 10 GPa by using a catalyst of either Fe or Ni [10]. Diamond synthesis was carried out under a high pressure with various conditions, including static and shock-wave compression. Also, diamond crystals were synthesized by CVD of hydrocarbon gases, such as CH₄, C₂H₂, and C₆H₆, under a low pressure of about 1 Pa at around room temperature [73].
Diamond-like carbon (DLC) is synthesized via gas-phase carbonization of hydrocarbons. It is as hard as the diamond crystal, because of $sp^3$ covalent C–C bonding, but has an amorphous structure due to random orientation of carbon tetrahedra, as explained before. It contains a relatively large amount of hydrogen. A large number of DLC films were examined by using the X-ray absorption fine structure spectroscopy [131], most of them containing of $sp^2$-hybridized carbon atoms in a range of 20–60%.

5.1.7. Fullerenes

The carbon cluster $C_{60}$ was firstly found in the soot formed by laser ablation of graphite in He atmosphere and named as buckminsterfullerene [85]. $C_{60}$ was also found in the carbon deposits formed by arc discharge between graphite electrodes [83]. This carbon cluster $C_{60}$ could be extracted using solvents, such as toluene, carbon disulfide, etc. and also be separated from the deposits by vaporization at around 400 °C in vacuum [84]. Carbon clusters with different sizes, consisting of different numbers of carbon atoms, such as $C_{70}$, $C_{82}$, ..., $C_{960}$, were identified and this series of clusters was called fullerenes. Among the different gas atmospheres which were examined for the formation of fullerenes, helium atmosphere gives relatively high yield of fullerenes in most cases.

Fullerene inserted by $H_2$, $H_2@C_{60}$, was successfully synthesized through an organic synthesis process [78]. Also, Ar was inserted into fullerene to give Ar$@C_{60}$, and, by doping K into the interstitial sites of these clusters, $K_3Ar@C_{60}$, was shown to improve superconductive performance [139]. The Gd-metallofullerolol Gd$@C_{82}(OH)_{10}$, where Gd atoms are inserted into the $C_{82}$ cage and $-OH$ radicals are attached on the carbon atoms of $C_{82}$, has been reported to be useful as a contrast agent for magnetic resonance imaging (MRI) [74].

5.2. Solid-Phase Carbonization

Thermosetting resins, such as phenol-formaldehyde and furfuryl alcohol, and also cellulose can be converted to carbon materials by solid-phase carbonization. When precursors are carbonized quickly, the resultant carbon materials become porous, as being represented by activated carbons. If the carbonization is performed so slowly that the resultant carbonaceous solids can shrink completely,
the so-called glass-like carbons are produced, which contain a large amount of closed pores.

5.2.1. Activated Carbons

In activated carbons, a large amount of open pores are formed, which are usually evaluated by surface area and pore-size distribution [15,101]. They have been used as adsorbents since prehistoric age and are now used widely not only in our daily life but also in various industries in the processes for both production and waste treatment. For the predominant applications of activated carbons as adsorbents, the pore texture is the most important property to be controlled. Figure 21 shows a conventional process for the production of activated carbon; various carbon precursors, not only thermosetting resins, including natural biomasses, but also some thermoplastic resins, such as pitches and coals, are used. For thermoplastic precursors, the so-called stabilization, which is a slight oxidation, is required prior to carbonization. In order to develop micropores with a width less than 2 nm, mild oxidation of carbonized materials is usually carried out, which is called “activation”. For activation, different processes are employed, gas activation (e.g., diluted oxygen gas, air, steam, CO$_2$, etc.) and chemical activation using H$_3$PO$_4$, ZnCl$_2$, KOH, etc. During gas activation, the pores may grow from micropores to mesopores and finally to macropores. Figure 22 shows the changes in pore size during air activation of glass-like carbon spheres [60]. At the very beginning of oxidation, ultramicropores (<0.7 nm width) are formed, presumably by opening of closed pores existing in the glass-like carbon matrix, and then ultramicropores grow to supermicropores (0.7–2 nm width) and then mesopores (2–50 nm width). For chemical activation, the carbonization and activation processes are often carried out in one heat-treatment process. Recently, a great success to get a high surface area reaching 3000 m$^2$/g was obtained by using KOH for activation [100].

The pore structure of most activated carbons is illustrated in Figure 23a, where macropores (>50 nm width) and mesopores are usually coexisted with micropores. These large-sized pores, macropores and some of mesopores, work as diffusion paths for oxidizing agents to create micropores during activation process and also as those for adsorbate molecules to reach the micropores during adsorption application. The pore-size distributions of activated carbons used for the adsorption of small gas molecules and decolorization of water are shown in Figure 23b. By the control of micropore size, the activated carbons may be successfully used as molecular sieves for various gases. Mesopores are effective for the adsorption of gasoline vapor composed from large hydrocarbon molecules [69]. Pore-size distribution of the activated carbon for vehicle
5.2.2. Glass-Like Carbons

Glass-like carbons (glassy carbons) are produced by the pyrolysis of different precursors, such as phenol-formaldehyde resin, poly(furfuryl alcohol), and cellulose, through an exact control of the heating process \[28,114\]. They are characterized by amorphous structure and also by properties very similar to inorganic glasses, such as high hardness, brittle conchoidal fracture, as shown in Figure 24a, and gas impermeability. The most crucial parameter for their production is a very slow heating, slower than the rate of shrinkage of carbonaceous materials on the carbonization process to compensate the formation of open pores due to the evolution of decomposition gases from the precursor block. In Figure 24b, the changes in weight, volume, BET surface area, and adsorption of water vapor with carbonization temperature are shown for poly(furfuryl alcohol) condensates \[28\]. The rapid decreases in weight and volume up to 700–800 °C are due to the precursor pyrolysis and correspondingly the BET surface area and adsorption of water vapor increase. Above 800 °C, however, the BET surface area and water vapor adsorption decrease quickly under very slow heating rate, because of the fact that most of pores are closed.

The commercially available glass-like carbons have a bulk density of 1.3–1.5 g/cm³, suggesting the existence of a large amount of pores, and show very low gas permeability as \(10^{-12}\) cm²/s, almost gas impermeable, suggesting that all of the pores are closed. The formation of closed pores with rather homogeneous size of about 5 nm was confirmed on a glass-like carbon, the volume of which reached around one-third of the bulk \[112\].
5.2.3. Carbon Fibers

In Table 2, carbon fibers are classified on the basis of their precursor, and their characteristics are summarized, including vapor-grown carbon fibers (VGCFs) prepared through gas-phase carbonization for the comparison. Commercially available carbon fibers are classified into two grades, general purpose and high-performance grades on the basis of their mechanical properties. The latter is further divided into two types, high-modulus and high-strength types [16]. In order to produce carbon fibers from precursor polymers (for example, PAN) and various pitches, stabilization is essential after their spinning, which consists of chemical reaction by using different oxidizing gases, such as air, oxygen, chlorine, and hydrochloric acid vapor [16]. The stabilized fibers are then subjected to solid-state carbonization (extrinsic solid-phase carbonization) in inert atmosphere. When cellulose and phenol fibers are used as precursors, the stabilization process is not necessary, and the spun fibers are subjected directly to carbonization (intrinsic solid-state carbonization).

In Figure 25, the variety in nanotexture of different carbon fibers is illustrated, on the cross-sections parallel and perpendicular to the fiber axis [59]. Most of thermosetting precursors, such as phenol and cellulose, and also isotropic pitches, give carbon fibers with random nanotexture in both cross-sections. For mesophase-pitch-based carbon fibers, however, different nanotextures in the cross-section perpendicular to the fiber axis, from radial and coaxial to random, are possible (Figure 25) [51]. VGCFs have a co-axial alignment of small carbon layers on a carbon nanotube at the center (Figure 18). Figure 26 shows, on mesophase-pitch-based carbon fibers, that the cross-sectional nanotexture governs also the graphitization behavior of the fibers [51]; fibers with straight radial nanotexture in its cross-section have a relatively high graphitization degree, expressed by low d002 and high La values, but those with zigzag radial nanotexture have a poor graphitization degree. In VGCFs, the small layers coalesce with each other forming long and smooth layers after 3000 °C treatment; the degree of concentric alignment is greatly improved and the perpendicular cross-section of the fiber is polygonized, mainly due to crystallite growth [165]. The difference in the nanotexture in carbon fibers becomes more pronounced after high-temperature treatment.

<table>
<thead>
<tr>
<th>Precursor</th>
<th>Carbon fibers</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyaniline (PAN)</td>
<td>PAN-based carbon fibers</td>
<td>Different grades and types</td>
</tr>
<tr>
<td>Isotropic pitch</td>
<td>Isotropic-pitch-based carbon fibers</td>
<td>General purpose grade</td>
</tr>
<tr>
<td>Mesophase pitch</td>
<td>Mesophase-pitch-based carbon fibers</td>
<td>Random cross-sectional nanotexture</td>
</tr>
<tr>
<td>Cellulose</td>
<td>Cellulose-based carbon fibers</td>
<td>General purpose grade</td>
</tr>
<tr>
<td>Phenol</td>
<td>Phenol-based carbon fibers</td>
<td>General purpose grade</td>
</tr>
<tr>
<td>Hydrocarbon gases</td>
<td>Vapor-grown carbon fibers</td>
<td>High graphitizability</td>
</tr>
</tbody>
</table>

5.2.4. Carbon Films

Polyimide films with different molecular structures have been developed as thermoresistant polymers and used in different fields, especially in the field of electronics. The heat treatment of these films at high temperatures leads to a wide variety of carbon structures, from highly crystalline graphite to amorphous glass-like carbon films [53]. This is a typical case where the molecular structure of the organic precursors and the texture of their films govern the structure and texture of the resultant carbon films, i.e., crystallinity and preferred orientation of hexagonal carbon layers. As shown on a polyimide film (commercially available film Kapton) in Figure 9b [50], the yield after carbonization is relatively high, about 60 mass%, and the linear shrinkage along the film is slightly larger than 20%. Even if these changes in weight and size are not so small, it is noteworthy to mention that the film can keep its original form, although
a little care not to give mechanical shock is required. **Figure 27** shows a crane made from a polyimide film by a technique of paper folding, the form of which is maintained after carbonization up to 1300 °C and even after graphitization at about 3000 °C.

Carbon films were also prepared from other organic films, poly(phenylene vinylene) (PPV) [118] and poly-p-phenylene-1,3,4-oxadiazole (POD) [110,111,134].

From the polyimide films with different molecular structures, either consisting of a large number of phenyl rings or containing fluorine, highly microporous films were synthesized by a simple carbonization at around 1000 °C in an inert atmosphere [119,120]. The polyimide film having 31.3 mass% fluorine in its repeating unit of molecules gave a highly nanoporous carbon film having a microporous surface area of 1340 m²g⁻¹ and consisting of micropores with a width of ca. 0.55 nm [119]. By controlling carbonization temperature, microporous carbon films were prepared, which had molecular sieving function [27,33]. Carbon films containing macropores with a diameter less than 1 µm were prepared from a polyimide by the introduction of phase inversion technique during the gelation of poly(amic acid) [32]. Co-polymers of polyimide and polyurethane gave macroporous carbons [140].

The preparation of microporous carbon films and their properties were recently reviewed [149].

5.3. Liquid-Phase Carbonization

Liquid-phase carbonization occurs for some precursors, such as pitches, which become viscous fluids before carbonization. Through this process, the cokes are prepared from petroleum and coal tar pitches, which are used as the raw materials for industrial production of various polycrystalline graphite blocks as the electrodes for steel refining and electrical discharge machining, the jigs for the growth of semiconductor crystals, the structure components of the nuclear reactors, etc. Polycrystalline graphite blocks are fabricated by using a filler coke, which is prepared from pitches through liquid-phase carbonization, and a binder pitch, which is carbonized in between the particles of the filler coke in a formed block also through liquid-phase carbonization. Spheres with optically anisotropic texture are formed at the beginning of the liquid-phase carbonization of pitches and are separated from the isotropic pitch matrix, which are called mesocarbon microbeads (MCMBs).

5.3.1. Cokes

Cokes are the products of liquid-phase carbonization of pitches. In the course of this process, optically anisotropic spheres are firstly formed, which are called mesophase spheres. By further heating, the spheres grow and coalesce with each other to give liquid crystals with different
textures composed of optically anisotropic units, called bulk mesophase \([9,38]\). A series of polarized-light microscopy images in Figure 28 shows the sequence from the formation of small mesophase spheres to their growth and partial coalescence. By further heating, their bulk mesophase becomes solid with anisotropic textures.

By applying a shear stress during liquid-phase carbonization, the so-called needle-like cokes are produced, which are now important raw materials in the production of large-sized graphite electrodes for metal refining \([59]\). In the needle-like coke, basic structural units are well oriented along the shear direction, as shown by polarized-light micrograph in Figure 29. The structure improvement, i.e., graphitization, occurs markedly in the cokes with increasing heat-treatment temperature, as shown by plotting XRD parameters against HTT for a needle-like coke in Figure 13. In Figure 30, 002 lattice fringe image of TEM is shown on a coke with different HTTs, revealing that the size of the layer planes observed as 002 fringes grow and their parallel stacking is improved with increasing HTT.

5.3.2. Mesocarbon Microbeads

The anisotropic mesophase spheres shown in Figure 28b can be separated from the isotropic pitch matrix by using a solvent \([160]\). The separated spheres are called mesocarbon microbeads (MCMBs) and have been used as anode material of lithium-ion rechargeable batteries \([150]\). Because of their nanotexture of radial point orientation, many cracks in radial direction are formed, accompanying certain structure improvement, after the high-temperature treatment, as shown in Figure 31, which seem to make intercalation/deintercalation of lithium ions easier.

FIGURE 26 Nanotexture in the perpendicular cross-section of mesophase-pitch-based carbon fibers and their lattice parameters after 3000 °C treatment.

FIGURE 27 Crane made from a polyimide film by paper folding technique and heat-treated at different temperatures. For color version of this figure the reader is referred to the online version of this book (Courtesy of Dr. A. Hatori of AIST, Japan)

(a) Polyimide film (b) After carbonization (c) After graphitization
5.3.3. Polycrystalline Graphite Blocks

The production procedure of polycrystalline graphite blocks is shown in Figure 32 [59]. For the filler, cokes derived from petroleum and coal tar pitches are usually used, but natural graphite, carbon blacks, and also recycled graphite particles are sometimes added. In advance, the cokes are carbonized in order to avoid the evolution of a large amount of volatiles, which introduces some shape distortions and cracks in the product and also reduces the density. For the binder, petroleum and coal tar pitches are used in most cases, because they have relatively high carbon yield as about 60 mass%, and also because they give a carbon similar to the filler cokes after carbonization. The particle size of the filler and the

![Figure 28](image1.png) Formation, growth, and coalescence of mesophase spheres in a pitch.

![Figure 29](image2.png) Appearance and cross-section of a needle-like coke. For color version of this figure, the reader is referred to the online version of this book.

![Figure 30](image3.png) 002 lattice fringe images of a needle-like coke heat-treated at different temperatures. (Courtesy of Mme A. Oberlin, France)
mixing ratio of the filler to the binder have to be controlled, in accordance with the requirements from the applications. The filler and the binder are mixed at a temperature higher than the softening point of the binder. The mixture thus prepared, which is usually called carbon paste, is formed after warming up at a temperature around 150 °C by either extrusion, molding, or cold isostatic pressing. The formed blocks are carbonized at a temperature of 700–1000 °C (called calcination in industries) and then graphitized at a high temperature above 2500 °C. The blocks heat-treated at a high temperature are polycrystalline, and often called “synthetic graphite” or “artificial graphite”. For producing graphite blocks, liquid-phase carbonization is involved in two steps: i) during the process of filler coke preparation and ii) during the carbonization of the binder pitch in a carbon paste.

Forming is an important process for the fabrication of polycrystalline graphite blocks, because it governs the preferred orientation of the filler particles, which are often anisotropic. The extrusion gives a preferred orientation of either flaky or needle-like filler particles along the direction of extrusion. Electrodes for metal processing with a large diameter up to 80 cm, carbon rods with different sizes, and also leads for automatic pencil with a diameter as thin as 0.3 mm are fabricated by this forming process. In graphite electrodes used in steel refining, needle-like cokes are essential to attain a high thermal shock resistance. Sudden and large expansion due to abrupt increase in temperature can be absorbed into a crack formed in needle-like coke particles as shown by arrow in Figure 29b. In the molding process, which is applied for the fabrication of carbon brushes for electric motors and electric contacts, the filler particles are statistically aligned perpendicular to the compressing direction. By isostatic pressing, the filler particles are randomly oriented, which lead to producing isotropic high-density graphite blocks, which will be described in detail later.

5.4. Novel Carbonization Processes

The recent developments in science and technology require a more exact control of structure/nanotexture and properties of carbon materials. In order to meet these requirements, various novel carbonization processes have been proposed: template method, polymer blend method, defluorination of fluorinated hydrocarbons, and carbonization of organic aerogels [57].

5.4.1. Template Methods

The template carbonization technique was firstly developed for the preparation of thin oriented graphite films using two-dimensional spaces in clay minerals with a layered structure [88,138]. Template carbonization reveals a high possibility to control the dimensionality of morphology by selecting suitable template materials: one-dimensional carbon nanofibers were synthesized using anodic aluminum oxide films, two-dimensional graphite layers using layered compounds, and three-dimensional nanoporous carbons using zeolites, silicas, MgO, and some surfactants [63].

Carbon nanofibers were prepared by carbon deposition from propylene gas at 800 °C on the inner walls of nanosized channels in an anodic aluminum oxide film [89]. Aluminum oxide is subsequently dissolved either by HF at room temperature or by NaOH aqueous solution at 150 °C in an autoclave. The procedure is schematically shown in Figure 33 with a SEM image of the resultant nanofibers. The tubular carbon deposits in the channels of aluminum oxide could be preferentially modified by fluorine and oxygen [34,92], and also easily filled by metal and metal oxide to prepare their nanowires [11,91,121,126,127]. The tubular carbons thus prepared could be easily converted to multiwalled carbon nanotubes by the heat treatment at high temperatures in inert atmosphere.

In order to control the pore structure in carbon materials, various templates were developed, as summarized in Table 3 by listing template, carbon precursors, and characteristics of the pore structure in the resultant carbons.

Carbons with a high BET surface area, even higher than 3000 m²/g, were prepared using the three-dimensional channels of zeolite, of which micropores had homogeneous size and were ordered [90,93,97,98,103]. By using mesoporous silicas, mesoporous carbons were formed, pores being possible to be either ordered or disordered and also to be channels [70,71,86,94,130]. By using some surfactants, mesoporous carbons could be prepared directly from furfuryl alcohol and polyacrylonitrile [81,95,147].

MgO was also used as template for obtaining mesoporous carbons [106,107]. A thermoplastic carbon precursor (such as poly(vinyl alcohol), poly(ethylene terephthalate), and pitch) is mixed with a MgO precursor.
(MgO, Mg acetate, Mg citrate, and Mg gluconate), and heat-treated at 900 °C in inert atmosphere. Then MgO formed through the pyrolysis of its precursor is dissolved by a diluted acid solution, liberating the mesoporous carbon. It was proved that the mesopore size of the resultant carbon is governed by the size of MgO thus formed. A large amount of mesopores with a size of about 5 nm is easily obtained using Mg citrate [104]. Figure 34a and b show the pore volume and size distribution, respectively, as a function of the mixing ratio between Mg citrate and PVA. From the mixture of Mg citrate and gluconate, a carbon with bi-modal mesopore sizes was obtained [105].

Microporous and/or mesoporous carbons were proposed to be used as adsorbents for various molecules with a wide range of sizes, electrodes for electrochemical capacitors, etc. [63]. By impregnating a polyimide into organic foams, commercially available polyurethane and melamine foams, carbon foams were prepared after carbonization [58]. Micropores could be easily created in the walls of macropores.

5.4.2. Polymer Blend Method

By spinning of a mixture of two kinds of carbon precursors, the one giving a relatively high carbonization yield and the other having a very low yield, and following carbonization, carbon nanofibers were successfully prepared.
Poly(urethane-imide) films were easily converted to macroporous carbon films [66,140,141,142]. The films prepared by blending poly(amide acid) with phenol-terminated polyurethane prepolymers were heated up to 200 °C for the phase separation to polyimide (PI) matrix and small polyurethane (PU) islands. By heat treatment up to 400 °C, PU component was pyrolyzed to gases and resulted in porous polyimide films, which were able to convert easily to porous carbon films by carbonization in inert atmosphere. Pore size in these polyimide and carbon films was controlled by changing the blending ratio of PI to PU and also the molecular structure of PU. The carbon films prepared were shown to be suitable as a medium for culturing biological cells, with macropores in the range of 0.6–3.0 μm sizes being shown to be preferable for the cells [82].

### 5.4.3. Defluorination of Polymers

Mesoporous carbons with a high surface area are obtained through the defluorination of poly(tetrafluoroethylene) (PTFE) with alkali metals [17,67,145,146,161]. PTFE film with 100 μm thickness was pressed with lithium metal foil with 200 μm thickness under 4 MPa in Ar atmosphere for 48 h to defluorination of PTFE. Excess lithium metals and finely dispersed LiF were eliminated by washing with methanol, heating up to 700 °C, and washing again by dilute HCl [161]. Defluorination of PTFE was also possible through heating a mixture of PTFE powders with alkali...
metals (Na, K, or Rb) in vacuum at 200 °C in a closed vessel [141,146]. Defluorination of PTFE with Na metal was found to give mesopore-rich carbon and very high $S_{BET}$ as 2225 m$^2$g$^{-1}$.

5.4.4. Carbon Aerogels

Mesoporous carbons were also prepared by the pyrolysis of resorcinol-formaldehyde organic aerogels [4,30,31,124,125,143,144,163]. Primary carbon particles have the size of about 4–9 nm and interconnected with each other to form a three-dimensional network, which gives predominantly interparticle mesopores. Adsorption isotherms of the resultant carbon aerogels belong to type IV and have a clear hysteresis [31]. Pore structure of carbon aerogels was known to be governed by that of precursor organic aerogels, which was controlled by the mole ratios of resorcinol to formaldehyde (R/F), to water (R/W), and also to basic catalyst Na$_2$CO$_3$ (R/C) [143]. Instead of supercritical drying of aqueous gels, freeze-drying method was also applied [144]. On the gels prepared through freeze drying (cryogels), much smaller shrinkage in pore size during carbonization was observed.

6. IMPORTANCE OF TEXTURES IN CARBON MATERIALS

6.1. Nanotextures

Nanotexture in carbon particles, which is established in the process of carbonization, governs their structure and properties. As shown in Figure 13, structural improvement in carbon materials with the heat treatment under atmospheric pressure depends strongly on their nanotextures; needle and regular cokes with planar orientation can approach graphite structure above 2500 °C, but glass-like carbon with random orientation is almost impossible to be graphitized even at very high temperature as higher than 3000 °C. Thermal and furnace blacks with point orientation have an intermediate behavior between the carbon with

FIGURE 34 Pore volume and pore-size distribution of mesoporous carbons prepared by MgO-templated method.

FIGURE 35 Preparation scheme of carbon nanofibers through polymer blend method. (Courtesy of Prof. A. Oya of Gunma University, Japan)
planar orientation and that with random orientation, the former being improved its structure with heat treatment more markedly than the latter. Graphitization behavior for vapor-grown carbon fibers, which has an axial orientation, is similar to the cokes, but that for other carbon fibers is governed by their axial orientation schemes along their fiber axes, as exampled by using mesophase-pitch based carbon fibers in Figure 26.

In the case of intercalation reactions, the nanotexture of the host carbon materials has a strong effect. The intercalation of sulfuric acid into natural graphite can proceed at room temperature in concentrated acid. The resulting intercalation compound is commonly used in industry for the preparation of exfoliated graphite [56]. However, in order to intercalate sulfuric acid into carbon fibers, the electrolysis in the acid is needed [152,153,155]. Potassium as a vapor, on the other hand, can be easily intercalated in various carbon materials, even into low-temperature-treated carbon fibers.

The nanotexture once formed is extremely difficult to change; in other words, it needs very severe conditions to destroy the nanotexture formed. Glass-like carbons, which have random orientation, cannot change to graphite under normal pressure even above 3000 °C, as shown in Figure 13. In order to change them to graphite, either melting or heat treatment under a high pressure is needed. Melting point of carbon is located at a temperature around 4000 K and a pressure above 10 MPa. When a rod of glass-like carbon was heated by direct passing the electric current under Ar pressure, it could be melted at its middle part and quenched suddenly to a low temperature, because of its breaking. After melting, a ball was found in a crater formed by melting. It was shiny and lubricant graphite with exactly the same lattice parameters as natural graphite. However, the wall of the crater was far from graphite, of which 002 diffraction line was broad and unsymmetrical, and of which d 002 spacing was about 0.34 nm, even though it was heated close to melting point [113]. Glass-like carbon spheres could be graphitized under a high pressure as high as 0.5 GPa at 1600 °C [52,72]. Change in 002 diffraction profile with HTT under 0.5 GPa pressure was shown in Figure 36a. The 002 diffraction line has a composite profile, consisting of two structural components, the one located at a low angle side (T component) is due to the pristine glass-like carbon and the one at high angle side having d002 of about 0.335 nm (G component), which was found at the contact point between two spheres of glass-like carbon, supposing to be stress accumulated point, as shown by polarized-light micrographs in Figure 36b. With increasing HTT under pressure, relative intensity of the G component to the T component increases, revealing gradual increase in graphitized parts, and correspondingly optically anisotropic parts increase their areas. Finally, the single peak of the G component in X-ray diffraction and totally anisotropic area in optical micrograph were obtained at 1600 °C, revealing a completion of graphitization.

FIGURE 36 Change in 002 diffraction profile and polarized-light micrograph of a glass-like carbon with heat-treatment temperature under the pressure of 0.5 GPa.
Graphitization of glass-like carbon due to stress accumulation was also observed in the composite of carbon fibers with glass-like carbon matrix [35]. Polarized-light micrographs of the composites heat-treated at different temperatures under normal pressure are shown in Figure 37. Optically anisotropic area is formed preferentially at the interfaces between fiber and glass-like carbon matrix formed from furfuryl alcohol condensate after carbonization at 1000 °C (Figure 37b), although the composite before carbonization was optically isotropic (Figure 37a). This formation of anisotropic areas is reasonably supposed to be due to a large shrinkage of the matrix resin during pyrolysis and carbonization, although the carbon fibers do not shrink because they have been prepared at around 1300 °C. These anisotropic areas grow with increasing HTT (Figure 37c); finally, almost all area of the composite became anisotropic (Figure 37d), and the development of graphitic structure was proved by X-ray diffraction.

Carbon fibers with random orientation, such as PAN-based carbon fibers, are difficult to be graphitized by the heat treatment under normal pressure, but they can be graphitized when their nanotexture was destroyed by exfoliation [154,156]. In Figure 38, changes in Raman spectrum are shown on a PAN-based carbon fiber from the pristine fiber to that after exfoliation, 900 °C-annealing, and then the 3000 °C-treatment. The pristine fiber shows relatively strong D-band, although it was already heat-treated at 3000 °C, but the exfoliated carbon fiber shows very weak D-band after the heat treatment at 2800 °C.

**6.2. Microtextures**

Most particles with planar and axial orientation, such as cokes, carbon fibers, and carbon nanotubes, are also anisotropic and as a consequence their agglomeration in various carbon blocks can create a further variety in microtexture, mainly due to preferred orientation of anisotropic particles. Therefore, in addition to the nanotexture and also graphitization degree of each particle, it is necessary to take into consideration the microtexture formed.

---

**FIGURE 37** Change in polarized-light micrograph of carbon fiber/glass-like carbon composite with heat-treatment temperature. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 38** Change in Raman spectrum of a PAN-based carbon fiber with exfoliation, 900 °C-annealing, and 2800 °C-treatment.
The microtexture is usually formed during the forming process of bulky carbon materials. In large-sized graphite electrodes for metal refining, for example, the particles of needle-like coke tend to be oriented along the extrusion direction, i.e., the rod axis, during their forming process through extrusion with pitch binder. To prepare carbon fiber reinforced plastics (CFRPs), different microtextures based on the orientation of carbon fibers, some examples of which are shown in Figure 39, have been applied in order to get high strength and high modulus of the composites.

Three methods have been employed for realizing the isotropy of carbon materials which fundamentally consist of anisotropic structural units or crystallites: i) the random aggregation of micrometer-sized particles, even though these particles are anisotropic, ii) the aggregation of spherical particles of carbon, and iii) the random agglomeration of nanometer-sized crystallites in the bulk. The first method is realized in the so-called isotropic high-density graphite blocks where small size coke particles are formed by using isostatic pressing, often called rubber pressing. The second method was carried out by using mesocarbon microbeads. The last method, i.e., aggregation of nanosized carbon layers, was realized in glass-like carbons, which are isotropic, but have a large amount of closed pores and nongraphitizing nature.
The presence of pores in the carbon blocks influences the bulk properties of the material. In such a case, the microtexture, including the shape and size of pores, has to be taken into account in addition to that due to the orientation of anisotropic particles. Optical micrographs on polished sections of isotropic high-density graphite blocks having different bulk densities from 1.735 to 1.848 g/cm³ are shown in Figure 40 [122]. Although the difference in bulk density looks rather small, a marked difference is easily observed in the shape, size, and distribution of pores in the cross-section. Different pore parameters of these carbon materials, such as density, average cross-sectional area, roundness, and fractal dimension, were determined with the help of image analysis. Figure 41 shows the plots of mechanical properties (elastic modulus, bending strength, and fracture toughness $K_{IC}$) versus the average cross-sectional area of pores. A dependence of the mechanical properties of these carbons on pore size is clearly seen.

FIGURE 41  Dependences of mechanical properties on average pore area in isotropic high-density graphite blocks.

FIGURE 42  Appearance of the macropores in the worm-like particles of exfoliated graphite and the distribution histogram of cross-sectional area of pores.
Macropores in exfoliated graphite particles (customary called worm-like particles) can sorb a large amount of heavy oil (more than 80 g per 1 g of exfoliated graphite) [157]. An example of the fractured surface of a worm-like particle of exfoliated graphite is shown in Figure 42a. The histogram example of the fractured surface of a worm-like particle of oil (more than 80 g per 1 g of exfoliated graphite)[157]. A n called worm-like particles) can sorb a large amount of heavy oil[157].

The size and volume of these macropores govern the sorption capacity of heavy oil[157].

REFERENCES


Novel Carbon-Based Nanomaterials: Graphene and Graphitic Nanoribbons

E. Gracia-Espino, F. López-Urías, Y.A. Kim, T. Hayashi, H. Muramatsu, M. Endo, H. Terrones, Mauricio Terrones** and Mildred S. Dresselhaus

1. INTRODUCTION

In 1962, Hanns-Peter Boehm et al. [1] reported the isolation and identification of single graphene sheets, “Thinnest Carbon Films”. Recently, the two-dimensional monolayer crystalline allotrope of carbon, known as graphene, was isolated using the so-called “scotch tape method”, in which an ingenious technique for its observation under an optical microscope was also described [2,3]. The relative ease of the sample preparation using theNovoselov’s method and the fascinating properties of this 2D atomic crystal have stimulated extensive experimental and theoretical studies on the novel linear E(k) graphene carbon electronic structures in monolayer graphene exhibiting sp² hybridization in bilayer and few-layer graphene. Their related graphitic nanoribbons have subsequently emerged, each with novel and distinguishable characteristic properties. At present, there has been a surge in the number of reviews and dedicated journal issues on the synthesis and properties of graphene [4,5]. Concurrently, new developing routes for the synthesis of few graphene layers have emerged. The first one consists of the growth of ultrathin epitaxial graphite films by thermal decomposition on the (0001) surface of hexagonal faces of SiC single crystal [6,7]. The second one synthesizes films with single- to few-layer graphene (from 1 to 12 layers) using an ambient pressure chemical vapor deposition (CVD) method on a polycrystalline Ni substrate [8]. This last method was subsequently improved by Sun Z. et al. [9]; here the graphene layers were grown from different carbon sources, such as polymer films and small molecules deposited on a metal catalysis substrate.

Graphitic nanoribbons, which are quasi-1D graphene nanostructures, can exhibit a band gap between the valence and conduction band states. The band gap depends on both the edge morphology and ribbon width [10,11], which is typically a few nanometers, making graphene nanoribbons a very interesting material for potential electronics applications [12]. Graphite nanoribbons less than 10 nanometers wide are expected to be semiconductors, independent of...
their edge patterns [12]. Narrow nanoribbons are thus excellent candidates for use in electronic devices, such as field-effect transistors (FETs), which form the basis of microchips in computers. A thorough exploration of the chemical and mechanical properties of nanoribbons will undoubtedly suggest other applications for these structures, perhaps as sensors, catalysts, and scaffolds for tissue regeneration or components of composite materials.

This chapter reviews studies of graphene with special attention given to their related graphitic nanoribbons. Different production methods, characterization techniques, and theoretical calculations for nanoribbons are discussed. In particular, we emphasize the role of defects and chemical doping in the electronic properties of graphene nanoribbons. Results of the conductance of undoped and doped armchair nanoribbons are also described. Finally, applications and future work on graphene and graphene nanoribbon materials are discussed.

2. THEORETICAL STUDIES ON PHYSICAL—CHEMICAL PROPERTIES OF GRAPHENE NANORIBBONS

The graphene structure consists of a basic two-dimensional monolayer formed by sp² hybridized carbon atoms (three coordinated neighboring carbon atoms), which belongs to one of the five 2-D Bravais lattices called the hexagonal (triangular or honeycomb) lattice. It is noteworthy that by piling up graphene layers, in an ordered way with AB interlayer stacking order [13], one can form hexagonal 3-D graphite, which belongs to the P6/mmc (194) space group. Graphene was initially considered as a theoretical building block used to describe the graphite crystal, and to study the formation of carbon nanotubes (rolled graphene sheets), and to predict the electronic properties of sp² carbons generally. This 2-D atomic (one atom thick) crystal of sp² hybridized carbon has one dominant fingerprint: a unique electronic structure with a linear E(k) dispersion close to the Fermi level. Such an electronic structure describes massless fermions [3], which result in new physical phenomena. Graphene nanoribbons consist of finite size graphene crystals, thus forming doubly coordinated atoms at the edges, in contrast to the triply coordinated interior atoms. If the width of the ribbon is of the order of nanometers, we have a ribbon-like nanostructure that exhibits properties different from those observed in an infinite size graphene sheet.

2.1. Electronic Properties of Zigzag and Armchair Graphene Nanoribbons

Theoretical studies on graphene nanoribbons started with the work of Fujita et al. [10,14]. These authors studied the electronic states of graphene ribbons with armchair and zigzag edges, and found that graphene ribbons presented different properties depending on the edge shape. The zigzag ribbon shows a remarkably sharp peak in the electronic density of states at the Fermi level, which is not present in infinite 2-D graphene. The authors also found that the singular electronic states arise from the partly flat bands at the Fermi level, whose wave functions are mainly localized on the zigzag edge. They also applied the Hubbard model within the mean-field approximation, finding a stable magnetic order in zigzag nanoribbons. The atomic morphology of zigzag and armchair nanoribbons is depicted in Fig. 1a-b. In order to show the environmental dependence of the local density of states, non-spin-polarized density functional calculations were carried out on a zigzag and armchair nanoribbon (see Figure 1c and d). These calculations do not yield electronic states at the Fermi level for armchair nanoribbons (Figure 1c). However, zigzag nanoribbons do show a high density of electronic states at the Fermi level, with the main contribution coming from atoms located at the zigzag borders (see Figure 1d), as previously predicted by Fujita et al. [10].

2.2. Width and Length Dependence of the Electronic Properties of Graphene Nanoribbons

From the first studies on the electronic properties of graphene nanoribbons, within the limitations of the tight binding (TB) approximation, it has been shown that armchair graphene nanoribbons (A-GNR) exhibit semiconducting or semimetallic behavior depending on the ribbon’s width. It was also determined that they do exhibit a metallic behavior for N = 3p−1, where p is an integer (N is the number of zigzag chains perpendicular to the grown axis), and semiconducting behavior otherwise [10].

DFT calculations preformed by Son et al. [15] revealed that hydrogen passivated armchair and zigzag GNRs always have nonzero and direct band gaps. They also found that A-GNR can be grouped into three different families: Nₐ = 3p, Nₐ = 3p + 1 and Nₐ = 3p + 2, exhibiting in all cases different electronic band gaps (Eₐ² < Eₐ³>p+1 > Eₐ³>p+2). The band gaps for each family decrease with increasing p and eventually tend to zero as p → ∞. These results have been proved to have the same trend by using first-principles calculations based on many-body perturbation theory, such as the GW approximation [16]. Compared with previous TB and DFT studies, the quasiparticle band gaps calculated with the GW approximation show significant self-energy corrections for both armchair and zigzag nanoribbons; in all cases the band gaps were increased by self-energy corrections.
For Z-GNRs, local spin density approximation (LSDA) and GW calculations have both shown an antiferromagnetic insulating ground state with ferromagnetic coupling at each zigzag edge, and antiferromagnetic coupling across the ribbon. Two notable characteristics are found in the electronic structure of Z-GNRs: (1) the top of the valence band and the bottom of the conduction band are composed of mainly edge states and (2) the spin–orbit interaction introduces a finite band gap in the Z-GNRs [16]. A band gap engineering investigation by M. Y. Han et al. [17] in lithographically patterned graphene nanoribbons demonstrated that the energy gap scales inversely with the ribbon width. Inspired by this work, Sols et al. [18] have revealed that the existence of an energy gap in graphene nanoribbons may be understood in terms of Coulomb blockade effects.

A systematic study on the electronic properties of graphene nanoribbons was carried out by V. Barone et al. [19]. Based on DFT calculations, these authors analyzed the band gaps and the relative stabilities of semiconducting carbon nanoribbons with and without hydrogen termination, and widths up to 3 nm. They are able to predict the nanoribbon properties using an extrapolated inverse power law, that is, nanoribbons with band gaps on the order of the band gaps of bulk Ge and InN should have a width between 2 and 3 nm. If a larger band gap material is needed, with a gap comparable to that of Si, InP, or GaAs, the width of the carbon nanoribbons should be between 1 and 2 nm.

Zero-dimensional graphitic systems (i.e. finite width and length) have also been studied by different groups. Using DFT calculations, Shemella et al. [20] found that in addition to quantum confinement along the width of the ribbon, finite size effects emerge along the length of the ribbon, which modifies the electronic states of A-GNRs with zigzag terminated (length confined) edges. In this context, K-T. Lam et al. [23] investigated armchair bilayer graphene nanoribbons in the DFT framework. Their results show that the band gap ($E_g$) of a bilayer A-GNR, in general, is smaller than that of a monolayer A-GNR and they show that bilayer A-GNRs exhibit two distinct groups, metals and semiconductors, whereas monolayer A-GNRs display different electronic properties.

2.3. Multilayered Graphene Nanoribbons

Developing a fully general tight binding Hamiltonian from first principles, Finkenstadt et al. [22] studied the effect of the thickness (i.e., more than one layer) on the electronic structure of graphene nanoribbons. The authors found that layered nanoribbons of graphene approach the familiar band structure of graphite with increasing ribbon width, except for zigzag edge states near the Fermi level. In this context, K-T. Lam et al. [23] investigated armchair bilayer graphene nanoribbons in the DFT framework. Their results show that the band gap ($E_g$) of a bilayer A-GNR, in general, is smaller than that of a monolayer A-GNR and they show that bilayer A-GNRs exhibit two distinct groups, metals and semiconductors, whereas monolayer A-GNRs display different electronic properties.
purely semiconducting behavior. Moreover, the band gap $E_g$ of bilayer A-GNRs is highly sensitive to the interplanar distance. Inspired by recent experimental reports, indicating that joule heating can atomically sharpen the edges of CVD grown graphitic nanoribbons [24,25], Cruz-Silva et al. [26] using quantum molecular dynamics (MD) calculations, studied the role of defects and interstitial atoms in the coalescence of graphene edges. The authors addressed the issue of the effect of electron irradiation on graphene and explained why loops (or coalesced adjacent graphene edges) do not form when fast electrons first irradiate the sample before joule heating is introduced. These authors simulated the electron irradiation by putting several types of defects in a bilayer graphene nanoribbon and determined that both vacancies and interstitials are key for keeping graphene layers parallel and for avoiding bilayer edge coalescence (or loop formation).

2.4. Magnetism and Half-Metallicity on Zigzag Nanoribbons

Since the first studies by Nakada and Fujita et al. [10,14], it was realized that edge states in Z-GNRs could exhibit interesting magnetic properties. By studying the spin structure at edges, Lee et al. [27] found that magnetic order could be established at edges of graphitic fragments. These authors concluded that zigzag nanoribbons exhibit ferromagnetically ordered spins at each edge but with the opposite spin direction (FM-A). However, the spin configuration corresponding to the ferromagnetically ordered spins at each edge with the same spin direction (FM-F) is quasi degenerate with an FM-A spin configuration at the opposite edges. Figure 2 illustrate the spin-polarized density of states and the FM-A and FM-F spin configurations of zigzag carbon nanoribbons. When the width of the nanoribbon is increased, the FM-A and FM-F spin configurations become energetically degenerate. Wimmer et al. found that an ideal GNR (with symmetric edges) has zero spin conductance but has a nonzero spin Hall conductance. Moreover, only GNRs with imperfect edges (asymmetrically shaped edges) exhibit a nonzero spin conductance [28]. It has been demonstrated that applying external electric fields across the zigzag nanoribbons results in the appearance of half-metallicity [29]. In this case, states with only one type of spin (up or down) are present at the Fermi level, and such a handle to control spin ordering could be important for the design and fabrication of spin-valve devices. T. Nikolaos et al. [30] reported the observation of spin transport on a micrometer-scale distance in a single graphene layer. They measured the spin transport in a four-terminal spin valve device, finding that the spin transport is relatively insensitive to the temperature and that the observed spin relaxation length, and relaxation times in their device, are limited by intrinsic impurity scattering and not by the intrinsic properties of graphene.
3. DEFECTS IN GRAPHENE AND GRAPHENE NANORIBBONS

3.1. Classification of Defects

In general, defects within graphene-like structures can now be classified, isolated, and investigated in detail. Defects play a crucial role in the properties of crystals and diverse nanostructures, and the sp² carbon nanostructures are no exception. In 1996 a systematic compilation of defects was made by Terrones and Terrones [31]. Graphene-like systems are so versatile that they can accommodate different kinds of defects, thereby altering the basic structure of the host material and its physico-chemical properties. In particular, defects could change the atomic arrangement of carbon atoms in different ways: preserving the topology (not changing the total curvature), introducing curvature, altering the detailed sp² carbon connectivity, and accepting foreign atoms into the lattice (doping). According to this classification, defects can be categorized into five groups: (1) structural defects [32–35], (2) pentagon–heptagon pairs and bond rotations generating grain boundaries or extended lines [33–40], (3) doping-induced defects [41–44], (4) non-sp²-carbon defects [45,46], and (5) high-strain folding of graphene sheets [47–51]. It is still a challenge to study defects at the atomic scale, control their location, and pinpoint their particular properties in order to tailor the electronic properties of graphene-like systems (see Fig. 3). Recent experimental evidence, at the atomic scale, of some types of defects in graphene obtained by different research groups [52–55] is depicted in Fig. 4.

1. **Structural defects** are related to imperfections that distort the hexagonal carbon lattice by introducing curvature, but preserving the sp² connectivity of the carbon atoms (carbon atoms connected to three neighbors). These defects are caused by the presence of non-hexagonal rings (e.g., pentagons, heptagons, or octagons) surrounded by hexagonal rings. For example, if a single pentagon or a few pentagons are embedded into the graphene lattice, nanocones with different apex angles are obtained [56–60]. A 30° angle in a single-walled carbon nanotube (SWCNT) could also be explained by the presence of a pentagon on one side of the tube and a heptagon on the opposite side [61]. The reactivity of pentagons, heptagons, or octagons with specific acceptor or donor molecules still has to be determined from theoretical and experimental standpoints. As will be explained in the following section, the presence of adjacent pentagons and heptagons in the same proportion could produce local alterations without changing the total curvature, since the positive curvature from the pentagons is canceled by the negative curvature of the heptagons, and the morphological perturbation of a 5–7 pair is only local (see Fig. 3).

2. **Pentagon–Heptagon pairs, bond rotations (grain boundaries and extended lines of defects)**. The presence of pentagon–heptagon pairs in graphene distorts the lattice locally, not changing the total curvature of the sheet. In particular, these defects could be 5-7-7-5 pairs embedded in the hexagonal network (also known as Stone–Thrower–Wales (STW-type) defects [35,36]) that could be created by rotating a carbon–carbon bond by 90° within four neighboring hexagons, thus resulting in the generation of two pentagons and two heptagons while preserving sp² connectivity [61,62]. 2-D planar graphene-like systems containing pentagons, hexagons, and heptagons, called pentaheptite or Haeckelites, have been proposed and found to be metallic in theoretical studies [63,64]. Nanoribbons constructed from Haeckelites could be considered as a new hypothetical nano-architecture with fascinating properties that could be applied in electronics. Isolated pentagon–heptagon pairs could also be introduced to form a grain boundary or an extended line of defects in graphene [37–40,65] or in a graphitic nanoribbon, thus changing their edge termination and electronic properties and forming...
a hybrid graphitic nanoribbon. These particular hybrid nanoribbons exhibit half-metallicity in the absence of an electric field, and could be used to transport electrons with one type of spin; such nanoribbons could be a step forward in designing new spintronic devices [37].

It is noteworthy that the electronic and chemical properties of these 5–7 or 5-7-7-5 pairs are different from the structural defects (see above), and their reactivity and detection need to be investigated theoretically and experimentally as specific defect types. Zettl and co-workers were able to observe 5–7 and STW (Stone–Thrower–Wales) defects directly on isolated graphene surfaces using aberration-corrected transmission electron microscopy (see supplementary material of Reference [65]). Figure 4f shows such a defect.

3. Doping-induced defects, arising from substitutional non-carbon atoms, can also be embedded into a sp² carbon lattice. In this case, it has been demonstrated that nitrogen and boron atoms can be introduced into the hexagonal sp² hybridized carbon lattice. With both N and B dopants, the chemical reactivity of the graphene surface increases, in one case due to the fact that N has one electron more than C (making the material n-type), and in the other because B has one electron less than C and the material becomes p-type. Therefore, these types of defects could be used to tune the type of conduction in graphene-like materials, ranging from n-type transport (substitutional nitrogen doping) to p-type conduction (substitutional boron atoms in the lattice) [41]. Recent studies have demonstrated that other elements such as P, S, and Si and paired dopants such as P-N
could also be introduced in the hexagonal lattice of carbon nanotubes [42–44,66]. Therefore, the insertion of non-carbon atoms into graphene-like materials could tailor the chemical reactivity and electronic transport of the layers and further work along this line needs to be carried out [67].

4. **Non-sp² carbon defects** caused by the presence of highly reactive carbons, such as dangling bonds, carbon chains, interstitials (free atoms trapped between SWCNTs or between graphene sheets), edges (in open nanotubes or in graphene nanoribbons), adatoms, and vacancies. These defects are usually observed in an HRTEM, when the adsorbed atoms on these reactive sites are removed by means of the energy of the electron beam. It has been demonstrated that the creation of such defects could promote the formation of covalent nanotube junctions [45] and could also trigger the coalescence of nanotubes [46]. However, for nanoribbons, the dynamics of these defects has not been fully investigated either as a class nor have the differences between members of the class been investigated for nanoribbons. Further research is required to move forward in both of these directions.

5. **High-strain folding of graphene sheets** (loop formation) can be induced by the thermal annealing of two adjacent graphene layers. These types of “loops” have often been observed through TEM measurements after open edges are thermally annealed at constant temperatures above ∼1500 °C [47–50], but the chemical reactivity and electronic properties of the loops or the loop-formation process have not yet been investigated in detail. In fullerene (or carbon cage) chemistry, the main type of defect is *structural*, which deals with the presence of pentagonal rings that induce a highly curved fragment caused by the mixture of sp² and sp³ orbitals. In nanotube chemistry, these defects mainly occur on the caps (where pentagons are located). Narrow or highly curved nanotubes (usually <1 nm diameter) would also have more reactive surfaces. In highly crystalline SWCNT samples, the caps are a small portion of the structure, and other highly curved areas are not present since the nanotubes usually have diameters larger than 1 nm. Therefore, graphene and nanotube chemistry require a better understanding of the defects mentioned above, in order to induce and control strong (or weaker) covalent (or π−π) interactions of the host carbon atoms with other molecules, atoms, or clusters.

It should be noted that buckling and folding can be induced deliberately in graphene. Bao et al. have reported the controlled rippling of suspended graphene sheets [68]. Li et al. [69] have also reported the deliberate formation of ripples in graphene flakes deposited over polymethylmethacrylate (PMMA), by heating the flakes above the glass transition temperature of PMMA; upon rapid cooling, the polymer shrinks more than the graphene does, and the stresses lead to periodic sinusoidal (λ = 550 nm) buckles in the edges of graphene sheets. Raman spectroscopy shows a blue shift of all spectrum features consistent with a compressive stress on the graphene sheet. The authors also report that this process can lead to the folding of graphene flakes into few-layer graphene but this does not seem to be a controlled process [69]. Other processing approaches are likely to appear in the literature extending the possibility of controlling the folding and the buckling of GNRS. Multiply folded graphene called “grafold” has been recently identified by Kim et al. [70]. According to density functional theory calculations, the electronic properties change depending on the folding; for example, double folded graphene exhibits localized states in the folded region [70]. Control over these kinds of architectures is needed in order to use them for practical purposes.

**Figure 3** depicts the most common type of defects in carbon nanoribbons: (a) Vacancies originating from the removal of one or more carbon atoms from the hexagonal lattice. The presence of vacancies strongly modifies the electronic structure of the carbon system; for instance, localized electronic states are created around the vacancies making the unsaturated carbon atoms (surrounding the vacancy) chemically active [71,72]. (b) Heptagon–pentagon pairs (termed here as Stone–Thrower–Wales type transformations) [35,36]. (c) Antidots, which are “holes” in the graphene structure, shown in Fig. 3c, are defects that exhibit interesting electronic and magnetic properties [71–75]. Recent studies of the quantum conductance of armchair nanoribbons with different antidot morphologies demonstrated that the electronic band gap depends strongly on the period of the holes and the internal geometry of the hole structure [76]. Figure 3d depicts one-dimensional (1-D) defects composed of a chain of pentagonal and heptagonal rings, showing that defects can act as a grain boundary [37]. (d) Vacancies and bond rotations are shown in Figure 3e. (e) A hexagonal Haeckelite structure, which is made of multiple pentagonal, hexagonal, and heptagonal rings in seen in Figure 3f. Other types of defects (not shown) are loops (high strain folding of a graphene sheet) and interstitials. While heptagon–pentagon (7–5) pairs and loops preserve the three coordination connectivity of the nanoribbon, the interstitials and vacancies do not. Certainly, a challenge for the future will be the use of defects to design new graphitic nanoribbons with surface specificity for sensing different kinds of molecules or anchoring specific polymer chains in order to produce stable and well-dispersed composites.

### 3.2. Carbon Haeckelite Nanoribbons

Nanoribbons constructed by using a Haeckelite sheet could be energetically stable and energetically competitive with the graphene nanoribbons (see **Figure 5**). The Haeckelite
structures exhibit a combination of pentagonal, hexagonal, and heptagonal rings paired symmetrically. The entire 2D monolayer can be formed by the creation of Thower–Stone–Wales type defects on the graphene system, transforming pyrene-like rings into two pairs of heptagons and pentagons. In the past, two-dimensional Haeckelite structures consisting of pentagons and heptagons were already predicted by Crespi et al. [63]. These authors found that these structures are metallic and metastable, with a binding energy comparable to that of a $C_{60}$ molecule. On the other hand, using the tight-binding method, Terrones et al. [64] studied Haeckelite layers and nanotubes made from Haeckelite sheets by introducing hexagons with different periodic symmetries. The authors demonstrated that all Haeckelite sheets formed by incorporating pentagons, hexagons, and heptagons, in an ordered way, are energetically competitive and they are more energetically favorable than the $C_{60}$ fullerene molecule. They also found that all Haeckelite nanotubes exhibit a metallic behavior independent of their diameter and chirality. Density functional theory calculations performed by Mark. T. Lusk et al. [77,78] have demonstrated that a variety of complex, stable landscapes of defect domains can be monolithically nano-engineered from graphene to make arbitrary structures. Preliminary results are shown on the calculated spin-polarized electronic density of states (Figure 5a) and quantum conductance (Figure 5b) for a rectangular Haeckelite structure [79]. These results show that the Haeckelite ribbon exhibits a metallic behavior with a binding energy comparable to that of graphene nanoribbons. In addition, it has been observed that the spins located at the edges of these ribbons have an FM-A (ferromagnetic/anti-ferromagnetic) spin arrangement at opposite edges, as is depicted in Figure 5c. Note that only the doubly coordinated atoms located on the pentagonal rings exhibit a magnetic moment. However, further studies are needed in this direction in order to determine the role of the ribbon’s width, and its edge morphology on the electronic properties of Haeckelite nanoribbons [79].

4. SYNTHESIS METHODS OF GRAPHENE AND GRAPHITIC NANORIBBONS

In this section, various methods for synthesizing graphene and graphitic sp$^2$ nanoribbons are discussed.

4.1. Synthesis of Graphene and Carbon Nanoribbons by Chemical Vapor Deposition

The chemical vapor deposition (CVD) method has been extensively used for the synthesis of carbon nanotubes [8,43,45], and nowadays the CVD method has also become very useful for producing graphene and graphene nanoribbons. One of the first reports on graphene synthesis was presented by Murayama and Maeda [48] who used the CVD method. They used the disproportionation of carbon monoxide in the presence of catalytic metal particles. The resulting material consisted of ribbon-like filaments, 10 µm long, 0.1–0.7 µm wide, and 10–200 nm thick. Later in 2008, another CVD synthesis method of graphitic nanoribbons was reported by Campos-Delgado, et al. [80]. Here, a mixture of ferrocene, thiophene, and ethanol generated rippled carbon nanoribbons, several micrometers long, 20–300 nm wide and <15 nm thick. These ribbons (see Fig. 6a) were grown with the (002) planes parallel to the main ribbon axis.
Scanning transmission microscopy (SEM) images and high-resolution transmission electron microscopy (HRTEM) image are depicted in Fig. 6a–c. [81]. During the same year, Subramanyam, et al. reported that the pyrolysis of ferrocene and tetrahydrofuran resulted in crystalline carbon nanoribbons with an orientation of the (002) planes perpendicular to the growth axis. Templates have also been employed during the CVD process in order to obtain few-layered graphene ribbons [82]. It has been reported that the decomposition of CH\textsubscript{4} on ZnS ribbons (templates) followed by acid treatments in order to dissolve ZnS resulted in few-layer graphene nanoribbons (FLGNRs) of ∼3.4 nm in thickness, 0.5–5 μm in width, and several microns long. TEM images of the graphitic ribbons are depicted in Fig. 6e. The CVD method developed by Reina et al., [8] where they obtain large amount of CVD graphene, may perhaps be extended to nanoribbons.

4.2. Chemical Routes to Synthesize Graphene and Carbon Nanoribbons

Alternative chemical routes, consisting of the exfoliation of commercial graphite [12] or by organic chemistry approaches, have also been developed. Small length GNRs have been synthesized by linking tetra- and hexaphenylbenzenes through the Suzuki—Miyaura reaction [83]. The resulting ribbons exhibit lengths of 8–12 nm. It remains to be seen whether these or similar chemical reactions could create wider and/or longer GNRs or even branched GNRs. Hydrothermal processes have been used to obtain either amorphous or crystalline carbon ribbons involving Teflon-lined autoclaves, and TEM images of the graphitic ribbons are depicted in Fig. 6f, g [84,85]. Other synthesis approaches involve electrophoretic deposition on highly oriented pyrolytic graphite (HOPG) followed by a heat treatment in order to obtain graphene and nanoribbons [86]. Here diamond
nanoparticles are graphitized into layers forming \(sp^2\) carbon sheets and ribbons.

### 4.3. Synthesis of Graphene and Nanoribbons from Carbon Nanotubes as a Precursor

Attractive methods able to obtain graphene or nanoribbons from carbon nanotubes have been also developed. The first method to obtain graphitic nanoribbons from nanotubes was published in 2009 [87]. Subsequently, two additional papers related to the unzipping of nanotubes appeared two weeks later [88,89]. A fourth method, first suggested by Terrones, involved the use of metal catalytic particles as nanoscissors [90], and the method was soon successfully implemented, see Fig. 6j [91]. Finally, it was also demonstrated that nanotubes could be easily unzipped by passing high electrical current (Joule heating) inside a TEM microscope [92].

The first method relies on the intercalation of lithium followed by thermal exfoliation treatments [87]. Cano-Marquez reported that the resulting material ("ex-MWNT") contains up to 60% nanoribbon material, and the rest of the resulting material was found to be partly opened MWNTs (unzipped in parts), as well as damaged tubes, and graphene flakes that were peeled off the carbon tube walls. TEM images of the thermally exfoliated nanotubes are depicted in Fig. 6h, i [87]. It should be noted that partially unzipped ex-MWNT material could be of interest for composites (with the unwrapped edges interfacing to a matrix). Such a material could even have useful properties for constructing electronic nanodevices [93].

The James Tour group [89] has reported an improved unzipping method, which is highly efficient and results in high yields of nanoribbons (both nearing 100%), and a schematic representation of their process is shown in Figure 7c. The resulting oxidized nanoribbons can be reduced by hydrazine in order to remove the oxygen functional groups, thus restoring the good electrical conductivity of graphene to these nanoribbons. Other approaches involve the oxidation of MWNTs in air, followed by sonication [94]. This process yields large amounts of nanoribbon material (e.g. 60%) with thicknesses of 2 or 3 layers.

Another method to unzip carbon nanotubes makes use of catalytic cutting. This method is attractive due to its simplicity when compared to other chemical methods. Nevertheless, the actual amount of nanoribbons produced by this method is still very low (approximately 5% of the original CNT sample). Previous reports [95–97] have shown that certain catalytic metal particles (Co or Ni) could cut graphene layers only along armchair or zigzag atomic lines. In particular, for metal decorated MWCNTs, the unzipping mechanism consists of carbon atom dissociation on the catalyst (Ni or Co nanoparticles) and, since the process takes place in an Ar-H\(_2\) atmosphere, the dissociated carbon atoms further react with H\(_2\) to form methane (CH\(_4\)), and a SEM image of this process is depicted in Fig. 6j [98]. It also appears that at high temperatures (~1100 °C), the encapsulated Fe nanoparticles inside carbon nanotubes start to react with the surrounding carbon atoms, causing their dissociation and eventually resulting in the longitudinal breakage of the tube. These partially opened carbon nanotubes have been predicted to possess novel magneto-resistance properties [93] and further experimental work is required to obtain detailed information. A schematic representation of the method is shown in Fig. 7b.

An alternative technique for unzipping nanotubes consists of applying an electric current along the length of a nanotube inside a transmission electron microscope [92], and a representation of this process is depicted in Fig. 7e.
Unfortunately, this method is only able to yield a very small amount of unzipped carbon nanotubes. However, experiments involving the passage of high currents through bulk nanotube samples should be tested and further evaluated. Another physico-chemical method was also developed by Dai's group in 2009 [88]. They used an Ar-plasma etching method for synthesizing narrow graphene nanoribbons from nanotubes with a relatively narrow width distribution of 10–20 nm. Here, pristine MWCNTs are used as the starting precursor materials. Firstly the MWCNTs are deposited on a Si substrate and then coated with a poly(-methylmethacrylate) (PMMA) film. The PMMA-MWCNT film was peeled from the Si substrate, turned over and then exposed to an Ar plasma (an illustration of this method is shown in Figure 7d). The embedded part of the tube remains intact and by washing away the polymer, graphene nanoribbons are obtained. The method is much more laborious and the yield is relatively low, ca. 20%, with some MWCNTs remaining; but these nanoribbons appear to have a high degree of crystallinity. A summary of the methods to obtain graphitic nanoribbons from carbon nanotubes is given in Fig. 7.

Finally, the degree of crystallinity of multilayered nanoribbons could be improved by subsequent high temperature treatments (1500–2000 °C). However such heat treatments could also be promoting loop formation at the edges of adjacent graphene layers [49]. When using Joule heating and high electron irradiation, concurrently, Jia, et al. [24] observed that the most stable edges in graphene ribbons are either zigzag or armchair edges. However, loop formation between adjacent sheets is promoted only after thermal annealing treatments are made on graphitic nanostructures at temperatures >1500 °C in the absence of electron irradiation [49,50].

5. ROLE OF CHEMICAL DOPING IN GRAPHENE NANORIBBONS

The chemical doping of graphene and graphene nanoribbons is an interesting topic since the electronic and quantum transport properties could be significantly altered depending on the species of the dopant, the location of the dopants in the structure, and their concentration. Schematic representations of the substitutional and pyridine-like nitrogen doping in graphene nanoribbons and nanotubes are shown in Fig. 8a and b, respectively. Theoretical studies were also performed by Cervantes-Sodi, et al. [99] on graphene nanoribbons doped with N, B, and O atoms. Edge-type and substitutional doping could induce a half-metallic behavior and it is also clear that the band gap could be tuned by doping [100]. Roche and coworkers predicted the behavior of the electronic transport of nanoribbons doped with different concentration of dopants, and they observed that different materials, exhibiting different band gaps, could be obtained in this way [101]. Therefore, doped graphene nanoribbons appear to be excellent candidates for the design of controlled novel electronic devices and transistors, and further theoretical and experimental work is needed on this topic.

5.1. Quantum Conductance in Doped Graphene Nanoribbons

We have carried out electronic calculations using density functional theory [102,103] with local spin density approximation (DFT-LSDA) using the Ceperley–Alder parametrization [104] as implemented in the SIESTA code [105]. The wave functions for the valence electrons were represented by a linear combination of pseudo-atomic

![FIGURE 8 Representation of two possible ways of doping in graphene nanoribbons and nanotubes. (a) Substitutional doping; in this case one carbon atom is removed and replaced by a specific dopant atom. (b) Pyridine-like doping; in this case, one carbon is removed or one vacancy is introduced and the dopant atoms replace the double-coordinated atoms surrounding the vacancy. These two doping types are in plane (inside the structure lattice); however, the dopant atoms can also remain out of plane. For color version of this figure, the reader is referred to the online version of this book.](image)
numerical orbitals using a double-$\zeta$ polarized basis (DZP) \cite{106}, while core electrons were represented by norm-conserving Troullier–Martins pseudopotentials in the Kleinman–Bylander non-local form \cite{107,108}. The real-space grid used for charge and potential integration is equivalent to a planewave cut-off energy of 150 Ry. The pseudo-potentials (pp's) were constructed from the corresponding valence electrons for each element (H:1s\textsuperscript{1}, B:2s\textsuperscript{2}2p\textsuperscript{2}, C:2s\textsuperscript{2}2p\textsuperscript{2}, N: 2s\textsuperscript{2}2p\textsuperscript{3}, O:2s\textsuperscript{2}2p\textsuperscript{4}, Si:3s\textsuperscript{2}3p\textsuperscript{2}, P:3s\textsuperscript{2}3p\textsuperscript{3}, and S:3s\textsuperscript{2}3p\textsuperscript{4}). The systems were constructed by using a supercell of 100 atoms taking 7 unit cells of the 8 as armchair graphene (20 carbon atoms per unit cell). Periodic boundary conditions were used and the inter-graphene distance was kept to a minimum of ~10 Å to avoid lateral interactions. All systems were relaxed by conjugate gradient minimization until the maximum force was less than 0.04 eV/Å. The electronic transport properties were studied by non-equilibrium Green’s function (NEGF) techniques, within the Keldysh formalism, based on density functional theory (DFT) as implemented in the TranSISTA code.

Structural relaxation and transport calculations have been carried out using density functional theory on doped armchair nanoribbons \cite{109}. The nanoribbons were substitutionally doped by replacing one carbon atom by B, N, O, Si, P, or S atoms, and the edges were passivated with hydrogen. The optimized structures containing B, N, O, and Si are slightly modified and preserve the honeycomb lattice structure (see Fig. 9a, b). Only the bond lengths surrounding the dopants are slightly altered. However, the structures doped with P and S reveal that the structures are significantly modified; the dopant atoms in these cases prefer to remain out of the plane, and carbon and hydrogen atoms located near the edges also remain out of the plane, since these carbon atoms are pushed in the opposite direction of the dopant atom (see Figure 9b).

The calculated quantum transport properties of chemically doped armchair carbon nanoribbons are shown in Figures 10 and 11 for a variety of dopants. We have used the calculated quantum conductance for these doped systems at zero bias (see Figure 10) as a reference to study the effect of doping. For the sake of comparison, we have also added the conductance value of undoped nanoribbons (vertical dashed line). The quantum conductance plots reveal similar features at and around the Fermi level, and the doped nanoribbons exhibit a semiconductor behavior. Results for the current—voltage (I–V) curves are shown in Figure 11. For low doping concentrations (1 doping site per ~100 atoms), we observed a linear response for the B-, N-, Si-doped and undoped nanoribbons. We have also found that these doped systems exhibit a higher electronic current with a turn-on voltage of 0.2 V. However, the nanoribbons doped with O and P need a higher voltage to start
conducting and the resulting currents are always smaller than that for the remaining doping cases. The effect of doping for the armchair nanoribbons reveals a very sensitive quantum transport effect. Therefore, we envisage that ribbons doped with Si could be used as inter-connector devices, whereas nanoribbons doped with P could be used for the fabrication of sensors. However, further studies concerning the location and concentration of the dopant atoms are necessary to aid experimentalists in fabricating efficient electronic devices for specific applications.

5.2. Synthesis of Nitrogen-Doped Graphene Nanoribbons

Experimental evidence of N-doped graphene has been obtained via a chemical vapor deposition (CVD) process, in which both edge-type and substitutional N doping are produced [110]. These authors found that N-doped graphene-like material exhibited n-type semiconducting behavior, and consisted of 2–6 graphitic layers with a variable N concentration (in the range 1.2–8.9 at%). Subsequently, Dai’s group was successful in synthesizing graphene nanoribbons doped with nitrogen via joule heating in the presence of NH₃ [111]. This process generated both edge-type and substitutional N doping, in addition to the presence of NH₂ molecules bonded to the ribbon edges. It was also observed that doping with O atoms generated a p-type material caused by the presence of C=O bonds. Finally, undoped and N-doped graphitic nanoribbons have been produced by unzipping MWNTs and CNx-MWNTs via their catalytic hydrogenation process [91]. The authors deposited metal nanoparticles on the surface of the carbon nanotubes using chemical and physical methods. The resulting samples were exposed to an Ar-H atmosphere at different temperatures, promoting the formation of pure carbon nanoribbons, partially open tubes, and nanosheets. They claimed that during the experiment, carbon atoms are dissociated on the catalyst (Ni or Co nanoparticles) and, since the process takes place in an Ar-H atmosphere, the dissociated carbon atoms further react with H₂ to form methane (CH₄). Although the literature regarding doped graphene nanoribbons is very scarce, it is clear that doping appears to be a reliable way to control the electronic and chemical properties of the ribbons, and further investigations along these lines are needed. In addition, B doping of nanoribbons could be interesting since B atoms within the graphene nanoribbon lattice could introduce holes; thus, the system could exhibit p-type semiconducting behavior. To the best of our knowledge, there are no experimental reports on B doping of GNRs. However different theoretical studies have addressed this important issue [112,113].

6. EXPERIMENTAL DETECTION OF EDGE-STATES IN GRAPHENE NANORIBBONS

6.1. STM, STS, and NEXAFS Techniques for the Detection of Edge States in Graphene Nanoribbons

Different experimental investigations have detected the presence of edge-states and have revealed some of their properties using several techniques, such as scanning tunneling microscopy (STM), scanning tunneling spectroscopy (STS), near edge X-ray absorption fine structure (NEXAFS), and electron spin resonance (ESR) [114–121] used STM and STS techniques to probe graphene edges, and edge-states were found by the Enoki group in zigzag ribbons (see Fig. 12a), whereas armchair ribbons exhibit
only a low density of edge states (see Fig. 12b). For investigation of the magnetic properties of carbon nanoribbons, the electron spin resonance (ESR) technique can offer a sensitive probe for detecting the localized spins of the edge-state [118], and previous reports have shown a spin degree of freedom for edge-states [119,120]. The electronic structure and spin magnetism have been reported using NEXAFS and ESR experiments for few-layer graphene nanoribbons produced by the chemical vapor deposition (CVD) method [121], confirming the existence of magnetic edge-states which are due to open nanographene edges. The disappearance of edge-state after annealing at higher temperatures is explained by the decrease in the population of open edges due to loop formation (see Fig. 12c).

6.2. Raman Characterization of Graphene and Graphitic Nanoribbons

Optical characterization studies of graphene started with the use of the optical microscope, and the resulting identification of graphene flakes was simple but very tedious. Quantifying the number of layers implied the use of techniques that required prior sample preparation, and these processes were time consuming (AFM and TEM). Cançado et al. in 2004 reported pioneering studies on the use of Raman spectroscopy in the determination of step edges in graphitic materials [122] and reported a method for distinguishing armchair from zigzag edges. Also, in 2006, Ferrari et al. reported the Raman spectra of single-, double-, and few-layer graphene produced by the Novoselov’s “scotch-tape” technique [123]. Their work provided a clear signature for each type of single-, double-, and few (<5)-layer graphene flake. Hence, Raman spectroscopy has become a very useful and powerful technique for characterizing graphene and GNRs and for distinguishing one edge structure from another.

Therefore, an understanding of the phonon dispersion of graphene is essential to interpret the Raman spectra of graphene. Since the unit cell of monolayer graphene contains two carbon atoms, A and B, there are six dispersive phonon bands (see Fig. 13a [124]), in which three are acoustic branches (A), and the other three are optic (O) phonon branches. For one acoustic branch (A) and one optic (O) phonon branch, the atomic vibrations are perpendicular to the graphene plane, and they correspond to the out-of-plane (o) phonon modes. For the other two acoustic and two optic phonon branches, the vibrations are in-plane (i). Traditionally, the directions of the vibrations are considered with respect to the direction of the nearest carbon—carbon atom vibrations and, therefore, the phonon modes are classified as longitudinal (L) or transverse (T), according to whether the vibrations are respectively parallel with or perpendicular to the A-B carbon—carbon directions. Therefore, along the high symmetry ΓM and ΓK directions in the Brillouin zone of graphene, the six phonon dispersion curves are assigned to LO, iTO, oTO, LA, iTA, and oTA phonon modes (see Figure 13a). For 3D graphite, a weak phonon dispersion also exists along the c direction perpendicular to the AB plane.
The most prominent features in the Raman spectra of sp² hybridized carbon materials are the G-band appearing at ca. 1580 cm⁻¹, and the G'-band at ca. 2700 cm⁻¹ for a laser excitation energy of 2.41 eV (see a typical Raman spectrum depicted in Figure 13b). When the carbon samples have a certain amount of disorder within the structure, a disorder-induced band appears, the so-called D-band, at ca. 1350 cm⁻¹ and the edge terminations also show vibrations at the D-band frequency. The G-band (see Fig. 13) is associated with the doubly degenerate (iTO and LO) phonon modes (E₂g symmetry) at the Brillouin zone center (Gamma point). On the other hand, the G₀- and the D-bands originate from second-order processes, involving two iTO phonons near the K point for the G₀-band, or one iTO phonon and one defect-induced scattering process for the D-band [125]. What is remarkable about this spectrum is the large intensity of the G₀-band relative to the G-band, and the Raman spectrum for monolayer graphene is a unique spectrum for all sp² carbons (see Figure 13b). Furthermore, zigzag edges do not contribute (because of selection rules) to a Raman signal while armchair edges do, thereby providing a convenient method for distinguishing between these two types of edges [125].

Raman spectroscopy is a useful technique to determine the number of layer (<5) in graphene samples. In Figure 14a the G’/Raman bands are depicted, as measured with a laser energy of 2.41 eV, for five graphene samples, (i) monolayer, (ii) bilayer, (iii) tri-layer, (iv) four-layer graphene, and (v) highly oriented pyrolytic graphite (HOPG), showing the evolution of the Raman spectra with the number of layers [128]. For a monolayer graphene sample, the G’ band at room temperature exhibits a single Lorentzian feature (Figure 14a(i)) with a full width at half maximum (FWHM) of ~24 cm⁻¹.

In the case of 2-layer graphene with a Bernal AB layer stacking, there are four different processes for an electron scattered from (to) each conduction band. These four different scattering processes give rise to four Raman peaks in the G’ spectrum. Figure 14a-ii shows the Raman spectra of a 2-layer graphene sample, and here the spectra were fitted with four Lorentzians [126], each with a FWHM of ~24 cm⁻¹.

For 3-layer graphene, we have an unit cell with six carbon atoms, and as a result the number of allowed Raman peaks in the G’ band is fifteen [125], but the energy separations of many of these fifteen different processes turn out to be very close to each other in energy from an experimental standpoint. Experimentally it is found that the minimum number of peaks with an FWHM of ~24 cm⁻¹ necessary to correctly fit the G’ lineshape in this case are six. The Raman spectrum is shown in Figure 14a-iii, and the six characteristic Lorentzians peaks for a tri-layered graphene are also shown. An important point here is the fact that the same FWHM of ~24 cm⁻¹ is always used for the G’ band which comes from the monolayer graphene feature.

As we can observe, the G’/band continues to evolve with the number of graphene layers until we reach the final material that we must consider which is bulk graphite. This limit is here represented by HOPG (highly oriented pyrolytic graphite). For HOPG, we can describe the G’ band with two peaks (see Figure 14a-v). The turning point of the Raman spectra, distinguishing the G’ band for HOPG from that for few-layer graphene, starts to happen at 4-layer graphene (Figure 14a-iv), where the experimental analysis is here given in terms of 3 peaks. It is possible to see that now the high-frequency side of the G’ band has a larger intensity than that for fewer graphene layers, indicating the passage to HOPG, which has its higher intensity peak at higher frequencies, as contributions away from the K point become more important.
Recent Raman experiments on disordered graphite films heat-treated at different temperatures showed that the G' Raman band can be used to infer information about the out-of-plane lattice parameter (c) and the interplanar crystalline coherence length (Lc). Figure 14b shows the Raman spectra in the G' spectral band region, for four different disordered graphite samples treated at different temperatures. The sample heat-treated at 2200 °C shows a typical spectrum of a turbostratic graphite (turbostratic graphite does not exhibit any long-range stacking order), consisting of only one Lorentzian component (G'_{2D}), emphasizing planar behavior. By increasing the temperature up to 2700 °C, the Raman spectrum for this sample exhibits the characteristic G' bands of graphite, with two Lorentzians (G'_{3DA} and G'_{3DB}), suggesting an improved long-range stacking order starting at this heat treatment temperature. The two samples with intermediate heat treatment temperatures in Figure 14b, 2300 and 2500 °C, show the coexistence of G'_{2D}, G'_{3DA} and G'_{3DB} Raman bands, which describe the evolution from a G' Raman band characteristic of turbostratic (disordered) graphite (2200 °C) to a G' band characteristic of HOPG (2700 °C). By using this information, Cançado et al. [127] derived an empirical formula to calculate the out-of-plane lattice parameter (c) and c-axis crystallite size Lc for laser excitation energies in the visible range (for further information see [128]).

Also, in graphene systems where stacking disorder can occur, i.e., for epitaxial or CVD-grown graphene, this approach can be used in the future to determine the interplanar distance (c/2) and the interplanar crystalline thickness Lc. Moreover, careful inspection of the G' band lineshape for samples with c-axis stacking disorder can be useful to determine the contributions to this spectral feature from the AB and non-AB stacking order regions of the sample.

Raman spectroscopy also has been a useful technique for monitoring the development and motion of edges in graphitic nanoribbons [80]. Figure 15a depicts an SEM image of as-prepared graphitic nanoribbons with widths of 20–300 nm, and thicknesses of 2–40 graphene layers. Figure 15b depicts an HRTEM image showing the disordered edges of the graphitic material. These carbon nanoribbons were studied under heat treatment in a furnace at constant temperature or by Joule heating [24,25], and in both cases the results of the annealing of defects within the interior of the ribbons can be seen. Here adjacent edges tend to form loops to passivate their reactive dangling bonds, as is shown in Figure 15c [49]. Raman spectra for the as-prepared material and for the material heat treated at various heat treatment temperatures are shown in Figure 15d, where a decrease in D-band intensity can be identified with a decrease in the in-plane disorder, but a residual D-band intensity remains at a high heat treatment temperature (2800 °C) and this effect is attributed to the
presence of edge loops [49]. This approach to the prep-
ration of graphene nanoribbons is at an early stage of
development and offers promise as a method to study the
development and motion of edges and their electronic and
transport properties. Raman spectroscopy is expected to
play a major role in guiding the future development of this
graphene nanoribbon material.

7. GRAPHENE APPLICATIONS

In this section, we give an overview of some of the most
important current and near future applications of few-layer
graphene and graphene nanoribbons. In conjunction with
fullerenes and carbon nanotubes, graphene and graphitic
nanoribbons could be the new building blocks of a new
carbon age with several important applications. This is just
the tip of the iceberg and there is no doubt that in the long
term these novel nanostructures will find more uses which
will be seen in our everyday life, in: electronics, smart
composite materials, health, energy, etc.

7.1. Graphene Films as Transparent Electrodes

There is a significant worldwide interest in the development
of high quality and large-area graphene because of its
scientific and industrial importance. Especially, graphene is
a promising material for replacing indium tin oxide in
transparent conductive flat panel displays. Optoelectronic
devices, such as displays, touch screens, light-emitting
diodes, and solar cells require materials with low sheet
resistance and high optical transparency, should benefit
from graphene. Graphene, as a two-dimensional single
atomic layer of crystalline carbon, is a zero-band gap
semiconductor that exhibits a unique Dirac electronic band
structure, high room-temperature carrier mobility [129],
a tunable band gap [17], ballistic transport [130], and
a visible transparency of 97.7% [131,132]. These charac-
teristics make graphene an ideal material for applications in
electronics and optoelectronics, including its use as a tran-
sistor, photodetector, and saturable absorber of light [133].

Graphene thin films exhibit a high optical transparency
in both the visible and near-infrared regions of the spectrum,
a low electrical resistivity, and high chemical and thermal
stabilities making graphene an excellent choice for trans-
parent electrodes in various optoelectronic devices [134]. As
an example, Blake et al. [135] fabricated a liquid crystal
display (LCD) by using a sheet of peeled graphene as the
transparent conductor. They determined that each layer of
graphene absorbs about 2% of the incident light intensity,
which is significantly lower than that of conventionally used
ITO (15–18%) [136]; this characteristic makes graphene an
excellent material for transparent electrode applications.
Specially, graphene has shown great potential for creating photovoltaic solar devices owing to its high optical transmittance, electrical conductivity, and surface area. The graphene film serves not only as a transparent electrode for light transmission, but also as an active layer for electron/hole separation and hole transport [137,138]. Graphene microsheets have been dispersed into conjugated polymers to improve exciton dissociation and charge transport [139–141]. Solution-processed thin films were used as conductive and transparent electrodes for organic [142] and dye-sensitized [134] solar cells, although the cell efficiency is still lower than those with ITO and fluorine tin oxide (FTO) electrodes [134]. Further work is needed to realize the full potential of graphene for these applications.

The formation of Schottky junctions between graphene sheets and Si has been shown to provide an efficient solar cell [143]. A solar cell fabricated by Li, et al. is shown in Figure 16. A schematic representation and a photograph of a graphene sheet (GS) used in a GS/n-Si Schottky cell with a 0.1-cm² junction area are depicted in Fig. 16a [143]. The graphene film was obtained by a CVD process, and a TEM image of this graphene is shown in Figure 16c. After deposition on a Si substrate, they determined that the surface of the graphene film was smooth but with many wrinkles, and these wrinkles covered the entire substrate. In this case, surface coverage by graphene is 100%, compared with porous carbon nanotube films. This film consisted of multiple layers of graphene that are overlapped and interconnected, which ensures a conducting pathway even if there are cracks formed in one of the layers. The multilayer structure is expected to provide higher carrier mobility based on a recent study on graphene layers decoupled from bulk graphite [144]. The photovoltaic properties of the graphene/n-Si solar cells show that the photo-generated carriers are separated by the built-in field, and holes are diffused to the junction where they are swept to the graphene side. The two-dimensional, highly conductive graphene will reduce or eliminate the lateral potential drop along a graphene sheet, thereby enabling uniform carrier separation and collection. Light current-density vs voltage (J–V) data for two different cells, and with an area of 0.1 cm² and 0.5 cm², are shown in Fig. 16c [143]. These two cells exhibit an overall solar energy conversion efficiency of 1.65% and 1.34%, respectively.

Graphene thin films also exhibit great mechanical properties. For this reason, the preparation of flexible electrodes has been seriously studied. Kim, et al. [145] fabricated graphene films on a polyethylene terephthalate (PET) substrate coated with a polydimethylsiloxane (PDMS) layer. These authors found good optical and
electrical properties (electrical resistance = 280 Ω/sq and 80% of optical transparency), and the graphene film exhibited excellent mechanical properties when used to make flexible and stretchable electrodes.

Bae et al. [146] recently demonstrated large-scale films synthesized by a roll-to-roll production method and used wet-chemical doping of predominantly monolayer 30-inch graphene films grown by chemical vapor deposition onto flexible copper substrates. The graphene film prepared by the CVD process is transferred to a polyethylene terephthalate (PET) substrate. The films exhibit sheet resistances as low as ~125 Ω/sq with 97.4% optical transmittance. However, further layer-by-layer stacking to fabricate a doped four-layer film made it possible to obtain sheet resistance values as low as ~30 Ω/sq at ~90% transparency, which is superior to commercial transparent electrodes, such as are made from indium tin oxides [146]. They also demonstrate the incorporation of graphene electrodes into a fully functional touch screen panel device capable of withstanding high strain. The electromechanical properties of this graphene/PET touch screen are compared to those from indium tin oxide (ITO); see Figure 17c.

Unlike an ITO-based touch panel that easily breaks under 1–2% strain, the graphene-based panel stands up to 5% strain, which is limited not by graphene itself, but rather by the printed silver electrodes [147].

The optical transparency, electrical conductivity, and mechanical properties of thus prepared graphene films strongly depend on the synthesis process used to obtain the raw graphene material. Films made from exfoliated graphite exhibit an electrical resistance of 8 KΩ/sq and an optical transparency of 83% [148]. However, in the case of reduced graphene oxide, values of 1 KΩ/sq and 80% of transparency have been reported [149]. However, a CVD graphene film has shown an electrical resistance of 700 Ω/sq and 80% optical transparency [8]. It may be that an important step in the process is also the film preparation method (spray coating, dip coating, inkjet printing, etc.). By modifying the electronic properties and improving the processing techniques, graphene films may be shown to have the potential for use as conductive, flexible electrodes, and thus as an alternative to indium tin oxide (ITO). However, in order to realize graphene-based transparent conductive films for ITO replacement practically, graphene films should exhibit

![Graphene/PET touch panel film](image-url)

**FIGURE 17** (a) An assembled graphene/PET touch panel film showing outstanding flexibility. (b) UV-Visible spectra of roll-to-roll layer-by-layer transferred graphene films on PET substrates. (c) Electromechanical properties of graphene-based touch screen devices compared with ITO/PET electrodes. The inset shows the resistance change by compressive and tensile strain applied to the upper and lower graphene/PET panels, respectively [146]. (d) A graphene-based touch screen panel connected to a computer with control software. The operation movie is provided as an online supporting material [146]. For color version of this figure, the reader is referred to the online version of this book.
optical transparency of 90% (λ = 550 nm) as well as a sheet resistance below 80 Ω/sq, and they should be sustainable to bending without degradation of performance.

7.2. Graphene-Based Field-Effect Transistors

Due to its outstanding properties, graphene has excellent opportunities for use in high-speed and high-frequency electronics. High-frequency transistors are devices that benefit from the exceptionally high electron mobility, electron–hole symmetry and quantum Hall effect, [3,150] and the strong suppression of weak localization effects [151,152]. Graphene nanoribbons with a width below 10 nm were produced via a chemical synthesis route. The energy gap from the ribbon with a width of 2 nm was estimated to be 0.4 eV and \( I_{on}/I_{off} \) ratios of up to \( 10^7 \) were obtained from graphene nanoribbon-based field-effect transistors (FETs) [148]. A top-gated graphene FET operating at high frequencies (gigahertz) was developed in 2009 by Lin et al. [153], and soon thereafter, transistors operating at 100 GHz were demonstrated in 2010 [154] and claims of 300 GHz operation were reported only a few months later by Liao et al [155]. These authors reported the fabrication of graphene transistors using a Co\(_2\)Si–Al\(_2\)O\(_3\) core–shell nanowire as the self-aligned top-gate, and a schematic representation is depicted in Fig. 18a, b [155]. To make the device, graphene flakes were first mechanically peeled onto a highly resistive silicon substrate. Figure 18c shows an SEM image of a self-aligned graphene transistor and an optical microscope image of the overall device layout (Figure 18c, inset) [155]. The cross-sectional SEM image of a typical device shows the platinum thin-film source and drain electrodes well separated from the nanowire, but the graphene below the nanowire gate is not clearly visible in Figure 18d. The unprecedented transistor performance achieved in such self-aligned devices should open up exciting opportunities in high-speed, high-frequency electronics.

7.3. Graphene Nanocomposites

Graphene has a 200 times greater breaking strength than steel [156], as well as high quality electrical and thermal properties, and now graphene is considered as an ideal reinforcing filler in structural and electrical nanocomposite applications. When graphenes are incorporated homogeneously, they showed the ability to improve the physical properties of polymers, metals, and ceramics at extremely small loading levels [157]. Graphene–silica composites were prepared via a spin coating method to generate transparent conductive films [158]. In addition, graphene-filled conductive polymers acted as a superior acceptor layer in organic photovoltaics [141], while TiO\(_2\)–graphene showed excellent photocatalytic ability [159]. The mechanical, electrical, and thermal properties of graphene-filled nanocomposites allow them to be utilized in...
automotive, aerospace, electronics, catalysis, and packaging applications.

7.4. Graphene and Graphene Composites for Sensor Devices

Graphene’s gas sensing properties have been investigated by different researchers. Schedin et al. [160] showed that mechanically exfoliated graphene flakes can detect a single molecule of NO2 gas. They fabricated micrometer-size sensors made from graphene, and such a device was capable of detecting individual events when a gas molecule attaches to or detaches from a graphene surface. The adsorbed molecules change the local carrier concentration in graphene by one electron, which leads to step-like changes in resistance. The high sensitivity that is achieved is due to the fact that graphene is an exceptionally low-noise material electronically, which makes it a promising candidate not only for chemical detectors but also for other applications where local probes sensitive to external charge, magnetic field, or mechanical strain are required. Graphene-based gas sensors allow ultimate molecular sensitivity since the adsorption of individual gas molecules could be detected. Large arrays of such sensors would increase the uptake area [161], allowing higher sensitivity for short-time exposures and the detection of active (toxic) gases in as low concentration as would be practically desirable.

Fowler et al. [162] reported the fabrication of graphene-based chemical sensors for NO2, NH3, and 2,4-dinitrotoluene. They deposited graphene using a spin-coating method to create single-layer films on inter-digitized electrode arrays. The time response was relatively slow at room temperature but this can be accelerated at elevated temperatures at the expense of sensitivity. Also Joshi, et al. [163] reported the gas sensing properties of graphene films and ribbons toward O2, CO, and NO2, and they determined that graphene films showed better sensing response than graphene ribbons for the tested gases. Pt/graphene/SiC devices proved to be an excellent sensor configuration for the detection of hydrogen gas. These experiments also show that hydrogen gas is readily desorbed from the sensor (after exposure), which results in a remarkable repeatability [164].

Chemically modified graphene has been used in biosensor applications; for example, immunoglobulin E (IgE) aptamers with an approximate height of 3 nm were successfully immobilized on a graphene surface. The aptamer-modified graphene field-effect (G-FET) transistor showed selective electrical detection of IgE protein, exhibiting good affinity and the potential for G-FETs to be used in biological sensors [165]. Hou et al. [166] found that graphene modified with ethylenediamine triacetic acid (EDTA) can be used as an ideal electrode material to fabricate biosensors [166]. They fabricated a dopamine electrochemical biosensor with a high selectivity and sensitivity, which might allow its potential use in the investigation and diagnosis of dopamine-related diseases.

7.5. Graphene in Energy Storage Devices

Graphene has been examined as an electrode material for lithium ion batteries and electric double layer capacitors, because of both its high specific surface area (2630 m2/g) and its unique two-dimensional morphology. As shown in Figure 19, the discharge capacity of graphene in a lithium ion battery was found to be 540 mAh/g, and this could be enhanced up to 730 and 784 mAh/g by adding carbon nanotubes and fullerenes to graphene [167]. Graphene also showed a good ability to coat silicone particles, and metal or metal oxide particles in order to prevent the volume

![Figure 19](https://example.com/figure19.png)

**FIGURE 19** Lithium insertion/extraction properties of the various graphene families. (A) Charge/discharge profiles of (a) graphite, (b) graphene, (c) graphene + nanotubes, and (d) graphene + fullerenes at a current density of 0.05 A/g. (B) Charge/discharge cycle performance of (a) graphite, (b) graphene, (c) graphene + nanotubes, and (d) graphene + fullerenes [167]. For color version of this figure, the reader is referred to the online version of this book.
expansion of electrode materials during the charge and discharge cycles in batteries [168–170]. The supercapacitor is recognized as one of the important next-generation rechargeable power sources due to its intrinsic ability to decrease the gap between the dielectric capacitor and the battery. In addition, the edge site is known to have a differential capacitance 10 times higher than that of the basal plane [170,171]. Thus, graphene has been examined as a potential electrode material for supercapacitors. As shown in Figure 20 a graphene-based supercapacitor exhibited specific capacitances of 135 F/g and 99 F/g with aqueous and organic electrolytes, respectively [172]. Simultaneously, in order to improve the performance of these graphene-based supercapacitors, several approaches have been examined by controlling their surface area and edge sites [173,174] and by introducing metal oxide with graphene [175]. Moreover, graphene showed a high ability to support metal nanoparticles (Au, Pt, and Pd) and the graphene-Pt complex exhibited a good catalyzing capability in methanol fuel cells [176].

7.6. Graphene as a Two-Dimensional Catalyst Support

It has been recognized that materials such as graphene oxide possess interesting and potentially useful reactivity, such as their capacity to function as oxidants and/or as acids [177]. Bielawski and co-workers revealed the potential of harnessing the reactivity of graphene oxide (GO) for various synthetic reactions [177]. The authors demonstrated the efficient oxidation of benzyl alcohol to benzaldehyde in the presence of GO as a heterogeneous catalyst. Another successful oxidation reaction that was shown was for cis-stilbene to benzenal which indicates that GO may find a role in the Wacker process (this process was originally referred to the oxidation of ethylene to acetaldehyde by oxygen in water in the presence of a palladium tetrachloride catalyst), which currently requires the use of metal-based catalysts [178]. The use of GO for “carbocatalysis” is a truly novel application of graphene-based nanomaterials and opens up a host of possibilities for chemical synthesis.

The catalytic application of graphene has focused primarily on the use of graphene as a support for catalytically active transition metals due its high superficial area. Mühlaupt and co-workers demonstrated that palladium nanoparticles dispersed on graphite oxide were able to catalyze Suzuki–Miyaura coupling reactions [179] (The Suzuki–Miyaura reaction is a valuable synthetic process for the construction of carbon–carbon bonds). Although the catalytic activity of this material was high, the supported metal was the active catalyst, not the carbon. Seger et al. [180] used graphene as a support material for the dispersion of Pt nanoparticles for the development of
a fuel cell electrocatalyst. In addition, graphene–TiO₂ nanocomposites can be employed as photo-anodes in a photo-electrochemical cell. In this context, the role of graphene is mainly to promote the collection and transport of photo-injected electrons [181]. Efforts are also being made to utilize graphene-based composites for visible-light-induced photo-electrolysis of water [181].

8. FUTURE WORK

The research on graphene and graphitic nanoribbons carried out so far has opened up a new line of investigation, which might impact our everyday life with several significant applications. However, there are experimental and theoretical challenges that remain to be addressed, such as:

1. Controlling the nanoribbons chirality and width;
2. Developing scalable methods for large-scale production of nanoribbons;
3. Controlling doping procedures for different species;
4. Studying chemical reactivity;
5. Controlling defects in GNRs;
6. Using GNRs as building blocks for different carbon nanomaterials (interconnections);
7. Synthesizing hybrid and composite materials based on GNRs; and
8. Controlling the self assembly of functionalized graphene nanoribbons in liquid media.

ACKNOWLEDGMENTS

The authors are grateful to K. Gómez for technical assistance. This work was supported in part by CONACYT-México grants: 60218-F1 (FLU) and E.G.E. acknowledges support from CONACYT-México grants (186203) and the Artificial Leaf Project Umeå (K&A Wallenberg foundation). MT thanks JST-Japan for funding the Research Center for Exotic NanoCarbons, under the Japanese regional Innovation Strategy Program by the Excellence. H.T. acknowledges support as visiting professor at the CNMS at ORNL. MSD acknowledges support from Navy/ONR MURI contract #N00014-09-1-1063.

REFERENCES


Chapter | 2.2 Novel Carbon-Based Nanomaterials: Graphene and Graphitic Nanoribbons


Nanodiamond—An Emerging Nano-carbon Material

Eiji Osawa
NanoCarbon Research Institute, Ltd, AREC (Asama Research Extension Center), Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda, Nagano 386-8567, Japan

1. DIAMOND AS CERAMICS

Is diamond a ceramic? We have seen diamond mentioned as an example of elementary ceramic. Strictly speaking, diamond and ceramics are very different, but they are also strikingly similar in that both are strong but suffer from sudden breakage—ceramics by brittle fracture and diamond by cleavage. Material scientists would be interested in removing these defects while retaining original advantages. One type of artificial diamond that is the closest to ceramics with regard to structural features will be nano-sized poly-crystalline diamond (NPCD), being developed by Sumiya and Irifune [1]. NPCD was synthesized by subjecting nano-carbon particles (usually nano-graphite) to high-pressure high-temperature conditions deep into the diamond-stable region, and it displayed the highest reported Knoop hardness of 140 GPa and excellent performance in cutting and polishing. However, the processability of NPCD is, although improved considerably compared with microsized single-crystalline diamond (MSCD), still incomparable to good ceramics.

Let us assume that the intention of our editor to incorporate one chapter on nanodiamond in this book will be for us to introduce the ceramics research community the status of developments in nanodiamonds so that the experts can assess the possibilities of nanodiamonds as a class of ceramic. Incidentally, we have long been wishing to study nanodiamond-ceramics composites, and in this respect, we would also like to communicate with them. Let us begin by mentioning the reason why we are interested in nanodiamond-ceramics composites. We have been visited by many amateur diamond admirers telling their long dreams. The three most popular dreams are [1] to use colloidal solutions of nanodiamond as the heat exchange media, [2] to use nanodiamond particles as the finest polisher, and [3] to make a composite with epoxy resin or other polymers and mold it into a heat sink. None of these ever worked due to the misconception that layman scientists too often bear. Generally speaking, composites of plastic polymers with nanodiamond never worked, probably because the two components are too different materials. However, we came to the idea of nanodiamond composites with ceramics, glasses, and metals—thanks to these conversations.

Before further discussing the significance of nanodiamond-ceramics composites, we must mention which nanodiamond we are going to use. Among artificial nanodiamonds, there seems to be unanimous agreement among the researchers in this field that the dispersed primary particles of detonation nanodiamond are by far the most promising. Interests are quickly rising to this new and mass-producible nanocarbon. In agreement with this trend, two good reviews are written recently [2,3]. In analogy with the naming by Sumiya and Irifune, we will call this class of nanodiamond as NSCD.

There are quite attractive features in the NSCD-ceramics, NSCD-glass, and NSCD-metal composites projects. First, because of the number density effect [4] (see below), we need to add only very small amounts of NSCD, say 0.1 to 0.01 wt%, to ceramics and other matrices. We
expect strengths in such composites to increase considerably if the Hall—Petch theory works with nanoparticles [5]. As will be shown later [2.3.4(6)], we have at least one piece of the preliminary evidence, which supports the nano-extension of Hall-Petch theory. The improved strength properties of composites cannot be as remarkable as that of pristine diamond, but we would like to remark that for many purposes we do not need the super hardness of diamond. In addition, the nano-Hall—Petch effect should come with increased resistance to brittle fracture. Finally, the processability of matrix will not be affected significantly by dispersing only small amounts of NSCD.

For very large or small numbers, exponential expression is the most convenient to write but often unimpressive and leads to underestimation. For convenience, exponential expression and conventional units is summarized in Table 2.

The NSCD-ceramics-composites project is pending in our laboratories because technical requirements for the composites have not yet been met: NSCD must be well dispersed in the solid matrix, processing temperature must be lower than about ca. 700 °C in order to avoid phase transition of nanodiamond into nano-graphite, and there should be no carbide formation between the components, to cite a few. None of these requirements can be fulfilled by our present knowledge on NSCD. Clearly, we must wait a little longer in the composite projects. Under such circumstances, the best thing we could do would be to briefly summarize the status quo in the development of NSCD.

### 2. DISCOVERY OF NANO SINGLE-CRYSTALLINE DIAMOND

NSCD was discovered in 1963 by Danilenko, Volkov, and Elin, physicists working at that time in Zababakhin Institute of Technical Physics, Russian Federal Nuclear Center, Snezhinsk, Chelyabinsk oblast, Russia, while detonating a military explosive called Composition B in water [7]. Unfortunately, this great discovery remained in anonymity for over 40 years for two reasons. One is that the Soviet military regime immediately classified the nanodiamond as one of the top secrets and never published any record. The second reason is that it was a premature discovery: there was no recognition of active surface in the single-nano particles nor was there any method known to evaluate dispersion of this size of nanoparticles in those days. We were simply lucky to get into this field after the science had acquired just enough experience on nanoparticles. We noticed in about 2001 that the imported detonation nanodiamond largely consisted of abnormally tight agglomerates. Being unable to identify the interparticle force acting

### TABLE 1 Particle Density and Other Basic Numbers Estimated for NSCD Particles

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average diameter $D (=2r)$, nm</td>
<td>4.8</td>
</tr>
<tr>
<td>Average volume $v (=4\pi r^2/3)$, cm³</td>
<td>$5.44 \times 10^{-20}$</td>
</tr>
<tr>
<td>Average mass $w (=\rho v)$, g</td>
<td>$1.63 \times 10^{-19}$</td>
</tr>
<tr>
<td>Average particulate weight $w' (=w/N)$</td>
<td>97,900</td>
</tr>
<tr>
<td>Average number of carbon atoms in a particle</td>
<td>8160</td>
</tr>
<tr>
<td>Average number of particles in unit weight $w' (=1/w)$</td>
<td>$6.15 \times 10^{18}$</td>
</tr>
<tr>
<td>Number of particles in 1 μl of 0.01 wt% colloid</td>
<td>$6.15 \times 10^{11}$</td>
</tr>
</tbody>
</table>

*a* $\rho = 2.99$ (exp), cf. 3.52 (natural bulk diamond) g/cm³.

*b* Concept corresponding to molecular weight.

*N* = $6.023 \times 10^{23}$.

### TABLE 2 Units of Very Large (Upper Four Lines) and Very Small (Lower Three Lines) Numbers and Their Exponential Expressions

<table>
<thead>
<tr>
<th>Exponential expression</th>
<th>$10^6$</th>
<th>$10^9$</th>
<th>$10^{12}$</th>
<th>$10^{15}$</th>
<th>$10^{18}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>China/Japan</td>
<td>千</td>
<td>亿</td>
<td>兆</td>
<td>1000兆</td>
<td>京</td>
</tr>
<tr>
<td>English</td>
<td>Million</td>
<td>Billion</td>
<td>Trillion</td>
<td>Quadrillion</td>
<td>Quintillion</td>
</tr>
<tr>
<td>SI</td>
<td>Mega</td>
<td>Giga</td>
<td>Tera</td>
<td>Peta</td>
<td>Exa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exponential expression</th>
<th>$10^{-6}$</th>
<th>$10^{-9}$</th>
<th>$10^{-12}$</th>
<th>$10^{-15}$</th>
<th>$10^{-18}$</th>
<th>$10^{-21}$</th>
<th>$10^{-24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>China/Japan</td>
<td>微</td>
<td>靜</td>
<td>漢</td>
<td>須</td>
<td>則</td>
<td>空</td>
<td>阿</td>
</tr>
<tr>
<td>SI</td>
<td>Micro</td>
<td>Nano</td>
<td>Pico</td>
<td>Femto</td>
<td>At</td>
<td>Zepto</td>
<td>Yotko</td>
</tr>
</tbody>
</table>
to form such tight agglomeration, we named it agglutination. In 2002, we found that the agglutinates could be disintegrated by attrition milling with 30-μm zirconia beads in water to give a stable dispersion of primary particles of NSCD for the first time after its discovery (Photo 1a) [8].

This incidence also marked the first emergence of typical nano-carbon particles before scientists. Whereas C_{60} was known since 1985 [9], it can be sublimed, vapor deposited, and dissolved in organic solvents to give molecularly dispersed solution. In other words, C_{60} behaves as a large molecule with 0.7 nm in diameter and hardly fits the definition of nanoparticles (1 < d < 100 nm) unless van der Waals diameters of carbon are included. In contrast, NSCD had much larger size of 4.8 ± 0.7 nm, and hence it is undoubtedly an authentic nanoparticle but challenging in that at first, every behavior seemed new and mysterious. There is no way to purify [10].

3. CHARACTERISTIC PROPERTIES AND BEHAVIORS OF NSCD

3.1. Interparticle Interactions

The first problem that everyone engaged in the research of detonation nanodiamond had to face was to explain why the agglomeration was so tight. We could destroy the unknown bonding among NSCD particles by brute force but the nature of bonding was still unknown. In 2007—2008, Barnard published two seminal papers that gave clear answers not only about the nature of interparticle bonding but also on a few other consequences. [11,12]. She used Seifert’s popular program for nanoparticles, self-consistent charge density functional tight binding (SCC-DFTB), to treat such large particles as NSCD—supposed to contain nearly 9000 carbon atoms in reality. The largest models she used contained about 1700 carbon atoms—large enough to lend high reliability in her interpretation. She discovered facet-dependent electrostatic fields, both negative and positive sign, and varying over a wide range in the charge densities (Figure 1e). Strong attractive Coulombic interactions should take place between the oppositely charged and parallel-oriented facets of the nearest neighboring particles. By taking such a pair and adjusting two parameters, vertical interfacial distance and angle of relative rotation around the common central axis of the interacting facets, she could identify all the prominent energy minima from among possible combinations of attractive facet—facet interactions and configurations.

Electrostatic charges on the facets of NSCD particles have been interpreted as arising from free spins generated on the new diamond surface when the atoms in a facet are transformed into graphene atoms (known to occur spontaneously and reproduced in Barnard’s calculations [11]), which then migrated to the surface by orbital interactions [13]. Thus, an NSCD particle acquires rather complex structure, in both geometrical and electronic terms (Figure 1d and e). Especially noteworthy is its permanent multipolar nature, which is unprecedented in the history of diamond research. These results explain the unusual
behaviors of NSCD displayed in a number of ways. For example, after the agglutinates were dispersed into colloidal solution of primary particles in wet bead milling, evaporation of water gives back considerably strong conglomerate, but this dried powder can be redispersed back to colloid by applying sonication alone—namely the agglomerate from the once-milled NSCD is not an agglutinate.

Why is agglutination irreversible? This behavior can be interpreted as follows. Immediately after the growth step ended under high-pressure, high-temperature (high PT) conditions created by detonation, crystals have enough time to relax into optimum orientation in the facet—facet attractive interactions, leading to tight agglutinates. Re-agglomeration of the milled material occurs by virtue of strong interfacial Coulombic attractions, but the temperature (below 100 °C) is too low to allow for nanoparticles to reach optimum orientation. The interactions that have been envisaged in agglutinates under very high PT conditions are named coherent interfacial Coulombic interaction (CICI), whereas all the other non-optimum interactions are called incoherent interfacial Coulombic interaction (IICI). The importance of interfacial electrostatic interaction in the aggregation of NSCD has been thereafter experimentally supported [14–16].

### 3.2. Solvation

Water is by far the most convenient and powerful solvent to disperse NSCD particles into stable colloidal solution up to about 8%. Naturally, water-soluble diamond is a tremendous surprise! Vicinal diols like ethylene glycol and propylene glycol, CH₃C(H)OHCH₂OH, are good solvents, dissolving up to 3% of NSCD, compared with monohydric alcohols like methanol and ethanol, which give only turbid colloid less than 1% in concentration. The contrasting behavior between monohydric alcohols and vicinal diols is still not understood. Other potentially interesting solvent type is dipolar aprotic. Dimethyl sulfoxide or DMSO, (CH₃)₂S=O, is the best solvent among all solvents tested for NSCD, giving up to 10% colloid without gel formation. In contrast, N,N-dimethyl formamide or DMF, (CH₃)₂NC(H)==O (I), is known to be similar to DMSO as polymer solvent, but it has no affinity to NSCD at all. Slightly hindered end of the negative pole of carbonyl in DMF compared to the corresponding pole in DMSO could be the reason. However, the popular industrial dipolar aprotic solvent, N-methylpyrrolidone (II), has a negative pole in very similar steric environments as DMF but dissolves NSCD up to 1%.

If we set aside small irregularities in the dissolving power of some organic solvents as mentioned above, we
may certainly infer that the remarkable stability of NSCD colloid is caused by a tight solvation, with OH groups oriented to negatively charged {111} facets and S=O or C=O dipole to positively charged {100} facets. The former H-bonding has been computationally confirmed [17]. On the other hand, NSCD is totally inert toward a number of popular solvents like hydrocarbons, halocarbons, aliphatic ketones, and esters including oils, contrary to our initial expectation that inherently nonpolar diamond nanoparticles should have affinity toward nonpolar solvents.

These observations and the knowledge on the permanent electrostatic fields on the surface of NSCD particles prompted us to expect that the surface charges in NSCD are also responsible for a few other peculiar behaviors. For example, the following reactions may take place on the surface of NSCD: (a) solvent or ligand exchange, (b) charge-transfer bond formation, and (c) linear or cyclic alignments of particles. Reaction (a) will be useful for application to catalysts, reaction (b) for the formation of metal–carbon bonds, and reaction (c) for percolation structure [18]. Note that the remarkable intraparticle charge migration mentioned above occurred because the diamond–graphite phase transition on the surface of single-nano particles brought the most electronegative (fullerene shell) and electropositive (diamond) carbons in close vicinity [13].

### 3.3. Shape of Particles

Natural diamonds display rich surface morphology [19]. For bulk diamond, the ratio of surface atoms to interior atoms is insignificantly small that the surface does not exert much influence on the property of crystal, but the situation is different in single-nano diamond particles like NSCD, wherein the surface properties define the behavior of whole particle. For this reason, it is essential to study the surface morphology of NSCD to understand its behavior. Although we cannot clearly see the shape of single-nano diamond particles under TEM due to poor contrast of carbon under electron beam (Figure 1a), the missing information can be covered to a considerable extent by calculations using SCC-DFTB theory. Our best guess for the shape of NSCD is still truncated octahedron (Figure 1b–e) [20].

### 3.4. Surface Structure

The surface of NSCD is believed to be covered with holey and polyhedral giant fullerene shell, wherein the holes cover [100] facets with large positive electrostatic potential, while the remaining polyhedral faces are covered with nanographene patches and have large negative to neutral charges [11]. The importance of surface graphitic and disordered sp²+ diamond layers is still controversial. According to SCC-DFTB calculations, both these upper layers are thick, each occupying about one-third of the whole particle. Recent aberration-corrected TEM images reveal much thinner upper layers than those calculated for smaller models.

Numerous reports have appeared that dealt with the analysis of carboxyl groups by simply taking IR spectra and performing titration, but these techniques are insufficient to characterize the complex surface of NSCD. Carboxyl groups are the results of oxidation to remove soot from the reaction mixture and hence have not much to do with the inherent nature of NSCD. We rather suggest to reduce the surface of crude NSCD with hydrogen at 500 °C to obtain nanodiamond hydrocarbon, wherein the surface is covered only with C–H bonds, and all the graphitic patches and functional groups are removed [21]. Then, we should be able to characterize the surface morphology including edge, face, and apexes. After this, it will be easier to introduce functional groups and identify their number and position on the surface.

### 3.5. Detonation Process

A surprising fact is that only one brand of NSCD agglutinates produced in Gansu, China, can be disintegrated perfectly into its primary particles while keeping high zeta potential throughout the milling operation. None of the other brands of NSCD agglutinates could be disintegrated due to quick decrease in the absolute values of zeta potential and subsequent increase in viscosity. Full development of electrostatic charges on the surface of primary particles appears an essential prerequisite to disintegrate agglutinates into primary particles. Regrettably, the secrecy prevailing in the manufacturers of detonation nanodiamond interferes with further investigation into the reason for failure in milling disintegration of agglutinates from the other manufacturers.

### 3.6. Purification

Although there is no straightforward and general way of purifying any nanoparticle [10], the situation in detonation nanodiamond is being improved. It is well recognized that bead milling contaminates nanodiamond with nanoparticles of zirconia, which were inevitably formed by contact of bead with bare diamond particles. At the
moment, we are thinking of one method to overcome the contamination problem, which is to use microdiamond as the crushing media in bead milling. Microdiamonds are now available at the cost range of less than $1 per g, and self-etching of the crude polyhedral diamond in the mill should provide spherical microdiamond particles. More difficult problem will be to coat the inner wall of the mill and rotating blades with diamond film by CVD technique. Although single-crystalline diamond film is more desirable than polycrystalline film currently available on market, this would be somewhat a too distant reality. Instead, CVD coating with high-quality polycrystalline diamond became recently a more realistic goal [22].

3.7. Color

The reason for the black coloration of colloidal solution of NSCD produced by bead milling has remained a mystery ever since the first successful isolation of its primary particles by milling. Dark color begins to appear in the initial phase of bead milling, as seen by transition from light brown color of suspension of agglutinates in water to dark gray to thin black color. Black color is quickly intensified, with milling time, until really strong but clear black. When the milling is complete, the concentration of NSCD is about 5–7%. Upon dilution of the colloid to 1%, black color becomes dim, and it turns brown to light and pale brown upon further dilution with water. The colloid becomes almost colorless at less than 0.01%, but upon shining with laser light scattering is clearly observable in such high dilution [Photo 1b]. Positive Tyndall effect excludes the possibility of self-extinction by supermultiple scattering with large number of nanodiamond particles as the cause of black coloration. It seems likely that some new physical phenomenon is occurring in the concentrated colloidal solution of single-nano diamond particles, a situation that never existed before us. At the time of writing this article (July 2011), we were informed that this problem was solved by a student in St. Petersburg.

Only a few physical constants have been measured on NSCD in the present form (Table 3).

4. POTENTIAL APPLICATIONS OF NSCD

With so many unique and promising features of NSCD as presented above, one would surely expect a lot of promising applications for this novel form of diamond. At the moment, major efforts are being directed to those applications that make use of the smallness of individual NSCD particles. Some of them are briefly mentioned below:

4.1. Nanospacer Lubrication in Water

‘Necessity evil’ means those recognized to produce environmental, health, or other kinds of hazards but difficult to stop because they are indispensable, and replacements cannot be readily found. Typical examples include fluorescent lights and gasoline-powered automobiles. Atomic energy may be cited as the most serious example but more disputes seem to be needed. Lubricating oil is an undisputable threat to our environments but so far, we have never been successful to find replacement. The used lubricating oils contain metallic oxide powder and cannot be burnt to dispose, and hence the recovered waste oils that amount to estimated $10^{11}$ tons per year in the world are being buried underground [23].

The mechanism of lubrication was first questioned by Leonardo da Vinci in the fifteenth century. He conducted the first experiment on lubrication in record and concluded that friction is not related with the contact area between two planes in relative motion but to the force applied to get moving. “Da Vinci riddle” remained unsolved until the middle of twentieth century when Professor F. Bowden of the Cambridge University finally gave the correct answer. According to his now famous fusion theory, no planes are truly planar so that microscopic ups and downs existing even in polished surface are inevitable, and hence the first contact between a pair of planes made of, for example, metal, approaching toward each other, takes place between a couple of asperities that happen to meet, hence the true contact area is much smaller than the apparent contact area (Figure 2). Under such circumstances, the force applied to move the planes is concentrated to the small number of contacting asperities, causing elastic deformation, fusion, and even bond formation between the colliding tips. Friction is equal to the force necessary to break the bonds formed in the true contact points. Bowden confirmed by careful experiments that the true contact areas are generally in the order of $\mu m^2$ and that bonds are really formed between the asperities [24].

Bowden’s fusion theory of friction became a guiding principle thereafter. The most important consequence of the theory is that oil is disqualified as the lubricating material. According Bowden’s theory, the primary function of lubricant must be to physically forbid direct contact between the moving planes. However, an oil film under the boundary condition is readily broken and not at all qualified for this purpose. One solution of this problem will be to use solid lubricants having sliding layered structure like molybdenum sulfide and graphite. Actually solid lubricants are being used exclusively for lubrication in space like the construction of space stations, but the cost is high and frictional coefficients are generally unsatisfactory. Therefore, solid lubricants are unsuitable for general purpose lubrication. We must find practicable replacements for oil.

The most straightforward and effective method to perfectly forbid contact between the interacting metal surfaces would be to use spacers, which have been used since Egyptian culture and display extra advantage of
causing very low friction by rolling. A typical spacer lubrication system is ‘ball bearing’, wherein small steel balls are immersed in grease and fixed around the journal of rotating axes. However, spacer lubrication seemed more costly than solid lubricants and must have been so far considered unsuitable for the general purpose lubrication system. It occurred to us that a colloidal solution of NSCD particles might work for this purpose.

At first, aqueous colloidal solutions were used as lubricating liquid, into which a sheet of silicon wafer was immersed, and a sapphire ball was pressed onto the wafer surface and given reciprocal movements by applying small force in a friction tester to obtain frictional coefficients (Table 4). Aqueous solutions were used due to our interests in Deformation, fusion and bond formation.
into animal articular super-lubricity by a combination of cartilages with complex structure and body water. To our surprise, we obtained an unexpectedly low dynamic fricational coefficient of 0.007 in 1% NSCD colloidal solution! We are at the moment in the process of optimizing experimental conditions, searching the best solvent and evaluating feasibility of the colloidal solutions of NSCD for the general purpose lubrication. Below in this section, let us discuss more about the mechanism of nanospacer lubrication in general.

We need to consider the boundary conditions on nanoscale, wherein tips of asperities of surface are approaching to within nanoscopic distance. Naturally the very top of an asperity consists of just one or a few atoms (Figure 3). As we go down from the tip, the number of atoms will rapidly increase to hundreds and thousands. Let us imagine a couple of asperities coming toward each other, and suppose a boundary condition starts when the distance between asperity tips reached 100 nm. Then, in a moment, the tips will touch each other unless there are certain numbers of NSCD particles floating in this region. In that case, they will be pinched between tips at an interasperity distance of 8 nm (diameter of NSCD + 2 times van der Waals distance of carbon atom). Note that the inter-asperity distance does not necessarily mean linear van der Waals distance of carbon atom). The spacers receive very high force when pinched between the asperities in the areas of the orders of nm² at the extremely small points of contact. In order to survive under such conditions, diamond would be the first choice.

**Nanospacers must roll well.** According to aberration-corrected TEM images and calculations of sphericity, the best model of NSCD is a truncated octahedron. As bead milling must have worn out sharp apexes and edges, NSCD can be considered to be quasi-spherical in shape and exhibit small frictional coefficient of rolling.

<table>
<thead>
<tr>
<th>Lubricant</th>
<th>Viscosity, $\rho$, $20^\circ$C</th>
<th>NSCD, w/v%</th>
<th>Dynamic frictional coefficient$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>0.01</td>
<td>0</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.005 – 0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>0.16</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Dimethyl sulfoxide</td>
<td>0.011</td>
<td>0</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.023</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Dynamic frictional coefficients were determined by Prof. S. Mori, Department of Applied Chemistry, Faculty of Engineering, Iwate University, on a reciprocating frictional tester, type KE-851 manufactured by Keya Precision Co. Ltd. (Mitaka, Tokyo). A sheet of 1 x 1 cm Si wafer was immersed in the lubricating colloidal solution and a sapphire semispherical ball (diameter 1–2 mm) was linearly reciprocated on the wafer surface under a load of 10–40 mN over a distance of 1.5 mm at a frequency of 10 Hz. Frictional force was monitored by using a highly sensitive transducer.

**FIGURE 3** Contrasting performance of oil and nanospacer lubrication. **Left:** The essence of experimental results reported in Bowden-Tabor textbook (ref. [24]) is condensed into a cone-shaped asperity. When the measurement began, fusion between touching asperities had already proceeded to create an eroded surface area that corresponds to a circle with 92 µm in diameter, and this loss increased to lose surface area corresponding to a circle with 426 µm in diameter when the friction was finally overcome, and the pair of metallic planes resumed relative movements. Alternatively, the contacting asperities are lowered by 0.03 to 0.7 mm in one frictional experiment between the two metallic planes. **Right:** In contrast, nanospacer lubrication takes place in the very initial stages of friction. There should be no measurable loss in the mass of tips as the closing asperities will be instantly separated by the rolling action of a few NSCD particles. For color version of this figure, the reader is referred to the online version of this book.
3. **Nanospacers should not damage the metal surface.** Rolling nanospacers should be safe with regard to the scratching of metal surface. Actually no damage was observed on the surface of Si wafer in the preliminary tests (Table 4). Theoretically, as long as the diameter of spacers is smaller than the smoothness of the metal surface, $\sigma_{\text{as}}$, they will not be trapped between the surfaces under direct contacts to cause scars on the surface.

4. **Water will quickly evaporate by the heat from friction.** Heat evolution should be much lower in nanospacer lubrication compared to oil lubrication because of the low friction in the former. Transition of machine materials from iron/steel to aluminum and ceramics should reduce rusting. Water may be replaced by alcohols, if higher viscosity is required for more efficient trapping of nanospacers at the contact points. Fortunately, NSCD particles dissolve well in ethylene and propylene glycol up to 2–3%. These popular glycols are nontoxic, cheap, high-boiling (about 190 °C), water-soluble, and low-freezing (−18 and −59 °C, respectively).

5. **What about the cost of NSCD?** The expired military detonators like Composition B have negative price.

The reason why Bowden could not find replacements of lubricating oil based on his own theory is that nanotechnology was unknown in midtwentieth century. Perhaps it is time for us to finish the long pending homework of our prescient pioneers (including da Vinci).

### 4.2. Fluorescent Bioprobes

Not all the gem diamonds are colorless and transparent, but a few of them are in many colors like red, pink, blue, canary yellow, and even black. These colors are known to arise from elements or their ions other than carbon trapped in the lattice defects of diamond crystal and absorbing visible light. Thus, the element-containing defects are called color centers. Since about 20 years ago [25], a pair of neighboring lattice defects trapping nitrogen molecule or ions in one of them are found to emit fluorescence, which are strong, stable, but not fading, nor blinking. These are called nitrogen-vacancy color centers or simply N–V color centers. Coupled with other advantages like high biocompatibility of diamond and the well-developed single-photon fluorescence spectroscopy, cell-imaging technique with fluorescent nanodiamond is advancing fast.

Recently Wrachtrup and his collaborators combined fluorescence microscope with magnetic resonance spectrometer for detecting the effect of magnetic field upon the spin state in a N–V center in diamond particle located in the vicinity of magnetic field, to give a sort of ultrasensitive scanning probe magnetometer [26]. They succeeded in observing 5 mT magnetic resonance line from a microparticle of nickel (Ni) with a space resolution of 8 nm in four minutes at room temperature by placing one particle of fluorescent nanodiamond within 1 µm of the Ni particle. A key to the success is to have combined atomistic spatial resolution of magnetic resonance imaging with fluorescence microscope, capable of catching and analyzing the luminescence from a single phonon. This observation logically prompted construction of an optically detected magnetic resonance (ODMR) microscope by a manufacturer of magnetic resonance spectrometers. In the near future, it will be possible to study dynamics of single molecule in a living cell in vivo at room temperature over a wide span of time as the design of ODMR microscope is advanced.

It is hoped that fluorescent nanodiamonds also develop, keeping pace with the ODMR microscope, but the progress seems slow at the moment. Preparative steps can be divided into two stages: growing nanodiamond crystals in desired sizes and generating vacant lattice defects followed by producing N–V centers by annealing.

1. **Growing nanodiamond crystals.** The probability of finding N–V center in natural diamond is quite low: one center for every ca. 300 nm³ volume of diamond crystal. This means that we need nanodiamond particles having at least 30 nm in diameter if the N–V concentration in nature is taken as standard. At the moment, the smallest fractions produced as by-products of the size-separation process of micro-diamonds are used, but more stable source of supply, preferably spherical particles, is desirable. At the moment, NSCD is the only source of nanodiamond particles outside the microdiamond industry, and hence the size is limited to only ca. 5 nm. Clearly, we need a lot of work to find out the proper conditions for creating proper concentrations of N–V centers.

For use in molecular biology and biomedical applications, those having simpler and more controllable surface structure than that of detonation nanodiamond are desirable. One realistic candidate is the hydrogenated NSCD, recently prepared by Williams [21] from commercial agglutinates of detonation nanodiamond under the flow of hydrogen gas at 500 °C for 4 h. During this clean reaction, the surface layers of graphitic carbon and oxygen-containing functional groups are reduced, vaporized, and removed, leaving the diamond core covered with C–H bonds. An additional merit of the hot hydrogen treatment is that the interparticle bonds by CICI and IICI are also cleaved. The simplest way of purifying the crude product from detonation of Composition B to obtain hydrogenated NSCD in one step would be to subject it to hot hydrogen treatment.

Here again many questions arise as to the properties of hydrogenated NSCD (HNSCD). Is HNSCD soluble in water? How about the possibility of spontaneous...
Graphitization? Is HNSCD stable enough to survive the photoluminescence experiments?

(2) Generating vacant lattice defects and producing $N$–$V$ centers. Chang’s method of irradiating solid nanodiamond particles with high-energy beam like $\alpha$- or $\gamma$-ray and annealing under nitrogen atmosphere at the highest possible temperature (perhaps 800 $^\circ$C) appears the most general and convenient one[27]. The color center that can be used for ODMR is at the moment restricted to $N$–$V$ types, as the splitting of energy levels for the triplet spins in the ground state (2.87 GHz) matches with that of magnetic resonance frequency of 2.75 GHz.

Other elements like Si and Ni have been claimed to have certain merits regarding luminescent characteristics and are worthwhile studying in view of the wide range of applications. However, incorporation of these elements into lattice defects poses new challenges. Irradiation with accelerated ion beams generated from these elements followed by annealing might be one possible way. Alternatively detonation of Composition B in the presence of these elements might work, especially for Si.

4.3. Seeding for Single-Crystal Diamond Film by CVD

Naturally, the final goal of homoepitaxial diamond growth by CVD technique is large-area single-crystalline diamond films, the starting materials for the diamond semiconductors. However, our NSCD appeared only recently, long after the grand project for single-crystalline CVD diamond film started. In the absence of single-nano diamond particles as the seeds for homoepitaxial diamond growth, the pioneering researchers in this field concentrated their efforts to the heteroepitaxial growth. They soon discovered that applying bias electric voltage on the growing crystal surfaces led to enhanced nucleation (Bias-Enhanced Nucleation). In this way, highly oriented polycrystalline diamond films grow (Highly Oriented Diamond). These were the hay days of heteroepitaxial diamond growth in the early 1990s, but thereafter the developments slowed down except for the ECR strategy for generating high-density plasma and achieving high-speed CVD process. It seems difficult for the heteroepitaxial strategy to reach the goal of mass-producing single-crystalline diamond films at least in the foreseeable future.

We slowly began testing our NSCD particles as the seeding for homoepitaxial growth of nanodiamond film by CVD. At first, we thought high two-dimensional seeding densities will be achieved only by using ink-jet printers. We indeed reached $2 \times 10^{11}$ particles/cm$^{-2}$ by a commercial industrial ink-jet printer, but this result was only marginally better than simple immersion method into 0.1–0.01% aqueous solution of NSCD [28]. Even at this preliminary levels of seeding density and orientation, the advantage of homoepitaxial over heteroepitaxial nucleation was clear (Figure 4). According to the latest results of Williams, where a seeding density exceeded $10^{12}$ particles/cm$^{-2}$, he obtained polycrystalline diamond films having Young’s modulus of 1100 GPa with thickness below 140 nm, while keeping the crystal grain size below 100–200 nm under a low feeding rate of methane at a high microwave energy density of 25 W/cm$^3$ [22]. This level of Young’s modulus corresponds to those of the natural bulk diamond. Such ultrathin films will be suited for coating intricately shaped objects like micro-electro-mechanical system (MEMS) made of Si, as well as internal wall of a bead mill and its stirring wings for diamond mill, an ultimatum milling system referred above.

We can readily think of a few practicable strategies to improve the quality of homoepitaxially grown diamond films or even to realize highly oriented and even single-crystalline diamond films. First of all, it should be definitely better to remove graphitic layers in NSCD before seeding, even though the surface graphitic layers in the pristine NSCD will be reduced sooner or later by heated hydrogen during CVD process. Actually, HNSCD mentioned above must have been conceived by Williams for this purpose [21]. Other tactics include more effective compression and alignment of seeds by means of Langmuir-Blodgett (LB) trough technique involving derivatization of NSCD particles for better formation of LB membrane and application of Bias Enhanced Nucleation (BEN) technique. In conclusion, we believe that homoepitaxial diamond growth by using our nanodiamond is a promising choice for diamond semiconductors.

4.4. Drug Delivery System Using NSCD Particles as Platform

Application of NSCD particles as the novel platform of drug delivery system (DDS) has been well documented [29] and hence will be commented here only briefly on the latest developments. A long-pending question on this new DDS system is the reason why the severe side-effects of anticancer drug doxorubicin could be suppressed so well. In our present system, drug molecules are absorbed (not bound) on the surface of NSCD particle and directly exposed to blood stream before reaching to ailing cells. In the case of doxorubicin hydrochloride, it seems to have been adsorbed to the surface of NSCD by a kind of H-bonding of positively polarized proton in the $N^+−H$ group of drug molecule with the negative charge in the $\{111\}$ facet of NSCD. It is likely that a small amount of
drug molecules will dissociate during transportation through the blood stream.

The drug molecules will be more firmly held to the carrier if it had been intercalated between layers of nanographene extending over the fused \{111\} facets. It was recently reported that thin layers of graphene oxide intercalate remarkably well with \(\text{C}_6\text{O}\) and water [30,31]. Edges of \{111\}nanographene between the 4/6 facets, which are likely oxygenated, serve as the entrance/exit for intercalation of guest molecules. Once intercalated between grapheme layers, it would be hard for the drug molecules to get out of the internal position to outside. In the slow release step, salting out agent like \(\text{Cl}^-\) or proton will quickly get in and get out of the intercalating space. Solvent water molecules will be readily intercalated and de-intercalated due to weaker force of intercalation compared with that of large drug molecules. Then, the nonfreezing water molecules observed in hydrogel of NSCD by DSC could also have been the intercalated water. Experimental proof for this intercalation hypothesis is going on in our laboratories and will be reported elsewhere.

An exciting case of anti-chemoresistance activities that surfaced during small animal tests [32] appears to have some relevance with the intercalation hypothesis. Here NSCD-mediated delivery of doxorubicin to mice that had been affected by breast and liver cancer survived 100%. The results are considered to represent a broadly applicable new strategy for overcoming the so-called ABC transporter-mediated resistance during cancer chemotherapy [33]. It is less likely that bare doxorubicin simply adsorbed on the surface of NSCD carrier escapes recognition by ABC transporter proteins.

Another interesting but serious case of drug resistance is malaria [34]. Here the intercalation hypothesis will be more critically tested.

### 4.5. Catalyst Carriers

In view of the large surface area expected for NSCD particles (Table 3), there should be many other applications as disposable carriers in addition to DDS. Several industrial catalysts [35,36], anti-electrostatics for small objects like micromotors and a few others have already been tested and gave positive results.

### 4.6. Strengthening Additive in Compound Electroplating

This final section pertains to our initial purpose of this chapter, namely exploration of the nanodiamond-containing ceramics composites as a new application of diamond to ceramics, which would maintain high strengths of matrix but do not suffer from brittle fracture. The simplest tests for this purpose would be to perform electroplating experiments that will give us some perspectives on the nano-extension of Zener Pinning theory to thin metal films. A preliminary experiment of Cr plating on Ni-coated bronze using commercial plating liquids gave significantly positive results, producing up to three times increase in Vickers’ hardness under high loads (Figure 5). These results are
highly encouraging because we had no way of observing the dispersion of NSCD particles in the plating baths containing high concentrations of electrolytes. Note that variation in the added amounts of NSCD gave visible effects not only on the Vickers hardness but also on the modes of brittle fracture (Figure 6). Effective suppression of cracks on the Cr film upon pressing with indenters can be interpreted to come from the number effect.

**Vickers indentation trace (load - 200g)**

![Vickers indentation images](image)

**FIGURE 6** SEM images of Vickers indentation marks on the Cr-plating film co-deposited with NSCD particles. The Cr film shows the well-known brittle fractures (Cr pure), whereas Ni film does not (Substrate). The brittleness of Cr film is significantly suppressed by adding less than 25 wt% of NSCD. It is likely that prolonged agitation of the solution before electrolysis promoted re-aggregation of the once dissolved NSCD particles under acidic condition.
5. SUMMARY AND PERSPECTIVES

Characteristics of diamond are briefly compared with ceramic material. If ceramics acquire the ultrahardness of diamond, the danger of brittle fracture in ceramics would be surely removed but at the same time processability would be lost. For this fatal defect, diamond cannot compete with the high processability of other ceramics. In order that diamond can perform better as an industrial material, we suggest studying composites of nanodiamond with other ceramics, glass, and metals, which should give the composites extra hardness, show considerable resistance to brittle fracture, and still maintain excellent processability of ceramics. One advantage of using nanodiamond as a reinforcing component is that its composition can be kept minimal due to the number density effect of single-nano particles. We chose dispersed primary particles of detonation nanodiamond as the target, and named it as nano single-crystalline diamond (NSCD).

In recent years, as new characteristics of NSCD became known among researchers of nanocarbon, a number of attractive applications that make use of the novel functions of individual particles of NSCD are being discovered. A few of them are mentioned here in some detail. Work on NSCD composites is delayed due to technical difficulties, but encouraging results have emerged from the work of compound plating of metal films. It is hoped that a new class of ceramics based on nanodiamond reinforcements will be reported in the next edition of this book series.

ACKNOWLEDGMENTS

We thank Dr. K. Uemura and Dr. K. Shalnov for assistance in the electroplating experiments and analysis, and Dr. A. Barnard and Dr. L.-Y. Chang for discussion and drawings. Kind invitation of Professor S. Soumiya and Professor T. Enoki to contribute this Chapter is gratefully acknowledged. I am particularly indebted to Professor Soumiya for his patience to wait for the completion of my slow writing.

REFERENCES


Chapter 2.4
Catalytic Carbons — Cathode

Catalytic Carbons

Jun-ichi Ozaki
Graduate School of Engineering, Gunma University, 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan

Chapter Outline
1. Introduction 103
2. Carbon Alloys 105
References 109

1. INTRODUCTION
1.1. Low-Carbon Economy and Hydrogen Economy

A low-carbon economy is recognized as an important policy for the sustainable development of mankind. Actually the terminology means an economy to reduce the greenhouse gas emissions into the biosphere, where the word “carbon” does not definitely mean “carbon materials” but “carbon dioxide”. I myself as a carbon scientist would like to propose a proper wording such as “low-carbon-dioxide economy”. However, I will use it reluctantly in this chapter, since the terminology is accepted worldwide.

Many ways are available in order to construct the economic or social systems to minimize the CO₂ emission. Among them, hydrogen economy is one of the most important and interesting options for this purpose. There are three important and necessary technologies to achieve the hydrogen economy: (1) production, (2) storage, and (3) conversion of hydrogen gas. We intend to invent carbon materials that cover the above three technologies.

Fuel cell is the core technology for converting chemical energy of hydrogen to electric energy. A simulation predicted that the 80% reduction of the CO₂ emission compared to the present one can be achieved by using fuel cell vehicles (FCVs) widely, which is accompanied by the following two technologies: H₂ extraction from renewable resources, and CO₂ capture and storage (CCS) [1]. In the beginning of 2011, thirteen Japanese leading companies including Toyota, Nissan, Honda, and energy companies issued a proclamation on their decision to the commercialization of FCV in 2015 [2].

A Japanese research agency, New Energy Development Organization (NEDO), presents a roadmap on fuel cell and hydrogen-energy technologies. According to this roadmap, the cost should be reduced from 2—2.5 million Yen/kW of the present to less than 0.4 million yen/kW by 2030. NEDO also analyzed the breakdown of the present cost of the fuel cell systems: approximately 60% of the cost accounts for the cost of the FC stacks; a half of the cost is due to the electrode catalysts because of the high platinum requirement of 0.5—1.0 g/kW and of the cost of the metal [3].

1.2. Fuel Cells

Fuel cell is an energy-converting device from chemical energy to electricity. Several types of fuel cells are available regarding to the types of fuel and electrolytes as shown in Table 1. The proton-exchange membrane fuel cell (PEMFC) is in the spotlight among the types of fuel cells. Figure 1 shows a scheme explaining the principle of PEMFC. Hydrogen gas fed to the anode reacts to form proton and electron expressed by the following chemical reaction:

\[ \text{H}_2 \rightarrow 2\text{H}^+ + 2e^- \] (1)

The two charged species, proton and electron, are transferred via different routes to the opposite electrode, cathode. The generated protons travel through the electrolyte (proton-exchange membrane), while the generated electrons travel through the external circuit to the cathode.
During the transportation of electron through the external circuit, the electric energy is converted to other energy at the load. Oxygen, usually supplied as air, fed to the cathode reacts with the two charged species to form H₂O as the final product expressed by the equation

\[ \text{O}_2 + 4\text{H}^+ + 4e^- \rightarrow 2\text{H}_2\text{O} \]  

(2)

The total chemical reaction that occurs in the fuel cell is the formation of water from hydrogen and oxygen:

\[ 2\text{H}_2 + \text{O}_2 \rightarrow 2\text{H}_2\text{O} \]  

(3)

A serious problem lies in the electrode catalysts, particularly the cathode catalyst that promotes the oxygen reduction reaction (ORR) expressed by Eqn (2). The rate of ORR is lower than that of the hydrogen oxidation reaction (HOR, anode reaction); the rate of ORR expressed by is lower than the rate of HOR expressed by (1) by six orders of magnitude [4]. This means that the rate of total reaction (3), namely the current to the external circuit, is limited by the rate of ORR (cathode reaction). A large amount of platinum catalyst is used to promote ORR, which makes the cost of the fuel cell expensive, and hence impedes the commercialization of PEMFC as analyzed by NEDO [3].

Lowering the amount of Pt and developing nonplatinum catalysts are the ways to lead the PEMFC to commercialization. Methods to reduce the platinum metal loading are improvements in the platinum-loading procedures [5,6] and enhancement of the activity by alloying platinum with other cheaper metals [7]. Besides these attempts, the use of nonplatinum catalysts has also been studied. The most famous example is the use of N₄−M complexes, such as phthalocyanines and porphyrins; in particular, the oxygen reduction activity was enhanced by heat-treating the complexes loaded on carbon blacks; these catalysts will be shown in the next paragraph [8−11]. Other candidates are found in carbides [12,13] and palladium alloys [14]. The possibility of carbonaceous materials has been revealed by my research group; the catalysts are the subject of this chapter.

Let us review the history of ORR catalysts relating to carbon materials according to Figure 2. In 1965, Jasinski found an ORR catalytic activity of cobalt phthalocyanine [15]. Jahnke et al. extended the types of the nitrogen-containing transition metal complexes, and they

<table>
<thead>
<tr>
<th>Type of fuel cell</th>
<th>Proton-exchange membrane fuel cell (PEMFC)</th>
<th>Phosphoric acid fuel cell (PAFC)</th>
<th>Molten carbonate fuel cell (MCFC)</th>
<th>Solid oxide fuel cell (SOFC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte</td>
<td>Ion-exchange membrane</td>
<td>Phosphoric acid</td>
<td>Alkali carbonates mixture</td>
<td>Yttria-stabilized zirconia</td>
</tr>
<tr>
<td>Operating temperature (°C)</td>
<td>70−90</td>
<td>180−220</td>
<td>650−700</td>
<td>800−1000</td>
</tr>
<tr>
<td>Charge carrier</td>
<td>H⁺</td>
<td>H⁺</td>
<td>CO₂⁻</td>
<td>O₂⁻</td>
</tr>
<tr>
<td>Electrolyte state</td>
<td>Solid</td>
<td>Immobilized liquid</td>
<td>Immobilized liquid</td>
<td>Solid</td>
</tr>
<tr>
<td>Catalyst anode</td>
<td>Platinum</td>
<td>Platinum</td>
<td>Nickel</td>
<td>Nickel</td>
</tr>
<tr>
<td>Fuels for cell</td>
<td>H₂</td>
<td>H₂</td>
<td>Reformate or CO/ H₂</td>
<td>Reformate or CO/ H₂ or CH₄</td>
</tr>
<tr>
<td>Oxidant for cell</td>
<td>O₂/air</td>
<td>O₂/air</td>
<td>CO₂/O₂/air</td>
<td>O₂/air</td>
</tr>
<tr>
<td>Cell efficiency (%LHV)</td>
<td>40−50</td>
<td>40−50</td>
<td>50−60</td>
<td>50−60</td>
</tr>
</tbody>
</table>

TABLE 1 Types and Features of Fuel Cells

Reproduced from Chunshan Song, Catalysis Today 77 (2002) 17−49

![FIGURE 1 Principle of PEMFC. For color version of this figure, the reader is referred to the online version of this book.](image-url)
also reported important findings to improve the complex catalysts: utilization of carbon supports and heat treatment at proper temperatures, which resulted in high activity and longer durability [8].

The active sites of the catalysts have been investigated by many researchers and three categories of the candidates have been proposed: N-coordinated metal species formed on the carbon surface [16–24], small metal particles [25–27], and modified carbon surfaces [28–33]. The first theory has been widely accepted by the most researchers. However, the materials are not exact carbon catalysts.

The author’s group developed a type of nanocarbon, nanoshell, and showed its electrochemical activity in the 90s [34–37]. We invented two types of ORR catalysts based on carbon materials: nanoshell (NS) [38,39] and B/N-doped carbons (B/N-Cs) [40,41]. The active sites of these catalysts have been assumed to be some types of defects formed on the nanoshell surfaces, not the surface metal complexes like in the case of the MN₄ catalysts. In the next section, I would like to describe the preparation, structural features, electrochemical properties, and the active sites of the carbon-based ORR catalysts that we have developed.

2. CARBON ALLOYS
2.1. What is Nanoshell?
Firstly, let me define the terms, nanoshell (NS) and nanoshell carbon (NSC). NS is defined as the product of catalytic carbonization of polymer with transition metals as mentioned below; it is a type of nanocarbon that has a characteristic shape of hollow sphere with a diameter ranging from 10 to 50 nm, while NSC is the carbon material that contains NS structure in their bodies.

NS structure is considered to be formed by catalytic actions of the transition metals on the carbonization of organic polymers. Usually, the preparation of NS is to carbonize the thermosetting resin, such as furan resin, phenol–formaldehyde resin etc. in the presence of metal complexes such as phthalocyanines, acetylacetonates, and metalloencenes of cobalt, nickel, and iron [34–39]. These transition metals catalyze the formation of the NS structures, which are formed by thermal decomposition of the complexes during the carbonization. Because of the incompleteness of the NS formation, amorphous carbon is also produced as a matrix that surrounds the NS structures [34–39]. The term NSC is used to refer to the carbon materials that contain both NS and amorphous carbon components. The prepared NSC is usually treated with hydrochloric acid in order to remove the metal species, which are exposed by the pulverizing process of the lump of the product carbon.

The structure of NSC is represented in Figure 3. Transmission electron microscopy (TEM) revealed the formation of the round-shaped carbon fringes ((a)), which was the origin of the naming of nanoshell [38,39]. However, it is very difficult to tell whether the fringe is just an entangled ribbon of elongated graphitic structures like the Jenkins’ model of glass-like carbons [40] or projections of closed 3D-spherical shapes that we imagined. Scanning electron microscopy (SEM) observation (fig 3(b)) proves that the fringes observed by TEM are the cross-section of the spherical structure by comparing the size of the two objects observed in these two different electron microscope images. Thus, the structure of NS was concluded to be the hollow spherical graphitic structure [41].

---

**FIGURE 2** History of non-Pt cathode catalysts related to carbon materials. For color version of this figure, the reader is referred to the online version of this book.
The degree of the NS development was estimated by 002 X-ray diffraction, which is represented by $f_{\text{sharp}}$ [37–39]. The 002 diffraction of NSCs has shown to be a superimposition of crystalline and amorphous components; i.e. the typical 002 diffraction pattern consisted of a sharp component and a broad component. The parameter, $f_{\text{sharp}}$, is defined as the areal ratio of the sharp component to the total 002 diffraction [37]. The sharp component has been revealed to represent NS, which was identified by the agreement of the 002 lattice constants obtained by both XRD and electron diffraction of the selected area [39]. Hence, $f_{\text{sharp}}$ can be interpreted as a parameter representing the degree of development of NS structure in NSCs [39].

2.2. Electrochemical Activity of NSC

Carbon materials have been used as the catalyst support for platinum catalysts in PEMFC, because of their electrical conductivity, chemical durability, and high surface area. Figure 4 shows the typical voltammograms for electrochemical oxygen reduction reaction (ORR) on various carbons. Since ORR is the cathode reaction of PEMFC as previously mentioned, the preferable cathode catalyst should have larger current density at higher positive potential. The ordinary carbon (a carbon black) virtually has no ORR activity according to the above criteria. In contrast, the NSC shows higher ORR activity. Further improvement is available when NSC was doped with nitrogen, which will be reviewed in the later section. As can be seen in the representative ORR voltammograms in Figure 4, the generation of nanoshell into the carbon matrix and the introduction of nitrogen element is effective for increasing the ORR onset potential, which approaches to that of a platinum catalyst, Pt/C. For the purpose of quantitative comparison of the onset potentials for ORR, a parameter, $E_{O_2}$, has been defined by our group by reading the potential that gives $-10 \mu A/cm^2$ on the voltammograms [38,39].

The ORR activity represented by $E_{O_2}$ was revealed to depend on $f_{\text{sharp}}$, the development degree of the NS structure in an NSC [39]. Figure 5 shows the dependences of the ORR activity ($E_{O_2}$) on the NS development degree ($f_{\text{sharp}}$) for two different NSCs. The difference between these two series was the types of the metal complexes, acetylacetonates (open symbols) and phthalocyanines (closed symbols). Both series of the samples show similar trends; i.e. the activity increased with $f_{\text{sharp}}$ and then decreased by further increases in $f_{\text{sharp}}$, indicating the existence of a proper development degree of NS for ORR. Another interesting point was observed when comparing the activities of the NSCs belonging to different groups but with similar $f_{\text{sharp}}$ values: the NSC made with phthalocyanines (closed symbol) showed higher activities than the one made with acetylacetonates. The difference is caused by the incorporation of nitrogen into the NSCs like the way that we observed in the enhancement of ORR activity by the nitrogen doping as briefly introduced in the next section.
2.3. B/N-Doped Carbons

For exerting ORR activity to carbon materials, we also revealed the importance of the introduction of nitrogen and/or boron into the carbon structure other than the introduction of NS structure as above mentioned [42–44]. The B/N-doped carbon alloys can be prepared by carbonizing a thermosetting resin, furan resin, in the presence of melamine as a nitrogen dopant and BF$_3$–MeOH complex as a boron dopant. Figure 6 shows the ORR voltammograms of the carbons thus prepared [44]. The pristine carbon, UN, showed the lowest ORR activity. By the introduction of nitrogen into the carbon material (N1 and N2), enhancements in the ORR activity were observed. The same enhancement was observed in the case of boron doping (B1). Increasing the nitrogen-doping level did not lead to further increase in the ORR activity; however, the simultaneous doping of nitrogen and boron resulted in further enhancements in ORR activity (BN1, BN2, and BN3).

Detailed X-ray diffraction analysis revealed the inhibitive effects of the dopants on the growth of graphitic structures in NS. The analysis of two-dimensional diffraction of graphite, 11, with a theory proposed by Diamond enables us to obtain size distribution of the crystallite size along the $a$-axis ($L_a$) of the graphitic structure. The results of the analysis indicated the distribution of $L_a$ shifted toward smaller sizes by the doping of B and/or N. The reduction in $L_a$ will increase the edge sites that are assumed to be active sites for ORR; hence, we observed enhancements in ORR activity by B and/or N-doping to carbon materials [44].

X-ray photoelectron spectroscopy (XPS) provided a clue to elucidate the ORR active sites on the N-doped carbon materials. The N-doped carbons mainly contained pyridinic and pyrrolic nitrogen atoms (edge nitrogen) embedded in carbon hexagonal structures and showed a good correlation with $E_{O_2}$. The correlation indicated that the nitrogen species located at the edge sites of the carbon layers should be responsible for the ORR activity [44].


FIGURE 6 Hydrodynamic voltammograms of the N/B-doped carbons for ORR. For color version of this figure, the reader is referred to the online version of this book. [J. Ozaki et al. Carbon 45 (2007) 1847 with permission of Elsevier].
We do not understand whether or not the nitrogen species is directly related to the ORR mechanism. Some authors pointed out that the pyridinic nitrogen atoms are merely an indicator of the presence of the edge sites of the carbon layers [30, 31]. Ikeda et al. examined a possibility of electronic perturbation due to the introduction of nitrogen into the graphene structure by employing a quantum chemical calculation [45]. They showed an activation of the carbon atoms on the zigzag edge of graphene, when the ternary-coordinated nitrogen atom is located at the immediately inner alignment of the zigzag edge plane.

When a boron atom is simultaneously added to carbon structure with nitrogen, we observed the formation of B–N–C moieties in the materials, which had also been detected by XPS analysis. The summation of the edge nitrogen and the B–N–C moieties has been shown to have a good correlation with the ORR activity, $E_{O_2}$ [44].

### 2.4. Improvements of ORR Activities of NSCs

Many ways to enhance the ORR activities of NSCs are possible. Here I would like to present two examples: activation and nitrogen doping of NSCs.

Activation is a process to produce active carbons, which has large surface area normally ranging $400−1500$ m$^2$/g. They are used as adsorbents and electrodes for double-layer capacitors. To obtain large surface areas, carbon materials are reacted with activators, such as H$_2$O and CO$_2$ to form gaseous products. The formation of gaseous products like CO$_2$ or CO results in the formation of pores by removing carbon atoms from the material. Another type of activation is performed by using caustic chemicals such as KOH and phosphoric acid, which are mixed with carbon precursors and heated together. Although the mechanism of the activation is not fully understood, different activation methods result in the different pore shapes. Commonly, the activation proceeds in two stages of pore formation: removal of disorganized carbon and exposition of aromatic sheets to introduce microporosity in the first stage; widening the existing pores or formation of larger-sized pores. We intended to introduce active sites by removing the disorganized carbons in NSCs and also by exposing the edge sites of aromatic sheets by the activation treatments [46].

Figure 7 shows the changes in the ORR activity with burn-off when two types of activation processes were applied to NSC. Different dependences were observed: CO$_2$ activation of NSC enhanced the ORR activity while KOH activation reduced the activity. We expected that the activation treatments will increase the ORR activity, since the process introduces an enhancement in the surface area and also surface defects to NSCs as mentioned above [47].

Another improving treatment is to introduce nitrogen onto NSCs. The increase in the ORR current by CO$_2$ activation followed by nitrogen doping is larger than that by CO$_2$ activation. The results indicate that the post-treatments like these are important idea for improving the activities of NSCs [48].

### 2.5. Cell Performance

The actual fuel cell contains several hundreds of anode—electrolyte membrane—cathode pairs (MEA, membrane electrode assembly), which are connected in series to obtain high voltages. In research laboratories, a single cell, which contains only one MEA unit sandwiched by two conductive plates that have flow channel of both fuel and oxygen, is often used to examine the performance of the catalyst in the conditions close to the actual states. Figure 8 shows the cell current—cell voltage ($i−V$) curves obtained for a single cell...
equipped with NSC in cathode and commercial Pt/C in anode [39]. For comparison, the $i-E$ curve for a single cell with a Pt/C catalyst in both anode and cathode is also shown in Fig. 8. Surprisingly, the NS cathode catalyst showed about the two-thirds of the maximum power density of the Pt/C catalyst. Further improvements are reported by two groups based on our idea [49].

2.6. ORR Activity and Structure of Carbons [39]

The fact that the preparation process of NSCs is similar to that of the partially decomposed N$_4$–M complex catalysts [18,21,50,51] reminds that the active sites of our NSCs are similar to the sites of the latter catalyst, i.e. surface metal complexes formed on the support carbon materials, like M–N$_4$ or M–N$_2$ (M = Co, Fe). The characteristic features of the N$_4$-complex catalysts are as follows:

1. The activity was reduced by acid-washing treatment [26,52]
2. The presence of nitrogen and metal elements was necessary, but the utilization of phthalocyanines was not an indispensable factor in forming the active sites [20,21]
3. The ORR activity was considerably reduced when the shell-like structures were formed [11,21,27]

Detailed comparison revealed that the active sites of NSCs are not the same as the ones found for N$_4$–M complex catalysts.

The ORR activity of NSCs has a profound relevance with the defects of NS structures. Figure 9 compares the structures of carbons with different ORR activities. Figure 9 compares the structures of carbons with different ORR activities. (a) amorphous carbon, (b) active NS, and (c) less active NS. The ORR activity of NSCs has a profound relevance with the defects of NS structures. Figure 9 compares the structures of carbons with different ORR activities. Amorphous carbons such as carbon blacks, shown in (a), are inactive for ORR; hence, amorphous carbons do not contribute to the formation of active sites in NSCs. (b) shows the typical TEM image of the surface of the ORR-active NS. The NS shown in the NSC has corrugated carbon hexagonal layers and also has defects on the surfaces of the NS structures. As seen in Figure 5, further increase in $f_{\text{sharp}}$, the development degree of NS in NSC, resulted in the decrease in the ORR activity of the carbons. The NSCs with lower ORR activity showed unbent carbon hexagonal layers and very smooth surfaces. The comparison shown here revealed that the carbons with 'proper disorders' show higher ORR activities; hence, the defects introduced into the NS surface are responsible for the ORR activity.

The importance of the defects or properly disordered structure has also been proven by the TEM observations of the activated NSC; the CO$_2$ activation introduced the density of the defects while the KOH activation completely destroyed the NS structures [47]. These results emphasized the importance of the defective surfaces of NS structures. Such disorders change the electronic structure of the surface of carbons to be suitable for the reactions. Further investigation from both theoretical and experimental points of view is required to elucidate the nature of the active sites as well as the reaction mechanism. I believe these studies will lead to an improvement of the catalytic performance of NSCs.

REFERENCES

Chapter | 2.4 Catalytic Carbons — Cathode Catalytic Carbons


[48] Unpublished data.


1. INTRODUCTION

Graphite and carbon materials are used for fission and fusion reactors in the nuclear energy field. These materials have excellent mechanical and thermo properties at high temperatures, and are used as moderator and reflector to energetic neutrons produced in the reactors.

In the fission reactors, they are used for structural components in gas-cooled reactors, i.e., CO2 gas-cooled reactors and helium gas-cooled reactors [1]. The former CO2 gas-cooled reactors, Magnox reactors, and advanced gas-cooled reactors (AGRs) have been developed in the United Kingdom since 1950s. Graphite is used as moderators for neutrons in the reactor core, and graphite components are located to keep fuel rods in the reactors. The reactors can obtain high temperatures of CO2 gas of about 390 °C (Magnox) and 635 °C (AGR), which are used to generate electricity by a steam turbine system. The latter is high temperature gas-cooled reactors (HTGRs). The HTGRs are graphite-moderated and helium gas-cooled reactors, and they have been under development since 1960s. Graphite and carbon materials are used for in-core structural components such as fuel blocks, reflectors, and core support structures. These blocks have different functions as the in-core components. HTGRs can provide high temperature of helium gas of about 950 °C at the reactor outlet, which is applicable not only for power generation by gas and steam turbine systems but also for process heat source in a hydrogen production system [2]. Inherent and passive safety features of HTGRs are also attractive ones.

In the fusion energy field [3–7], graphite and carbon materials are used for high-temperature plasma-facing components for fusion devices, e.g., JT-60 of Japan Atomic Energy Agency (JAEA) in Japan, TFTR in the United States, and JET in the European Union. The application of carbon material in the ITER has been studied under the international cooperation. The application to the fusion field is summarized in a reference [3].

This paper describes on the application of the graphite and carbon materials in the fission reactors, especially for the HTGRs. The functions of the in-core components, neutron irradiation effects on graphite, and R&D items are described.

2. OUTLINE OF HTGRs

2.1. Outline of HTGR Development in the World

For the development of the HTGRs, experimental and demonstration reactors had been constructed and operated in the Europe and the United States from 1960s to 1980s [1]. The experimental reactors are Dragon (Euratom, 20 MW thermal power), AVR (Germany, 15 MW electric power), and Peach Bottom (USA, 45 MW electric power), and the demonstration reactors are Fort Saint Vrain (USA, 330 MW electric power) and THTR (Germany, 300 MW electric power). The HTGRs use multilayered ceramic-coated fuel particles (TRISO) about 1 mm in diameter to keep the fission product gases within the particles. The graphite and carbon materials are used for the in-core structural components such as fuel blocks, reflector blocks, and core support structures. The temperature of the reactor coolant, helium gas, at the outside reactor vessel was more than 700 °C. The power generation system by steam turbine
was demonstrated through the plant operation of these historical HTGRs [1].

In 1990s, the JAEA had developed the HTTR, the first HTGR in Japan, which is a test reactor with 30 MW thermal power [2,8]. Prismatic block-type fuel elements are used in the HTTR. It was demonstrated in 2004 that the HTTR can produce high-temperature helium gas of about 950 °C at the reactor outlet for the first time in the worldwide HTGR developments [9]. A long-term high-temperature operation test, 50 days of continuous operation with the outlet temperature of 950 °C, had been successfully carried out in 2010 [10]. The goal of the HTTR project is to demonstrate a thermochemical hydrogen production cycle, iodine-sulfur (IS) process, with nuclear heat from the HTTR.

The VHTR is a promising candidate for the Generation-IV nuclear energy systems, and an international collaboration has been carried out through the GIF [11,12]. Regarding the commercial HTGR/VHTR projects for individual country, the United States focuses on the next generation nuclear plant (NGNP), and China started the construction of high-temperature reactor-pebble-bed module (HTR-PM). The nuclear hydrogen development and demonstration (NHDD) project has been studied in Korea. JAEA has been carrying out R&Ds on the design of an original VHTR based on the technologies established through the HTTR construction and operation. JAEA’s design is for a prismatic block-type reactor, gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C) [13]. It can generate electricity and produce hydrogen by the IS process.

The VHTR is a promising candidate for the Generation-IV nuclear energy systems, and an international collaboration has been carried out through the GIF [11,12]. Regarding the commercial HTGR/VHTR projects for individual country, the United States focuses on the next generation nuclear plant (NGNP), and China started the construction of high-temperature reactor-pebble-bed module (HTR-PM). The nuclear hydrogen development and demonstration (NHDD) project has been studied in Korea. JAEA has been carrying out R&Ds on the design of an original VHTR based on the technologies established through the HTTR construction and operation. JAEA’s design is for a prismatic block-type reactor, gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C) [13]. It can generate electricity and produce hydrogen by the IS process.

The VHTR is a promising candidate for the Generation-IV nuclear energy systems, and an international collaboration has been carried out through the GIF [11,12]. Regarding the commercial HTGR/VHTR projects for individual country, the United States focuses on the next generation nuclear plant (NGNP), and China started the construction of high-temperature reactor-pebble-bed module (HTR-PM). The nuclear hydrogen development and demonstration (NHDD) project has been studied in Korea. JAEA has been carrying out R&Ds on the design of an original VHTR based on the technologies established through the HTTR construction and operation. JAEA’s design is for a prismatic block-type reactor, gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C) [13]. It can generate electricity and produce hydrogen by the IS process.

The VHTR is a promising candidate for the Generation-IV nuclear energy systems, and an international collaboration has been carried out through the GIF [11,12]. Regarding the commercial HTGR/VHTR projects for individual country, the United States focuses on the next generation nuclear plant (NGNP), and China started the construction of high-temperature reactor-pebble-bed module (HTR-PM). The nuclear hydrogen development and demonstration (NHDD) project has been studied in Korea. JAEA has been carrying out R&Ds on the design of an original VHTR based on the technologies established through the HTTR construction and operation. JAEA’s design is for a prismatic block-type reactor, gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C) [13]. It can generate electricity and produce hydrogen by the IS process.

The VHTR is a promising candidate for the Generation-IV nuclear energy systems, and an international collaboration has been carried out through the GIF [11,12]. Regarding the commercial HTGR/VHTR projects for individual country, the United States focuses on the next generation nuclear plant (NGNP), and China started the construction of high-temperature reactor-pebble-bed module (HTR-PM). The nuclear hydrogen development and demonstration (NHDD) project has been studied in Korea. JAEA has been carrying out R&Ds on the design of an original VHTR based on the technologies established through the HTTR construction and operation. JAEA’s design is for a prismatic block-type reactor, gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C) [13]. It can generate electricity and produce hydrogen by the IS process.

2.2. In-Core Graphite and Carbon Structures in the HTTR

Graphite and carbon materials are used for the in-core components of the HTGRs in order to obtain high temperature of helium gas at the reactor outlet. The components are categorized based on their safety function, replaceability, etc. In the followings, the application of the graphite and carbon materials to the HTGR in-core components is described with reference to the design of the HTTR [8,15,16].

The arrangement of graphite components in the HTTR is shown in Figure 1 [8]. The reactor core consists mainly of hexagonal fuel blocks, control-rod guide blocks, and replaceable reflector blocks. The active core, which is 2.9 m in height and 2.3 m in diameter, consists of 30 fuel columns and seven control-rod guide columns. The active core is surrounded by replaceable reflector blocks and permanent reflector blocks and supported by the core support components.

The graphite components are categorized into two kinds taking their safety function and replaceability into account [15]. One is core graphite components, the other is core support components. These graphite components have different use and functions as follows:

i) The core graphite components are replaced at regular interval; on the other hand, the core-support graphite components are permanent components.

ii) The core support components have more serious structural functions than the core graphite components.

Therefore, the stress limit for the core support components is set to be more severe than that for the core graphite components.

2.2.1. Core Graphite Components

The fuel blocks are graphite hexagonal prisms with an array of fuel holes. The fuel blocks are 360 mm across flats and 580 mm in height as shown in Figure 2 [8]. Three dowel pins are installed on the top face, and they engage with dowel sockets in the bottom face of the block above. TRISO-coated fuel particles with UO2 kernel are dispersed in the graphite matrix and sintered to form a fuel compact. Fuel compacts are installed in a fuel rod made of graphite, 34 mm in outer diameter and 577 mm in length. Fuel rods are inserted into the vertical holes in fuel block. Coolant helium gas of the reactor flows through gaps between the holes and the rods.

The fuel blocks are fabricated from the IG-110 graphite, isotropic fine-grained nuclear grade. Table 1 shows typical thermomechanical properties of the IG-110 graphite [8,15]. Control-rod guide blocks and replaceable reflector blocks have the same external shape as the fuel blocks. They are also fabricated from the IG-110 graphite.
2.2.2. Core Support Graphite Components

The assembly of core support is also shown in Figure 1 [8]. The hexagonal hot plenum block array is made up of two axial layers. This structure provides lateral and vertical positioning and support of the core array. The hot plenum block assembly contains passages which collect the primary coolant flow from the outlet of the columns and distribute it into the high-temperature plenum beneath the hot plenum blocks. These blocks are fabricated from grade PGX, medium- to fine-grained molded graphite. Table 1 shows typical thermomechanical properties of the PGX graphite [8,15]. The core support posts and seats are
designed to structurally support the core and hot plenum block array. The posts and seats are made of the IG-110 graphite with higher strength than the PGX graphite.

The thermal insulation layer at the core bottom consists of three blocks: lower plenum block, carbon block, and bottom block. The main function of the carbon block, ASR-0RB carbon, is to keep the metallic core support structures below 500 °C. Table 1 shows typical thermomechanical properties of ASR-0RB carbon [8,15].

The permanent reflector is a graphite structure surrounding the replaceable reflector and control-rod guide columns. The permanent reflector is an assembly of graphite blocks making 12 circumferential segments and 8 axial layers. These blocks are made of PGX graphite.

2.2.3. Design Criteria for Graphite Components in the HTTR

2.2.3.1. Outline of the Design Criteria

For the construction of the HTTR, JAEA (formerly, JAERI) established the graphite structural design code [14]. In this code, the ASME-CE-code draft for graphite core support structure, which was prepared in 1980 but not established, was partially modified in the items of biaxial failure theory, buckling limit, and oxidation effect on the basis of the experimental data [8,15,16]. The JAEA’s design code contains a set of material property data for the IG-110 graphite, PGX graphite, and ASR-0RB carbon being used in the HTTR core. The specified minimum ultimate strength, Su, is determined from statistical treatment of strength data such that the survival probability is 99% at a confidence level of 95%, because the ultimate strength exhibits considerable statistical scatter. Table 2 shows the Su values for graphites and carbon specified in the structural design code [14,15]. As per the environmental effect on the strength, both the increase of strength by neutron irradiation and the decrease of strength by oxidation are considered. On the other hand, the increases of strength by temperature and strain rate are not considered due to a conservative design viewpoint.

For the requirements on the quality assurance of graphite and carbon materials for the in-core components, industry-wide standards for the raw material and processing have not been established. Hence, for the quality assurance of in-core components of the HTTR [21], it was carried out on the basis of the requirements from reactor design. Their quality has been controlled during production and after machining with respect to their properties and specifications. Figure 3 shows the flow diagram of acceptance test program for the manufacturing process of graphite and carbon components in the HTTR [15,21].

2.2.3.2. Irradiation Effects on Graphite Material Properties

For the application of graphite materials to the in-core components, one of the important subjects is to take into account the neutron irradiation-induced property changes. The fundamental mechanisms on the irradiation effects on graphite were studied and summarized in references [22,23]. However, the material properties of graphites are dependent highly on their grades based on raw materials and forming process. The irradiation effects on the graphite are also dependent on their grades.

Here, the irradiation effects on the IG-110, as a reference of the nuclear grade of graphite, are briefly described based on the above mentioned design code for the HTTR [14]. The properties, considering the irradiation effects, are the

---

### Table 1: Typical Thermo-Mechanical Properties of Graphite and Carbon Materials (Un-Irradiated and Un-Oxidized) [8,15].

<table>
<thead>
<tr>
<th>Property</th>
<th>IG-110 graphite</th>
<th>PGX graphite</th>
<th>ASR-0RB carbon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk density (kg/m³) (300K)</td>
<td>1.78 x 10⁻³</td>
<td>1.53 x 10⁻³</td>
<td>1.65 x 10⁻³</td>
</tr>
<tr>
<td>Mean tensile strength (MPa) (300K)</td>
<td>25.3</td>
<td>8.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Mean compressive strength (MPa) (300K)</td>
<td>76.8</td>
<td>30.6</td>
<td>50.4</td>
</tr>
<tr>
<td>Young's modulus (GPa) (300K)</td>
<td>7.9</td>
<td>6.5</td>
<td>8.7</td>
</tr>
<tr>
<td>Mean thermal expansion coefficient (293–673K) (10⁻⁶/K)</td>
<td>4.06</td>
<td>2.34</td>
<td>4.40</td>
</tr>
<tr>
<td>Thermal conductivity (W/mK) (673K)</td>
<td>80</td>
<td>75</td>
<td>10</td>
</tr>
<tr>
<td>Grain size (10⁻⁶ m) (Mean)</td>
<td>20</td>
<td>800 (Max.)</td>
<td>2000 (Max.)</td>
</tr>
</tbody>
</table>

(*) Determined from the cord joining two points (one point is the one-third of the specified minimum ultimate tensile strength and the other is the one-third of the specified minimum ultimate compressive strength) on the stress-strain curve.
dimensional change, longitudinal elastic modulus, mean coefficient of thermal expansion, thermal conductivity, strength, and steady-state irradiation creep coefficient. In the design code for the HTTR, they are expressed in the graphical expressions of material property data as a function of fast neutron fluence about several irradiation temperatures. The design curves in the graphical expressions were determined based on the irradiation experimental results at the JMTR of JAEA, etc.

### 2.2.3.2.1. Dimensional Change

The irradiation-induced dimensional change is a fundamental neutron irradiation effect. Figure 4 shows the relative dimensional change of the IG-110 as a function of fast neutron fluence for several irradiation temperatures [14]. It is used for the design of graphite components in the HTTR. The dimensional change depends on irradiation temperature. The dimensional change behavior at higher neutron fluence is expressed in Figure 5 [24]. The curves in this figure are obtained by the intrapolation and...
extrapolation of the experimental data with evaluation. Generally, at the first stage of the irradiation, the bulk of the graphite is in shrinkage. This mechanism is expressed based on the irradiation behavior of grain and pore structure of graphite [22,23]. For each graphite grain, the neutron irradiation causes expansion in c-axis direction and shrinkage in the a-axis direction. The expansion in the c-axis direction is accommodated by pores among the grains, since nuclear graphites have about 20% of porosity in general. As a result, the bulk of the graphite, polycrystalline structure with grains and pores, shows shrinkage by the neutron irradiation at the first stage. When the neutron fluence is increased, the shrinkage is attained to the minimum point, “turn around”. Then, the dimensional change has turned out to be increased. It was explained that the residual stress given by the shrinkage process causes micro cracks in the bulk of graphite, and the cracks work to increase the dimensional change rate. The fast neutron fluence of the turn-around point depends on the graphite grade and irradiation temperature. For the IG-110 graphite, it is about $2 \times 10^{26} \text{n/m}^2(E > 0.1 \text{MeV})$ at irradiation temperature of 600 °C [24].

2.2.3.2.2. Longitudinal Elastic Modulus Figure 6 shows the change of longitudinal elastic modulus of the IG-110 as a function of fast neutron fluence for several irradiation temperatures [14]. The change in elastic modulus due to irradiation is calculated from the change of velocity of longitudinal ultrasonic wave and the un-irradiated value of the longitudinal elastic modulus. At the initial stage of the irradiation, as shown in the figure, the modulus is rapidly increased by the pinning effect of dislocations in graphite crystals [22,23]. Then, at the next stage of the irradiation, the modulus is gradually increased by the irradiation-induced dimensional shrinkage, i.e. the decrease of porosity of graphite.

2.2.3.2.3. Mean Coefficient of Thermal Expansion The change of the mean coefficient of thermal expansion of IG-110 as a function of fast neutron fluence is shown in Figure 7 for several irradiation temperatures [14]. The mean coefficient of thermal expansion is increased at the initial stage of the irradiation and then is decreased by the irradiation.

2.2.3.2.4. Thermal Conductivity Figure 8 shows the change of the thermal conductivity of IG-110 as a function of fast neutron fluence for several irradiation temperatures [14]. The conductivity is rapidly decreased by the initial stage of the irradiation. Irradiation-induced defects reduce the mean free path of phonon, which dominates the thermal conductivity of graphite [22,23]. The lower irradiation temperature gives the greater decrease in the conductivity. It is because that high-temperature irradiation causes annealing recovery of irradiation-induced defects during the irradiation.
2.2.3.2.5. Strength

The change of the compressive and tensile strengths of IG-110 graphite by fast neutron irradiation is shown in Figure 9 [14]. In this neutron irradiation range, both strengths are increased by the irradiation.

2.2.3.2.6. Steady-State Irradiation Creep Coefficient

Irradiation-induced creep behavior is evaluated from the dimensional change due to irradiation with loaded condition. An irradiation-induced creep equation is expressed based on the model, which is the combination of the Maxwell model (steady-state irradiation creep) and the Kelvin model (transient irradiation creep) [15].

\[
\varepsilon_c = a \sigma \left\{ 1 - \exp(-b\gamma) \right\} + K\sigma\gamma
\]  

(1)

Here,

- \(\varepsilon_c\): Irradiation induced creep strain
- \(\sigma\): Stress (MPa)
- \(\gamma\): Fast neutron fluence \((\times 10^{26} \text{ n/m}^2, E > 0.1 \text{ MeV})\)
- \(a, b\): Transient irradiation creep parameters
- \(K\): Steady irradiation creep coefficient
- \(E_0\): Longitudinal elastic modulus before irradiation (GPa).

Figure 10 shows the steady-state irradiation creep coefficient of IG-110 as a function of irradiation temperature [14].

\(\bar{\alpha}\): Mean coefficient of thermal expansion after irradiation from 20 °C to irradiation temperature

\(\bar{\alpha}_0\): Mean coefficient of thermal expansion before irradiation from 20 °C to irradiation temperature

FIGURE 7 Change of mean coefficient of thermal expansion of IG-110 as a function of fast neutron fluence [14].

FIGURE 8 Change of the thermal conductivity of IG-110 as a function of fast neutron fluence [14].
3. R&Ds FOR VHTR DEVELOPMENTS

3.1. Use of Graphite Components for Long Service Life

Mechanical and thermo stability of in-core graphite components is important during their service life in reactor operation. The service life of the core graphite components in the HTTR is about $1.5 \times 10^{25} \text{ n/m}^2 (E > 0.18 \text{ MeV})$ in the fast neutron fluence. It is possible to extend the service life for commercial reactors based on the in-core design. The service life of the core graphite components in the GTHTR-300C is evaluated as the fast neutron fluence of about $5 \times 10^{25} \text{ n/m}^2 (E > 0.18 \text{ MeV})$. For the prismatic block type reactor, graphite components are replaceable. The fuel graphite blocks in the GTHTR-300C will be replaced at every four-year interval during the reactor operation. For Pebble bed-type reactors, the maximum fast neutron fluence may be in the order of $10^{26} \text{ n/m}^2 (E > 0.1 \text{ MeV})$ in case that core graphite components are not replaced throughout the reactor lifetime. It is important to design the graphite components to bear the severe fast neutron irradiation. Since the in-core graphite components of the VHTR are used subject to severe condition, it is important to assess their structural integrity during the service life.

Residual stress is accumulated in the graphite components by irradiation-induced dimensional change due to thermal gradient and neutron flux distribution within the component during the reactor operation. The lifetime of the graphite component is determined by the comparison of the residual stress and the stress limit. The allowable service life is determined so that the irradiation-induced residual stress does not exceed the limit [25]. For the stress limit of the graphite components in the design standard of the graphite components in the HTTR, it is set as one-third of the specified ultimate minimum strength $S_u$ for the core graphite components, which are replaceable components [15]. The graphite grade with high strength will give the high stress limit. It is important to use a graphite grade with high strength to use in the long service life. In this point, IG-110 graphite used in the HTTR has enough high strength.

On the other hand, the residual stress is evaluated by viscoelastic analysis taking uncertainty of in-core temperature and neutron irradiation condition into account [26]. The evaluated residual stress, hence, contains some safety margin. It is possible to extend the allowable components lifetime by reducing the margin with reasonable way, if we can measure the residual stress directly. JAEA is developing the nondestructive evaluation technique for the residual stress of graphite components by using a micro-indentation method. It uses the change of the indentation behavior loading/unloading at the residual stress condition.
of graphite. Since the indentation behavior is affected by the residual stress condition, it is possible to evaluate the stress from microindentation behavior. The details of this study are shown in references [27,28].

### 3.2. Standard for the Graphite Components of VHTR

For the design of the graphite components in the HTTR, the graphite structural design code for the HTTR [14] and inspection standard of graphite for the HTTR [21] were applied. It was so because the general standard system for the HTGR had not been established at that time. Recently, to prepare the general standard system, a “special committee on research on preparation for codes for graphite components in HTGR” at Atomic Energy Society of Japan (AESJ) had established draft of standard for graphite components in HTGR (including VHTR) in March 2009 [29]. In the United States, a subgroup in the American Society of Mechanical Engineers (ASME) is working to establish a new design code for graphite core components, which can be applicable to gas-cooled reactors including VHTR.

For the establishment of the standard of the AESJ, the standards for the graphite components of HTTR and ASME were reviewed. The established standard is based on the HTTR graphite design code [14] and the inspection standard of the graphite components for the HTTR [21]. The established “draft of standard for graphite core components in high temperature gas-cooled reactor” is composed of (1) general rules, (2) design standard, (3) material and product standards, (4) in-service inspection and maintenance standard, and (5) appendix and explanation.

The graphite components are classified into three categories (A, B, and C) in the standpoints of safety functions and replaceability. In the design section, the stress limits are determined by deterministic evaluation in principle. Alternatively, it is possible to determine the stress limits by probabilistic evaluation with an appropriate fracture probability. The general requirements for the in-service inspection and maintenance standard are determined. When a defect larger than a design assumption is found in graphite components by the in-service inspection, the cause shall be investigated, and the propriety of continuous use shall be judged by application of fracture mechanics, etc.

As an appendix of the design standard, the graphical expressions of material property data of IG-110 graphite as a function of fast neutron fluence are expressed. The graphical expressions were determined through the intrapolation and extrapolation of the irradiated data. The irradiation data for historical grades H-451 and ATR-2E were used for the evaluation of the intrapolation and extrapolation of irradiation data for IG-110. The details of the intrapolation and extrapolation method were reported in references [24,30].

### 3.3. Development of Control Rod Made of C/C Composite

The control rods of HTGRs are used at severe temperature condition. At the reactor scram, control rods are immediately installed into the reactor core with high temperature. For the HTTR, ferritic superalloy Alloy 800H is used for the metal parts of the control rod [8]. Its maximum allowable temperature to be used repeatedly after the scrams is 900 °C. On the other hand, for the VHTR, thermo condition of the core will be more severe than that of the HTTR. It is important to develop the control rod made of heat-resistant ceramics composite-material substitute for metallic materials. It is one of the major subjects for the VHTR development. The heat-resistant ceramic composite, C/C composite, and SiC/SiC composite are major candidates for this purpose.

Figure 11 shows the scheme of R&Ds at JAEA for the control rod made of C/C and SiC/SiC composites [31]. The R&D items are categorized in the following phases: (1) database establishment, (2) design and fabrication, and (3) demonstration test in the HTTR.

For the database establishment in phase (1), the data on material properties at un-irradiated and irradiated conditions are necessary. JAEA focuses especially on a two-dimensional C/C composite (2D-C/C composite) that has the layer structure of laminae composed of fibers and matrix [32]. Since the 2D-C/C composite has great anisotropy in thermal and mechanical properties in parallel and vertical to lamina directions, it is important to take the anisotropy into account for design of the components. Neutron irradiation effects on the properties are also anisotropic and important. A PAN-based and plain-woven 2D-C/C composite, CX-270G grade, is one of the candidate materials for the control rod application. It is graphitized at

![FIGURE 11 Scheme of R&Ds on application of C/C composite to control rod (31). For color version of this figure, the reader is referred to the online version of this book.](image)
TABLE 3 Typical Material Properties of CX-270G and IG-110 Graphite [32].

<table>
<thead>
<tr>
<th></th>
<th>Bulk density (Mg/m³)</th>
<th>Bending strength (MPa)</th>
<th>Tensile strength (MPa)</th>
<th>Elastic modulus (GPa)</th>
<th>Thermal expansion coefficient (10⁻⁶/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CX-270G (pl)</td>
<td>1.63</td>
<td>167</td>
<td>81</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>(2D-C/C) (pp)</td>
<td></td>
<td>133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG-110 (graphite)</td>
<td>1.77</td>
<td>39</td>
<td>25</td>
<td>10</td>
<td>4.6</td>
</tr>
</tbody>
</table>

pl: parallel lamina direction, pp: perpendicular lamina direction

2800 °C in its manufacturing process. Table 3 shows its typical material properties in comparison with fine-grained isotropic IG-110 graphite [32]. For the design and fabrication in phase (2), it is necessary to determine the control rod structure. Connecting method for the component parts is one of the key technologies for the fabrication of the control rod. Appropriate design methodology, e.g. fracture theory, for the fiber-reinforced ceramics has not been fully developed yet. It is also an important issue for the C/C and also SiC/SiC composite application to the control rod for the VHTR.

For the demonstration test by the HTTR in phase (3), it is the final stage for the composite control rod development. The large scale component of C/C and SiC/SiC composite can be irradiated in the HTTR, and the performance of it can be confirmed.

4. CONCLUDING REMARKS

HTGR/VHTR can provide high temperature of helium gas about 950 °C at the reactor outlet, which is applicable not only for power generation by gas and steam turbine systems but also for the process heat source in a hydrogen production system. Graphite and carbon materials are used for in-core structural components of the reactors, such as fuel blocks, reflectors, and core support structures. The in-core components are categorized based on their functions and replaceability. Development of general design code for the in-core components are been carrying out in Japan and the United States for the commercial HTGR/VHTR. Since neutron irradiation effect, depending on graphite grade, is one of the important subjects to be considered in the reactor design, it is necessary to establish the irradiation database for the candidate graphite grades. For the control rod of the VHTR, it is important to develop the control rod made of heat-resistant ceramics-composite materials, C/C and SiC/SiC composites, to substitute for metallic materials in order to use high temperature condition in the core. JAEA is carrying out the R&Ds for the ceramics-composite control rod, and it is possible to demonstrate it by using the HTTR in the future.

REFERENCES


1. INTRODUCTION

We use a number of electronic devices in our daily life, benefiting from them, and these electronic devices use a great many number of semiconductors; in other words, it may safely be said that they support our life.

In manufacturing processes of semiconductors that exist close to us, as we have just mentioned above, a number of carbon parts are used.

In this article, carbon parts used in various manufacturing processes of silicon semiconductors are described. The carbon mainly used in current silicon semiconductors manufacturing is isotropic graphite. The designation of isotropic graphite comes from the fact that the properties of a graphite block manufactured is about the same in all directions, and isotropic graphite was produced on a commercial basis for the first time in the world by Toyo Tanso (Japan) in 1974.

Having excellent properties of high density and high strength in comparison with conventional graphite, isotropic graphite is used extensively not only in the manufacture of silicon semiconductors but also in electric discharge machining electrodes used in manufacturing molds, dies for casting copper alloys and the like continuously, metal melting crucibles, hot-pressing molds, core materials for high-temperature gas-cooled reactors, and other similar uses.

Figure 1 shows the manufacturing processes of this isotropic graphite.

Isotropic graphite is produced through these processes, with a time of five to six months taken for the raw materials to be processed into graphite blocks.

Table 1 shows the characteristics of our typical materials, often used in the manufacture of silicon semiconductors, among a number of kinds of isotropic graphite.

2. SILICON SEMICONDUCTOR MANUFACTURING PROCESSES

Silicon semiconductors are manufactured roughly in the processes shown in Figure 2; the processes underlined in black use graphite.

The reason why graphite is used in these processes is that it generally has the following properties meeting the requirements for parts used in these processes:

1. Lightweight
2. Resistant to heat and not melting under normal pressure (sublimation temperature: about 3650 °C)
3. Low vapor pressure (3) (about 3.8 × 10⁻² mmHg at 2500 °C)
4. High emissivity (about 0.7 to 0.8)
5. Suitable electric resistivity (an important factor for use in heater applications)
6. An increase in strength proportional to temperature (flexural strength at 2500 °C in a nonoxidizing atmosphere about 1.5 times that at room temperature)
Figure 1: Manufacturing processes of isotropic graphite. For color version of this figure, the reader is referred to the online version of this book.

1. **Raw materials**
   - Filler: coke, Binder: coal tar pitch
   - Filler raw materials coming in the form of pellets or in lumps are pulverized to fine powder with a size of about 10 microns. (Some raw materials come in pulverized form, so they need no pulverizing process.)
   - Pulverized raw materials are stirred and homogenized to prevent them from separating by grain size and resulting in an uneven size distribution.

2. **Pulverizing**
   - With the fillers (aggregates), coke and graphite, not having binding force and hence not capable of being molded by direct pressurization, the binder (bonding agent), coal tar pitch, is added to the fillers, which are kneaded while being heated.
   - With kneaded raw materials forming pebble-like lumps on cooling and not capable of being molded directly, they are pulverized for a second time.

3. **Kneading**
   - Since excessively fine or coarse powder is not suitable for molding, particle size is adjusted by sieving it into grain size suitable to molding during the pulverizing process.

4. **Pulverizing**
   - Adjusted powder is charged into a rubber mold, which is sealed against water intrusion and immersed into water pressure tank, and water pressure is applied on the mold to mold the powder. Applying uniform force to the entire surface of the rubber mold in a water pressure tank causes grains to be oriented not in a given direction but in randomly oriented texture; this allows an "isotropic" block to be molded in which the difference in characteristic between the longitudinal and the lateral direction is reduced. Sometimes graphite produced using this molding device is called a "CIP" material because the molding device referred to as a Cold Isostatic Press (abbreviated to CIP).

5. **Molding (CIP)**
   - A molded block is put in a metal container, covered with packing coke to be insulated from air, and heated gradually to about 1000°C for carbonization. (This process usually takes about one month.)

6. **Baking**
   - A baked block is a porous body with microscopic pores, and pores in a block are impregnated with pitch with the aim of improving the apparent density and strength. (Some graphite grades are not impregnated.) A baked block is put in a pressurizing container, and pitch whose viscosity is lowered by heating is charged into the container and pressurized to impregnate pores in the block. After the impregnation, the block is baked again to carbonize pitch with which pores are impregnated.

7. **Pitch impregnation**
   - Allow a large current to flow through the baked block placed between the electrodes covered with coke grains to insulate the block from the air, and the coke grains and block will act as resistors and generate heat, heating themselves up to about 3000°C. This process takes about two months. A furnace based on the scheme is called an Acheson furnace.
7. High thermal conductivity and low coefficient of thermal expansion make it resistant to thermal shock and stable in dimensions.
8. Easy machining
9. Capable of being purified (ash contents can be reduced to the order of ppm.)

Silicon semiconductors are made from silicon purified to an ultra-high purity. That purity is called eleven nines, that is, the purity of silicon is 99.999999999%.

The use of ultra-high-purity silicon as raw material to manufacture semiconductors in this way naturally leads to the requirement for a high purity in graphite material used in the manufacturing processes of the silicon raw material.

Ordinary graphite contains impurities of the order of 400 mass ppm or so. Purifying this raw material at temperatures over 2000 °C in the stream of halogen gas causes metal impurity contents to volatilize in the form of halogen compounds with low boiling points and nonmetal impurities to evaporate; this makes it possible to obtain high-purity graphite with an impurity content of less than 20 mass ppm. In addition, improved processing methods will make it possible to manufacture ultra-high-purity graphite with an impurity content of less than 5 mass ppm.

### TABLE 1 Typical characteristics of some grades by Toyo Tanso

<table>
<thead>
<tr>
<th>Grade</th>
<th>Apparent density Mg/m³</th>
<th>Hardness HSD</th>
<th>Electrical resistivity μΩ·m</th>
<th>Flexural strength MPa</th>
<th>Compressive strength MPa</th>
<th>Coefficient of linear thermal expansion $\times 10^{-6}$/K</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG-110</td>
<td>1.77</td>
<td>51</td>
<td>11.0</td>
<td>39</td>
<td>78</td>
<td>4.5</td>
</tr>
<tr>
<td>IG-710</td>
<td>1.83</td>
<td>58</td>
<td>10.0</td>
<td>47</td>
<td>103</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Temperature range over which the coefficient linear thermal expansion was measured: 623 ~ 723 K.

FIGURE 2 Manufacturing processes for silicon semiconductors. For color version of this figure, the reader is referred to the online version of this book.
3. MANUFACTURING PROCESSES FOR POLYCRYSTALLINE SILICON

The manufacture of silicon as raw material for silicon semiconductors begins with the manufacture of metallic silicon by means of reducing relatively pure silica.

After this process, metallic silicon is subjected to the reaction with hydrogen chloride to remove impurities in metallic silicon and to form trichlorosilane, which is distilled repeatedly to produce high-purity trichlorosilane.

Using this highly purified trichlorosilane gas, polycrystalline silicon is grown around a silicon seed rod in a reaction vessel heated up to about 1200 °C by means of the CVD method, and a seed rod with a diameter of about 10 mm is grown to a polycrystalline silicon rod with a diameter of about 100–150 mm in the final process.

This method is called the Siemens method, with this device schematically shown in Figure 3. In this process, it is necessary to pass current through the seed rod for the rod itself to be a heating element, with the graphite part serving both as the electrode and as the seed rod holder to allow current to pass through the silicon seed rod.

With the CVD process requiring a high temperature between 1000 and 1200 °C, the characteristics of graphite of the absence of reduction in strength and the possession of adequate electrical conductivity under high temperatures are used; pieces of graphite used in this way are commonly called seed chucks.

In actual manufacturing processes, the graphite seed chuck undergoes an action to break it to collect polycrystalline silicon manufactured by the Siemens method, and, for such a use, a less expensive graphite material made of relatively coarse grains called an extruded graphite is used in the main.

In addition to these uses, graphite is used as the preheater to heat the silicon seed rod and the nozzle introducing gas into the furnace, and the like.

4. MANUFACTURING PROCESS FOR MONOCRYSTALLINE SILICON

This is the process by which monocrystalline silicon rods are produced with ultra-high-purity polycrystalline silicon manufactured in the above method as the raw material.

Two methods are available to produce a silicon single crystal, the CZ (Czochralski) and the FZ (floating zone) methods. In both methods, a single crystal is grown by contacting polycrystalline silicon with a monocrystalline seed. However, the techniques are different from each other, with the FZ method using few carbon parts.

In the CZ method, to grow a single crystal, on the other hand, a monocrystalline seed is brought into contact with melted polycrystalline silicon, and the seed is pulled upward gradually while being rotated slowly. Almost all of furnace internal components are made of purified graphite material or carbon heat insulator; this process is the one using the largest number of graphite parts in the silicon semiconductor manufacturing processes.

Figure 4 shows a silicon single-crystal manufacturing furnace based on the CZ method schematically. In the illustration, typical designations are used for different parts because they are called differently depending on the user and device manufacturer.

By means of such growing furnaces, huge silicon single crystals with a diameter of 300 mm and a length of over 2 m are manufactured.

Graphite parts are usually machined from a cylindrical block. However, the larger the diameter of a silicon single crystal becomes, the larger that of the graphite used; for this reason, the diameter of a crucible to make a 300-mm-dia. silicon single crystal exceeds 800 mm. As a result, the graphite heater to heat such a crucible and the heat shield member installed outside the heater to intercept heat from the heater have diameters as large as 1000–1300 mm.

In connection with this, the upper parts exposed to Si vapor and SiO gas at high temperature tend to suffer from cracking and chipping as a result of the reaction between carbon and silicon. In addition, carbon contents mixed in manufactured single-crystal silicon as a result of the occurrence of the above cracking and chipping are quite capable of leading to problems such as abnormal product quality.

As a means of resolving such a problem, graphite parts are used whose surfaces are coated with SiC or pyrolytic carbon.

Lightweight and high-strength C/C composite materials, which are manufactured by laminating carbon fiber in response to an increase in the diameter of parts in the
furnace, are finding their applications in crucibles and shielding materials.

5. PROCESSES OF MACHINING A SILICON SINGLE-CRYSTAL INGOT

After its perimeter has been machined to the desired diameter, a silicon single-crystal rod prepared by the CZ or the FZ method is provided with a plane called an orientation flat or a small notch to show the crystal orientation.

After this, the single-crystal ingot is sliced to a thickness of 1 mm with a slicing device shown in Figure 5.

In the past, an inside perimeter edge blade was used to slice off disks one by one. As a result of the difficulty in manufacturing an inside perimeter edge blade with a large diameter to match an increase in the diameter of a single-crystal ingot and an increase in the production efficiency, the saw slicing with a wire saw capable of producing a number of slices simultaneously has become the mainstream production method.

In carrying out these machining operations, the single-crystal ingot is bonded to the slicing beam that is made of graphite. Being porous, graphite is easy to bond and cut and allows a wafer to be cut and separated easily; this is the reason why graphite is often used.

6. MANUFACTURING PROCESS OF EPITAXIAL WAFERS

In some cases, as needs arise, the CVD method is used to grow thin single-crystal silicon film epitaxially on the
surface of a mirror wafer, finished in a mirror-like smooth surface by grinding a sliced wafer; this is called an epi-wafer.

Epitaxial manufacturing devices are roughly classified into the pan-cake type and the barrel type, which are shown schematically in Figure 6.

In this process, SiC-coated high-purity graphite is used for the platform, called the susceptor, on which to place a mirror wafer.

The reason why the platform is called the susceptor is that the platform, made of SiC-coated high-purity graphite and caused to generate heat by induction heating, serves also as a heater to heat the mirror wafer placed on it. The reasons for providing SiC coating are:

1. The purity of SiC is higher than that of graphite;
2. It does not easily react with the silicon wafer with which it comes in contact; and
3. Being porous, graphite adsorbs much gas, having the property of releasing the gas adsorbed when heated. The prevention of gas release by means of coating reduces the adverse effect on the silicon wafer grown epitaxially.

The difficulty of processing two or more wafers placed on the platform simultaneously due to upsized wafer diameters and the requirement for the quality improvement such as membrane thickness uniformity have led to the frequent use of single-wafer-processing susceptors that process a single large-diameter wafer placed on them.

7. CVD MEMBRANE FORMING PROCESS

With semiconductor manufacturing processes requiring many layers of circuits to be stacked, securing insulation between layers becomes necessary, and membranes made of SiO₂, SiON, SiN, and the like are used as insulating membranes.

Used as devices to form these insulating membranes are atmospheric pressure CVD systems (Figure 7) and plasma enhanced CVD systems (Figure 8). High-purity graphite and SiC-coated graphite are used also in platforms to carry wafers in the furnace and plates to hold wafers in the furnace, for example.
SiO$_2$ is used for interlayer insulating membranes in the circuit formed on the surface layer of a wafer; however, the unnecessary portion of the SiO$_2$ membranes must be removed after they were formed by means of the CVD method. For this purpose, the etching gas is turned into plasma, and, using its energy, the unnecessary portions of the insulating membranes are removed in a dry-etching process.

As shown in Figure 9, high-frequency voltage is applied between the upper and the lower electrode to generate discharge, which turns etching gas into plasma to etch only the desired surface portion of the wafer. The graphite is used as the cathode electrode in the upper part, and is shaped into a disk with a number of holes provided to blow out etching gas.

Graphite is used as the cathode electrode for the excellence of its good electrical conductivity, sputter resistance, gas-corrosion resistance, and the like. With
the cathode electrode positioned above the wafer, if particles like carbon atoms sputtered by the collision of machining powder and ions fall on a wafer, this will affect the circuit on the wafer adversely. For this reason, graphite is very rarely used alone, but it is used with SiC coating on it, or with its surface converted to SiC, or in the form of amorphous carbon (glassy carbon), and so on.

9. ION IMPLANTATION PROCESS

In manufacturing semiconductors from silicon, it is impurity elements, introduced in very small quantities into silicon (a fourth group element), of the third group, B, P, As, and others that allow the performance as semiconductors to be displayed.

In this process, the element introduced is evaporated and its vapor is hit by electrons and ionized, and then the ions generated in this way are drawn out by the force of the electric field, being further accelerated by the electric field to accumulate their energy, and finally they are shot into the target position on the wafer at a high speed.

Figure 10 schematically shows an ion implanter, in which several tens of high-purity graphite parts are used in the ion extraction electrodes, accelerating electrodes, shaping slits to adjust the beam shape, and the like.

10. THERMAL DIFFUSION PROCESS

This process also is intended to introduce impurity elements into a wafer, using the self-diffusion of atoms by heat.
Figure 11 shows the system schematically, in which graphite with its surface converted into SiC is used on the jig called a wafer boat used to vertically arrange wafers to be processed.

Although there are wafer boats made of quartz, those made of SiC-converted graphite are more resistant to high temperatures and allow the diffusion treatment to be carried out at higher temperatures, thereby providing the advantage of shortening the treatment time.

This concludes an explanation of carbon materials used in the silicon semiconductor manufacturing processes; we hope it would be a help for the readers to understand the roles of carbon.
Chapter 2.7

Isotropic Graphite for Electric Discharge Machining

Hiroaki Itami
Product Planning Department, Toyo Tanso Co., Ltd., 2181-2 Nakahime, Ohnohara-cho, Kanonji, Kagawa 769-1612, Japan

Chapter Outline

1. What is Electric Discharge Machining? 135
2. What is a Mold? 135
3. Choice Between Electric Discharge Machining and Cutting to Best Suit the Purpose 136
4. Principle of Electric Discharge Machining 137
5. Kinds of Electrode used in Electric Discharge Machining 137
6. Kinds of Graphite Electrode Material for Electrical Discharge Machining 139
7. High-performance Graphite 139
8. Concluding Remarks 140

1. WHAT IS ELECTRIC DISCHARGE MACHINING?

Electric discharge machining is a method of machining in which, as shown in Figure 1, when an electrode and a workpiece are brought closer to each other with pulse voltage being applied between them, dielectric breakdown occurring at a certain distance between them triggers discharge, whose high temperature and pressure in turn cause part of the workpiece to be melted and blown. Only if current flows through a workpiece, even one made of a hard metal that cannot be machined by cutting can be machined; this is what electric discharge machining can do.

Figure 2 shows a piece of graphite, the electrode, and the workpiece produced by transferring a shape by means of electric discharge from the electrode. This electric discharge technology is used to manufacture molds.

2. WHAT IS A MOLD?

Today, most of the products and their parts surrounding us such as automobiles, home appliances, and everyday items are manufactured by means of molds.

If you make and retain the shape of a product in a mold, simply closing the mold and pouring resin or molten metal into the mold enables one to make a great number of products of the same shape economically in a matter of a short time. Molds are now indispensable items to maintaining our comfortable life.

Molds are classified according to raw materials used to make products, manufacturing methods, and other factors. Table 1 shows the kinds of molds and methods of molding products.
3. CHOICE BETWEEN ELECTRIC DISCHARGE MACHINING AND CUTTING TO BEST SUIT THE PURPOSE

To make a mold, a desired shape is formed out of a metal lump in one of the two widely used methods that are described below:

1. Mechanical machining: Edged tools are used to cut metal directly.
2. Electric discharge machining: An electrode having the shape of a desired product is made, and electric discharge is generated between the electrode and the workpiece to melt the metal.

In the following, the criteria as to the choice between electric discharge machining and mechanical machining are described.

Metals with hardness of 55HRC (Rockwell hardness) or more are called difficult-to-machine materials because machining mold materials with hardness of 55HRC or more results in much wear in end mill cutting edges.

When a difficult-to-machine material is handled, a material with good machinability (graphite and copper, for example) is machined into the product shape, which in turn is used as the electrode installed on an electric discharge machine to carry out electric discharge machining (shape transfer).

In a case where the material is not difficult to machine, too, mechanical machining to depth requires the protrusion of the end mill to be longer, which increases the probability that vibration occurs during mechanical machining.

If \( \frac{L}{D} = \frac{\text{machining depth}}{\text{end mill diameter}} \) exceeds 10, mechanical machining will be accompanied by
the risk of vibration, and, to avoid this, electric discharge machining is chosen.

Figure 3 shows the criteria for choosing a machining method on the basis of the “hardness of mold material” and the “machining depth.”

4. PRINCIPLE OF ELECTRIC DISCHARGE MACHINING

1. Voltage increase: Pulsed high voltage is applied between the workpiece and the electrode.
2. Discharge column formation: When the electrode and the workpiece come closer to each other to a distance (several tens to hundreds microns), a discharge column, a kind of plasma, is formed.
3. Melting: Heat generated by discharge melts the electrode and the workpiece at the same time.
4. Evaporation: the surrounding working fluid is evaporated instantly.
5. Explosion: The pressure generated by the evaporation blows the melted electrode and workpiece away.
6. Cooling: The working fluid cools off the discharge point, with a hollow left on the workpiece surface.

With these phenomena, described in items 1 to 6, occurring several thousands to several tens of thousands of times per second, electric discharge machining proceeds. The quantity to be blown away is determined by the electric conditions.

Advancing machining steps to the closest point to the ultimate shape is called rough machining, and finalizing machining steps from the rough machining to the ultimate shape is called finish machining. In some cases, the electrical conditions are changed through a gradual reduction in discharge power in several steps.

Two important parameters determining the electrical conditions are the current value and the ON-time duration. In rough machining, in which the importance is placed on the machining speed, the current value is made large as shown in Figure 4(a), with a current waveform with a longer ON-time duration appearing. In finish machining, in which the importance is placed on the surface roughness, the current waveform is defined as one with a smaller current value and a shorter ON-time duration (Figure 5).

5. KINDS OF ELECTRODE USED IN ELECTRIC DISCHARGE MACHINING

Electrode materials used in electric discharge machining are graphite (isotropic graphite), copper, copper tungsten, and the like.

Copper is used in cases in which finer surface roughness is required. However, copper has a high bulk density of 8.9 Mg/m³, and, when the electrode area is large, the machining accuracy may be affected adversely as a result of the surpassing of the allowable suspension load of the electric discharge machine or of the bending of the electric discharge machine, at the time of jumping, caused by the electrode mass.

Copper is not suitable to high-speed electric discharge machining, because allowing large current to flow through the copper piece causes the discharge surface to soften due...
to heat and the surface becomes rough as a result of explosion.

With a bulk density of about 1.8–1.9 Mg/m$^3$, about one-fifth of that of copper, graphite is a light substance used for large electrodes (Figure 6). In addition, graphite has high strength at elevated temperatures, and the advantage of faster machining speed because it allows larger current to flow through itself than copper does. However, since holes called pores are present in graphite, graphite is inferior to copper in terms of threshold discharge surface roughness.

Extremely resistant to heat, copper tungsten is used in machining ultra-difficult-to-machine materials such as carbide. However, the extremely poor machinability makes it difficult to manufacture electrodes, and the price is high.

Table 2 shows the advantages of graphite as compared with copper in mechanical machining and electric discharge machining.

Having smaller cutting resistance, graphite is machined mechanically in a shorter time than copper. Mechanical machining of copper is accompanied by burrs and its removal needs much labor, but graphite is not accompanied by burrs and does not need the time to remove them.

After mechanical machining, graphite can be passed to electrical discharge machining, the immediately following process. In other words, graphite is the material indispensable to implement the automatic changeover from mechanical machining to electrical discharge machining.

In electrical discharge machining, the short time needed to start electrical discharge and the high machining speed allow the machining time to be reduced.

The coefficient of thermal expansion of graphite is as low as less than one-third of that of copper and its deformation due to heat in mechanical and electrical discharge machining is small; this contributes to high accuracy machining (Figure 7).

<table>
<thead>
<tr>
<th>Advantage of graphite in mechanical machining</th>
<th>Advantage of graphite in electric discharge machining</th>
</tr>
</thead>
<tbody>
<tr>
<td>No burrs occurring</td>
<td>Shorter time before the start of discharge</td>
</tr>
<tr>
<td>Low cutting resistance, hence easy to machine</td>
<td>Fast electric discharge machining</td>
</tr>
<tr>
<td>Lightweight, hence easy to handle</td>
<td>Lightweight, hence less burden on the main shaft</td>
</tr>
<tr>
<td>Integrated electrode construction possible</td>
<td></td>
</tr>
<tr>
<td>Small stress deformation</td>
<td>Small thermal deformation</td>
</tr>
</tbody>
</table>

FIGURE 6  Bulk Density. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 7  Coefficient of Thermal Expansion. For color version of this figure, the reader is referred to the online version of this book.
6. KINDS OF GRAPHITE ELECTRODE MATERIAL FOR ELECTRICAL DISCHARGE MACHINING

Graphite is black as ebony and all kinds of it look the same. However, differences in the parameters specified in manufacturing processes such as the grain size distribution of raw material, sintering temperature, and graphitization temperature yield kinds of graphite having general physical properties and discharge characteristics different from those of other kinds. Both materials suitable for rough machining accompanied by fast electrical discharge machining speed and materials suitable to machining with less electrode wear and finer discharge surface roughness are available; it is necessary to choose the graphite one uses depending on the purpose. Table 3 shows typical graphite materials for electrical discharge machining supplied by Toyo Tanso.

7. HIGH-PERFORMANCE GRAPHITE

The demand for high-speed machining coupled with high accuracy has become strong also in areas of industry that have been using copper electrodes, with the request for using graphite having increased. However, the electrical condition for electrical discharge machining in parts with smaller areas and finer surface roughness tends to lead to larger electrode wear with graphite electrodes.

For the reason mentioned above, a graphite electrode with satisfactory surface roughness and low electrode wear has been looked for eagerly.

In the past, a kind of graphite with low electrode wear was available, but although it exhibited excellent performance in electrical discharge machining, its performance in mechanical machinability with an electrode was poor.

Generally speaking, “mechanical machinability” and “electrical discharge machinability” of graphite are in a trade-off relationship, and satisfying both simultaneously is difficult. In manufacturing conventional high-grade graphite, mechanical machinability was traded off for electrical discharge machinability, that is, a decrease in the electrode wear as the top priority with the aim of controlling manufacturing process conditions to increase “flexural strength.” This sometimes posed problems of chipping during mechanical machining.

Against this background, the newly developed TTK material manufactured by Toyo Tanso is a product with inventions in materials design; the material is produced on the basis of a technique of homogenizing grain sizes in a small dimension as shown in Figure 8. As a result, a material was developed successfully that meets both mechanical machinability and electrical discharge machinability that had been in a trade-off relationship.

In developing this new material (Figure 9), we reduced and homogenized grain sizes so that the load on the end mill may be constant during mechanical machining; this leads to the realization of more accurate shapes (Figures 10 and 11).

---

**TABLE 3 Graphite Grade for EDM**

<table>
<thead>
<tr>
<th>Class</th>
<th>Grade</th>
<th>Recommendation</th>
<th>Bulk density Mg/m³</th>
<th>Hardness HSD</th>
<th>Specific resistivity microOhm·m</th>
<th>Flexural strength MPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium-Grade</td>
<td>ISEM-8</td>
<td>for Rough</td>
<td>1.78</td>
<td>65</td>
<td>13.5</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>ISO-63</td>
<td>for Rough and Finish</td>
<td>1.78</td>
<td>76</td>
<td>15.0</td>
<td>65</td>
</tr>
<tr>
<td>Excellent-Grade</td>
<td>TTK-5</td>
<td>for Finish</td>
<td>1.78</td>
<td>80</td>
<td>15.5</td>
<td>80</td>
</tr>
</tbody>
</table>

Source: (Toyo Tanso catalog)
In electrical discharge machining, a smoother electrode surface contributed to improving the surface roughness and waviness substantially.

Normally, however, the grain size may be reduced; the presence of air holes called pores is unavoidable because of the method of manufacturing graphite. For this reason, the new material cannot excel copper or copper tungsten in terms of the threshold discharge surface roughness. However, TTK material excels both of the conventional materials in that it allows ultra-precision and minute shapes to be realized.

When copper or copper tungsten is machined mechanically, cutting resistance causes thermal deformation. Graphite also is subject to thermal deformation; however, since its coefficient of thermal expansion is very small, one-third of that of copper, thermal deformation is slight, which allows high-speed feed and thereby reduces the machining time substantially.

In a typical case, a PCD (polycrystalline diamond) tool was used to machine a column with a diameter of 30 \( \mu \text{m} \).

When undergoing electrical discharge machining, copper and copper tungsten cannot maintain the shape as a result of the softening and thermal expansion due to the heat from discharge. On the other hand, highly resistant to heat and having a small coefficient of thermal expansion, graphite is capable of keeping the shape while being machined by means of electrical discharge.

8. CONCLUDING REMARKS

The current trend is reducing the number of electrical discharge machining processes. As described above, however, there are a number of shapes that cannot be
handled with mechanical machining, and electrical discharge machining is very excellent in terms of accuracy and machine service life.

In addition, there are also cases of machining in which shapes cannot be realized even by means of a copper or copper tungsten electrode. Graphite material is extremely suitable to such shapes and is compatible with shapes that have not been considered (high aspect ratio shapes).

The use of graphite material in electrical discharge machining of ultra-precision shapes promises a very great possibility.
1. BACKGROUND

Carbon fibers are the strongest and stiffest fibrous reinforcement known to date [1]. Carbon fibers were first used by Edison in 1870 as filaments for incandescent bulbs. Carbon fibers for structural applications became available starting only in the early 1960s and improvement of their properties took place steadily over the next three decades. Many grades possess a tensile strength of about 5 GPa, which make them almost five times stronger than steel. Other grades possess a thermal conductivity of 1,000 W/m·K, which is almost three times that of copper [2]. Coupled with their low density of 1.7–2.0 g/cm³, such fibers have found extensive use in high-performance aerospace and defense applications [3].

In the 1990s and later, significant efforts have been devoted to improving the production processes, which has led to not only improvement of properties but also a general reduction in their cost. Since then, carbon fibers have been used in industrial components, nuclear reactor components, rehabilitation of bridges and buildings [4], high-performance sporting goods, and packaging of electronics [5]. In addition to strength and stiffness, other carbon fiber characteristics that have led to above applications include [6]:

- high dimensional stability, low thermal expansibility, and low abrasion for missile cone and aircraft brakes;
- good vibration damping and outstanding stiffness for audio equipment, high-performance hi-fi equipment, and robotic arms;
- high electrical conductivity for EMI and RF shielding (electromagnetic shielding blocks radio frequency electromagnetic radiation); and
- chemical inertness and high corrosion resistance for chemical and nuclear industries.

Their outstanding strength to weight ratio makes carbon fibers ideal reinforcements for military and civilian aircraft applications. F-16 and Stealth bombers have been made of carbon–epoxy composites for almost twenty-five years now. Recently, Boeing 787 “Dreamliner” commercial airplane has been designed to use about 50% by weight of carbon fiber composites for its primary structure [7]. In land-based transportation applications, the high cost of carbon fibers (US$ 10–1500) somewhat limits its use to high-performance sporting goods (golf clubs and tennis rackets), and racing cars and high-performance automobiles such as BMW “E-series” electric vehicles [8]. Current, worldwide production of carbon fibers is estimated at 50,000 metric tons and the overall market is US$ 1.5 billion, which is expected to go up to about US$ 2 billion by 2015 [9,10].

Although commercial carbon fiber processing has been around for almost fifty years now, it is still not a mature technology due to the fact that carbon fibers are intractable. Therefore, thin fibers (nominally 10 μm) must be first obtained from an organic or hydrocarbon precursor. Then, the precursor fibers have to be converted to carbon fibers by thermal treatment to temperatures ranging from 1000 to 3000 °C [11]. The complex reactions lead to removal of noncarbon constituents, leaving behind carbon fibers. The chemical composition of the precursor plays a profound role on the microstructure and properties of the resulting carbon fibers. Therefore, the remainder of the chapter is divided into three main sections based on the precursors: (i) PAN-based, (ii) pitch-based, and (iii) rayon- and bio-based.
2. PAN-BASED CARBON FIBERS

2.1. Conventional PAN Precursors

Over 90% of commercial carbon fibers are produced from polyacrylonitrile (PAN) precursors [12,13]. The primary reason for this widespread use of PAN is the fact that PAN polymer does not melt, but can be dissolved in suitable solvents. Consequently, PAN precursors can be solution-spun into thin fibers that can be thermally processed into intractable fibers because PAN does not melt. It undergoes crosslinking and ladder-formation reactions before melting, and leads to the formation of intractable fibers, which can be further heat treated to temperatures in the 1000—1500 °C range to obtain carbon fibers [14].

A schematic diagram of the solution-spinning process is displayed in Figure 1. A solution of PAN (15—25 wt%) is first prepared using dimethyl sulfoxide (DMSO), dimethyl acetate (DMAc), dimethyl formamide (DMF), or another suitable solvent. The solution is wet-spun into a coagulation bath where the solvent diffuses out, leaving behind PAN fibers. Due to limited viscoelastic properties (strain hardening) displayed by PAN solutions, wet spinning cannot be performed at a fast rate or a large draw-down ratio. Also, the solvent must be allowed to diffuse slowly to prevent the formation of a hard skin that can lead to entrapment of solvent, which ultimately results in large cavities and voids in the fiber that result in a weak fiber. In any event, the out-diffusion of solvent leaves behind micropores that must be collapsed in a post-draw of the solidified PAN fibers.

As described in Figure 1, PAN-based carbon fiber can also be specified into three categories according to the heat treatment conditions and their properties: low-temperature heat-treated (300—1000 °C) type 1 fiber of low modulus and strength, intermediate-temperature heat-treated (1000—2000 °C) type 2 fiber of high-strength, and high—temperature-treated (2000—3000 °C) graphitized type 3 fiber of high modulus.

Next, the fibers are thermo-oxidatively reacted in a step, often called “stabilization,” to make them intractable and hold them together in such a way as to avoid extensive relaxation and chain scission during the final carbonization step. Stabilization is typically conducted in air at temperatures ranging from 200 to 300 °C, and leads to crosslinking and/or cyclization of PAN [15—18]. The cyclization is an exothermic reaction during which the pendant nitrile groups react with each other, transforming PAN backbone into a ladder-type structure, which is intractable. The reaction mechanisms for stabilization depend on the experimental conditions and type of copolymer used [19]. Numerous reactions can take place during heating of PAN, and many are still not well understood [20]. Some studies, such as of Burland and Parsons [21], showed that the first step of the stabilization is cyclization through reaction of the nitrile groups, with dehydrogenation being significant only above 300 °C. Others, such as Grassie and McGuchan [18], have proposed that dehydrogenation and cyclization reactions take place simultaneously, the former occurring both within the noncyclized polymer chain as well as within the condensed heterocyclic rings.

Cyclization reactions are highly exothermic and can lead to runaway reactions. So, this behavior can be considerably reduced if a comonomer such as methyl acrylate (MA), vinyl acetate (VA), or itaconic acid (IA) is introduced into the polymer chain. Also, the activation energy of the cyclization reaction is smaller for the copolymer, so the cyclization reaction sets in at a lower temperature and is completed over a broader range over a longer duration. Thus, the amount of comonomer in the
precursor affects the rate of oxidative stabilization [15,19], and thence the applied tension requirements [22].

At the molecular structure level, several studies have discussed the stereospecificity of the cyclization reaction [23–25]. It has been argued that the cyclization reaction should occur preferentially in isotactic sequences to form the ladder-like structure. Generally, it has been observed that intramolecular cyclization reactions occur at lower temperatures (175–230 °C) in the amorphous phase of the polymer [12,26]. The crystalline regions act as “bridge” points between the amorphous regions, holding the structure together. Above 320 °C, oxidation and intermolecular crosslinking take place, and that oxidative degradation reactions occur above 380 °C [12,26]. As noted earlier, a range of copolymer compositions are used to facilitate the dehydrogenation and ladder-formation reactions. This thermo-chemical conversion process is extremely complicated; so, precursor compositions and precise processing condition remain proprietary. However, this topic has also stimulated fundamental studies, which are largely cited in this chapter.

In addition to the chemical structure of PAN precursors, processing conditions play a key role in the transformation of organic precursors into carbon fibers and their microstructural characteristics. For instance, molecular orientation of PAN precursor significantly affects the properties of the polymer fibers, and orientation must be maintained as much as possible during stabilization if the final properties of the carbon fibers are to be maximized. Finally, heat treatment conditions play a pivotal role in obtaining fibers with a range of graphitic content and texture. “Graphitized” carbon fibers can be obtained from mesophase pitch precursors using heat treatment temperatures (HTT) of 3000 °C or higher, which yield a high modulus and high thermal conductivity. For high strength and intermediate modulus, HTT range of 1000–1500 °C is typically used for PAN precursors.

A representative scanning electron micrograph (SEM) of T300 carbon fiber is displayed in Figure 2. It is interesting to note that the precursor fibers and the resulting carbon fibers derived from PAN precursors are typically not circular, but are “kidney” shaped. The crease and crenulations along the length of the fibers are a consequence of the solvent out-diffusion during wet spinning.

A typical wide-angle X-ray diffractogram for T300 fibers is displayed next in Figure 3. The results were obtained from a Rigaku-MSC unit using an X-ray wavelength of 1.5406 Å (Cu target, 45 kV and 0.65 mA). The integrated azimuthal distribution is shown in Figure 3(a), which indicates a two-theta peak associated with (002) layer planes at about 25.4°. This is typical of
the “turbostratic” structure where the interplanar spacing is 0.344 nm, significantly greater than the 0.3354 nm found in crystalline graphite. Also, the registry of the planes is not the \textit{abab}... sequence found in graphite. Figure 3(b) shows the azimuthal distribution of the 002 planes; the full-width at half of the maximum height (FWHM) is almost 30°, typical of ordinary PAN-based carbon fibers. Later we will see that in highly oriented and graphitic carbon fibers, this distribution can be significantly narrower.

Typical mechanical properties of PAN-based carbon fibers are listed in Table 1. The largest producer of carbon fibers, Toray (based in Japan) and Cytec (based in US), both use fiber classifications such as T300. Historically, it represented a tenacity (tensile strength) of 300 ksi, which translated to about 2.1 GPa. With precursor and process improvements, T300 grade now possesses a tenacity of almost 3.7 GPa. The modulus of such fibers is typically 200–400 GPa, whereas the strain-to-failure is 1.5–2%.

Another mechanical property of great relevance in structural applications is the compressive strength of fibers, since composite structures can be subjected to compressive loads during actual use. Due to the microfibrillar orientation along the fiber axis, carbon fibers generally display much lower compressive strengths, typically less than about 2 GPa. It is also noted that the crease and surface defects present in carbon fibers cause the strength of such fibers to be much lower than the theoretical, in-plane strength of graphite that is predicted at almost 100 GPa, but one that has never been achieved.

As for transport properties of PAN-based carbon fibers, electrical resistivity is typically 16–20 \(\mu\Omega\)-m, whereas the thermal conductivity is about 10–20 W/m·K depending upon the HTT, with higher temperatures resulting in higher electrical and thermal conductivity. It should be noted, however, that the thermal conductivity of PAN-based carbon fibers is at least one order of magnitude lower than those derived from mesophase pitch (as will be discussed in the next major section).

### 2.2. Alternative PAN-based Precursors and Processing Routes

As noted earlier, the vast majority of commercial carbon fibers are produced from wet-spun polyacrylonitrile (PAN) precursor fibers [12,13]. Current research on PAN-based carbon fibers is directed towards either reducing the cost significantly or improving the properties greatly.

To get away from high costs associated with the wet-spinning process, the high-volume melt-spun PAN fiber route has been attempted by external plasticization of PAN and internal plasticization by making a copolymer of 10–15% comonomer content [27,28]. PAN-based precursor fibers, however, must be thermally stabilized to obtain an intractable structure before carbonization. Therefore, a suitable stabilization route is necessary for melt-spun PAN (before oxidation) to avoid melting of fibers.

In studies conducted at Clemson University, we have reported earlier that UV-assisted stabilization of PAN copolymer is an alternative; however, due to lack of UV sensitivity the process took prolonged stabilization time [30,31]. Subsequently, we used a PAN terpolymer containing a UV-sensitive monomer, acryloyl benzophenone (ABP) [32], which was synthesized by researchers at Virginia Tech. Fibers were melt-spun at 220°C and UV irradiated under a high-power (4.5 kW) mercury arc lamp for 1–5 min durations. After UV irradiation, the terpolymer was sufficiently crosslinked such that the fibers could be successfully oxidatively stabilized at 320 °C and then carbonized at 1500 °C. The modulus of the precursor fibers

<table>
<thead>
<tr>
<th>Classification</th>
<th>Carbon fiber</th>
<th>Modulus [GPa]</th>
<th>Tensile strength [GPa]</th>
<th>Electrical resistivity [(\mu\Omega)-m]</th>
<th>Thermal conductivity [W/m·K]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAN-based carbon fiber</td>
<td>T300</td>
<td>230</td>
<td>3.5</td>
<td>18.0</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>T650</td>
<td>255</td>
<td>4.3</td>
<td>15.2</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>T800</td>
<td>294</td>
<td>5.9</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>M35 J</td>
<td>343</td>
<td>4.7</td>
<td>11</td>
<td>45</td>
</tr>
<tr>
<td>Pitch-based carbon fiber</td>
<td>P25</td>
<td>160</td>
<td>1.56</td>
<td>12.1</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>P30</td>
<td>207</td>
<td>1.4</td>
<td>10.2</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>P55</td>
<td>220</td>
<td>0.9</td>
<td>9.4</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>P100</td>
<td>470</td>
<td>1.0</td>
<td>4.6</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>K1100</td>
<td>565</td>
<td>3.1</td>
<td>1.1–1.3</td>
<td>900–1000</td>
</tr>
</tbody>
</table>

**TABLE 1 Properties of PAN- and Pitch-based Carbon Fibers [1,43]**
was about 6 GPa and the tensile strength was about 230 MPa and strain-to-failure was about 16%. The carbonized fibers displayed tensile strengths and moduli as high as 730 MPa and 140 GPa, respectively.

Although the modulus of the carbonized fibers was acceptable, the tensile strengths were lower than those of commercial carbon fibers due to the presence of occlusions and other defects. Nevertheless, the study demonstrated that melt spinning, combined with UV-assisted stabilization, is a viable solvent-free route for producing PAN-based carbon fibers [32]. An interesting characteristic of these carbon fibers derived from melt-spun PAN terpolymer is that their two-theta peak associated with (002) layer planes in is observed at a low value of 24.9°, when typical T300 type carbon fibers display it at 25.6°. The low two-theta value indicates low graphitic crystallinity and suggests that such fibers may be used in high-temperature, insulation applications that require thermal stability.

Another recent study describes a continuous, plasma-based process for faster stabilization of precursor fibers [33]. PAN, pitch, or any other suitable organic/polymeric precursor fiber can be used as a feedstock in the microwave-plasma system. The authors expect about 20% reduction in processing costs.

Finally, in other studies, researchers are attempting to increase the properties of carbon fibers significantly. In a recent study [34], three different strategies were used simultaneously: a significant reduction in the diameter of carbon fibers, from 6 μm to 1 μm, was achieved using an island-in-sea bicomponent fiber spinning of a high-molecular-weight PAN precursor to which 1 wt% single-wall carbon nanotubes were added. The tensile strength was reported to be 4.5 GPa and modulus was over 450 GPa for the SWNT-modified fibers in comparison with 3.7 GPa tensile strength and 337 GPa modulus for pure PAN-based, ultrathin carbon fibers.

Ultimately, the translation of these enhanced fiber properties into properties of structural composites needs to be determined. If successful, such fibers can lead to next-generation composites that are either significantly stronger than the current ones (at a higher cost) or much less expensive than current composites (at similar property level).

3. PITCH-BASED CARBON FIBERS

3.1. Conventional Pitch Precursors

The second most common precursor in current use for production of structural carbon fibers is mesophase pitch (MP), a thermotropic liquid crystalline material. Thus, it can be melt-spun using processing temperatures above the softening point. Mesophase pitches are disk-like, polynuclear aromatic compounds [35]. MP-based carbon fibers have been recognized for their superior lattice-based properties, viz. tensile modulus and thermal conductivity [12,13,27].

Mesophase pitch is typically obtained from either thermal polymerization of isotropic pitches [36,37] or chemical synthesis of monomers such as naphthalene or anthracene [38]. Perhaps, the most easily processed MP pitch is one derived from polymerization of naphthalene. The reaction is catalyzed by a BF3 Lewis acid, and produces an MP pitch with a low softening point of about 270 °C due to a low molecular weight centered around 500–600 amu [39,40]. Corrosion issues associated with the catalysts lead to significant process costs, and such precursors have typically been sold for about US$ 50–100/kg. The other grade of MP pitches can be produced from coal-tar residues or from reaction of decant oils obtained during cracking of petroleum [41]. Coal-tar pitches are heavier in the polynuclear aromatics hydrocarbons (PAHs) and can lead to a much higher mesophase content; they also suffer from toxicity of the PAHs. Petroleum pitches offer a much safer and less toxic composition.

In a typical process, MP pitch is fed into continuous extruders fitted with gear pumps, as illustrated in the schematic of Figure 4. Extrusion temperatures typically range from 300 to 360 °C and pitch fibers are produced with diameters ranging from about 8 to 15 μm [42,43]. As with other precursors, these MP fibers are “stabilized” using oxidative crosslinking of the hydrocarbon. Aliphatic pendant groups attached to the discotic PAH facilitate this thermo-oxidative crosslinking reaction. The stabilized fibers are then carbonized in an inert environment to temperatures of ~ 1000 °C and ultimately graphitized at temperatures exceeding 3000 °C.

The properties of resulting carbon fibers are a consequence of the high degree of graphene plane orientation [28–31], which in turn is generated by flow-induced orientation of discotic mesophase precursor. Circular symmetry associated with round-hole spinnerets typically results in a radial structure of the carbon fiber [44,45] as displayed in Figure 5. As heat treatment temperature increases, stack height and plane orientation increase. Crystallites are axially well orientated in the outer region of the die where shear rates are high [46]. In contrast, crystallites in the fiber are less orientated due to low shear rates at the centerline of the circular die.

Figure 6(a) and (b) display wide-angle X-ray diffraction results for MP-based carbon fibers. Figure 6(a) indicates a two-theta peak associated with (002) planes at values between 26 and 26.4°. This corresponds to a large fraction of the (002) planes being stacked with an interplanar spacing of 0.3354 nm and the registry of the planes being *abab* sequence found in crystalline graphite. The highest two-theta value of about 26.4° for K1100 grade is consistent with the highest extent of interplanar packing and three-dimensional crystalline order resulting from an
extremely high HTT of 3000–3300 °C. For P25 grade MP-based carbon fibers, the two-theta value is about 26.1°, consistent with other lattice-dominated properties that indicate a lower HTT.

The (002) peak is very sharp for K1100 fibers with an FWHM of about 0.4°. For P25 fibers, the peak is less sharp with a higher FWHM of about 0.8°. These FWHM values of ~1° for MP-based fibers are significantly smaller than...
that of about 4.6° for PAN-based fibers. The stacking height, LC, for K1100, P25, and T300 is of the order of 200, 20, and 2 nm. This clearly illustrates the larger three-dimensional stacking developed in carbon fibers derived from disk-like MP precursors versus the chain-like PAN precursors. From the azimuthal orientation distribution of the (002) graphene layer planes within the fibers, displayed in Figure 6(b), the FWHM values were about 4° for K1100 and 14° for P25. These FWHM values for MP-based fibers are much smaller than those for PAN-based fibers (FWHM greater than 30° for T300).

Typical mechanical properties of commercial MP pitch-based carbon fibers are summarized in Table 1, in conjunction with those discussed earlier for PAN-based carbon fibers. The most significant difference is that the tensile moduli of MP-based fibers are significantly higher than those of PAN-based ones, but tensile strength and strain-to-failure values are lower. This trend is a consequence of the high degree of graphitic development, which enhances lattice properties such as the modulus. In contrast, strength is a defect-controlled property, governed by the “weakest link”. The formation of large crystallites within MP-based carbon fibers also results in significant weak areas or critical defects in the inter-granular region. This region is prone to rapid crack propagation and catastrophic failure within the fiber. Consequently, MP-based fibers display significantly lower strain-to-failure values of less than 1%. For very high crystalline development, such as for K1100, strain values are less than 0.5%. However, thermal conductivity, a lattice-controlled property, is highest in such fibers. For K1100, axial fiber conductivity is reported at about 1100 W/m·K, a value two orders of magnitude higher than that of 10 W/m·K for PAN-based carbon fibers. Thermal conductivity is the ease of phonon transfer within a material, which is greatly enhanced by the high graphitic crystallinity within MP-based carbon fibers.

Overall, fiber properties not only depend upon properties of precursor but also heavily depend upon the processing conditions, heat treatment temperature, holding time at that temperature, and the defect density generated during processing. For carbon fibers produced from three MP precursors (AR-MP from Mitsubishi Gas and Chemicals, experimental MP-1, and experimental MP-2), lattice-dominated properties are displayed in Table 2. Experimental MP-1 had a mesophase content of about 75% and displayed, as expected, the lowest modulus and highest electrical resistivity. From the Lavin—Issi correlation [2] proposed for mesophase pitch-based carbon fiber, \( k = \left[ 440,000/(\rho + 258) \right] - 295 \) where \( k \) [W/m·K] is the thermal conductivity and \( \rho \) [\( \mu \Omega \cdot \text{cm} \)] is the resistivity. The estimated thermal conductivity is about 250 W/m·K, lowest among the three. For experimental MP-2, under two different spinning conditions (draw-down ratios), thicker fibers display a higher predicted conductivity (400–500 W/m·K), and also display a “Pac-man” splitting of the fiber cross section as, illustrated in Figure 7. The hoop stresses generated during heat treatment and the highly anisotropic texture of the fiber section result in the splitting. Mochida suggested that the radial open wedge occurs due to the anisotropic manner of shrinkage along the circumference of the fiber in the outer area at high temperature [47]. The v-notches generated are also deleterious to fiber strain-to-failure values. However, the splitting of the fibers enables the graphene layers to pack and register more perfectly, which results in a higher modulus and conductivity.

Another feature of MP-based fibers that is different from the one for PAN-based fibers is their compressive strength. For MP-based carbon fibers with the highest tensile modulus and thermal conductivity, the compressive
strength is only about 0.7 GPa, a small fraction of its tensile strength of about 2 GPa, and also a similar low fraction compared with compressive strength of about 2 GPa for PAN-based carbon fibers. From a structural composite perspective, this is not desirable because most structures must endure not only tensile stress but also compressive and bending stresses.

3.2. Alternative Mesophase Pitch Precursors and Strategies

As discussed above, flow-induced orientation of discotic mesophase precursor typically results in a radial structure of the carbon fibers. However, this structure also results in flaw sensitivity, and reduces tensile and compressive strength of such fibers relative to those obtained from polyacrylonitrile precursors. Therefore, process-based control has been studied to determine its control of the microstructure of mesophase pitch [48–51]. Mochida was able to establish a strong correlation between microstructure and the selection of pitch feedstock and spinning temperature, observing a range of textures from planar radial to quasi-onion [48]. Hamada established that mesophase domain size correlated strongly with shear stresses which, in turn, depend directly upon viscosity and inversely upon diameter cubed [49]. Also, placement of a mixer directly before the spinning nozzle resulted in microstructures deviating from planar radial, namely random, onion, and quasi-onion textures [50]. Use of a fine mesh, effectively a static mixer, also cut domain sizes and produced a distorted radial microstructure [51].

In a composition-control-based study, we have demonstrated the modification of morphology of mesophase pitch-based carbon fibers by the incorporation of carbon nanotubes [52]. Carbon multiwall nanotubes (MWNTs) with various aspect ratios were used. “Carpet” type of MWNTs, obtained from the CVD process, provided a long aspect ratio of over 1000, whereas shorter MWNTs had an aspect ratio of about 30. A range of concentrations between 0.1 and 1.0 wt% MWNTs were melt-dispersed into molten

<table>
<thead>
<tr>
<th>Carbon fiber</th>
<th>Diameter [µm]</th>
<th>Apparent modulus [GPa]</th>
<th>Strain-to-failure [%]</th>
<th>Electrical resistivity [µΩ·m]</th>
<th>Thermal conductivity [W/m·K]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_828</td>
<td>10.2 ± 0.6</td>
<td>414 ± 186</td>
<td>0.3 ± 0.2</td>
<td>2.5</td>
<td>570</td>
</tr>
<tr>
<td>L1_806</td>
<td>10.6 ± 0.5</td>
<td>301 ± 33</td>
<td>0.4 ± 0.2</td>
<td>5.4</td>
<td>254</td>
</tr>
<tr>
<td>L2_828</td>
<td>9.4 ± 0.7</td>
<td>502 ± 66</td>
<td>0.3 ± 0.1</td>
<td>4.6</td>
<td>316</td>
</tr>
<tr>
<td>L2_444</td>
<td>17.2 ± 1.7</td>
<td>426 ± 101</td>
<td>0.2 ± 0.1</td>
<td>3.0</td>
<td>495</td>
</tr>
</tbody>
</table>

FIGURE 7  SEM micrographs of carbon fiber heat-treated at 2500 °C and derived from: (a) AR synthetic mesophase; (b) experimental petroleum pitch MP-1; (c) experimental petroleum pitch MP-2 melt-spun at a high draw-down ratio; and (d) MP-2 melt-spun at a low draw-down ratio.
AR-MP mesophase pitch. Breaking of filaments during extrusion was observed at MWNT concentrations of 0.5 and 1.0 wt%. However, at 0.1 and 0.3 wt%, the spinnability was good. Additionally, a batch of mesophase pitch was compounded without the addition of nanotubes (0 wt%) as a control, in order to verify that the processing step did not impact the finally fiber microstructure.

**Figure 8** displays SEM micrographs of the unmodified (0 wt%), 0.1 wt% and 0.3 wt% long aspect ratio, and 0.3 wt% short aspect ratio MWNT-modified carbon fibers in cross section. The unmodified fiber exhibits a strong radial orientation of the graphite planes around the fiber axis, which has resulted in the radial splitting typical of these types of fibers. Cross section of carbon fibers containing 0.1 and 0.3 wt% long aspect ratio MWNTs exhibited a random texture. The fiber containing 0.3 wt% short aspect ratio MWNTs appears to have more of a PanAm structure. Additionally, the graphite sheets of the nanomodified fiber show significantly more folding and shorter length of continuity, compared to the unmodified fiber. In conclusion, the addition of only 0.3 wt% MWNTs to the mesophase pitch precursor allowed for the suppression of radial splitting, which can adversely affect fiber mechanical properties. These results clearly indicate that carbon nanotubes homogenize or randomize the texture of the carbon fibers and prevent them from splitting. This nested texture also prevents the fiber from splitting.

For carbon fibers containing 0 wt% nanotubes (control samples), the compressive and tensile strengths were 0.5 and 2.3 GPa, for a ratio of about 0.2. For carbon fibers containing 0.1 wt% MWNTs, the compressive and tensile strengths were 1.0 and 2.5 GPa, for a ratio of about 0.4, whereas for carbon fibers containing 0.3 wt% MWNTs, the compressive and tensile strengths were 0.4 and 1.3 GPa, for a ratio of about 0.3 [53]. The estimated axial thermal conductivities for 0, 0.1, and 0.3 wt% MWNT-modified carbon fibers (all heat treated to only 2400 °C) were 600 ± 50, 700 ± 230, and 500 ± 30 W/m·K, with there being no statistically significant difference in these values. Thus, compressive strength was improved in the nanomodified fibers without sacrificing the axial thermal conductivity. It is hypothesized that the addition of nanotubes to mesophase pitch nucleates a larger number of smaller-size domains in the cross section, and resulted in different morphology and texture of the resulting fibers. However, along the longitudinal axis, the crystallite length is largely unaffected and helps in retention of axial thermal conductivity of the fiber.

Although synthetic AR-MP mesophase precursors are chemically most stable at spinning temperatures and can be thermo-oxidatively stabilized at moderate temperatures, the production of such grades has been suspended in recent years. Consequently, petroleum pitches will find more routine use in the future. However, like any naturally occurring substance, it is prone to compositional variation due to variability in petroleum composition obtained from different parts of the world. However, if suitable analytical techniques can be devised to characterize and standardize the raw materials, petroleum pitch is potentially a very inexpensive precursor that can significantly change the increased use of carbon fibers.

4. RAYON AND BIO-BASED PRECURSORS

As noted above, PAN and MP pitch precursors produce HCN and other toxic gases during oxidative stabilization.
and carbonization steps. Therefore, current research studies are investigating the replacement of PAN or pitch by environmentally friendly/sustainable precursor alternatives. Although numerous hydrocarbons can be converted to carbon, very few have been actually converted to structural carbon fibers. Besides PAN, mesophase pitch and rayon are the only other precursors that have produced carbon fibers. Mesophase pitch produces highly graphitic carbon fibers with high modulus, high thermal conductivity, but low strength. Rayon, on the other hand, uses naturally occurring cellulose, but its conversion to carbon is mechanistically limited to carbon fibers with poor strength and cannot be used in primary reinforcement applications.

Rayon fibers are produced from wet spinning of cellulose that is dissolved in CS₂ and NaOH [42,43]. Typically, rayon fibers do not have to be oxidized, but have to be stretched significantly during carbonization. Typically, the carbon fibers have a modulus of under 100 GPa and a tensile strength of under 1 GPa. The overall carbon yield from rayon precursors is only about 10–30%, when it is 40–50% for PAN, and about 70–75% for mesophase pitch. In contrast to the presence of significant polynuclear aromatic hydrocarbons in mesophase pitch that leads to a high carbon yield, the lack of aromatic structure within rayon results in this low carbon yield.

The lack of aromatic structure in rayon also leads to a low graphic crystallinity in rayon-based carbon fibers. As determined by wide-angle X-ray diffraction, integrated azimuthal profiles display a two-theta peak associated with the stacking of (002) layer planes at 24.5°, when PAN-based fibers display it at about 25.6°. This indicates a low degree of graphitic content in rayon-based carbon fibers, which in turn leads to a fairly low thermal conductivity, at least within the family of carbon fibers. The low thermal conductivity has led to their use as thermal insulation materials in ultra-high-temperature environment such as rocket nozzles. Rayon-based carbon fibers are typically used with a phenol–formaldehyde matrix as a composite in contact with hot exhaust gases. Such materials ultimately ablate, but due to their insulative properties protect the inside metal components from melting or degrading.

Even though rayon is derived from naturally occurring cellulose, ironically the chemical process used to make rayon utilizes solvents (xanthates) that are not environmentally friendly. Among other bio-derived feedstocks, lignin is an aromatic material that holds potential as a carbon precursor. Lignin encapsulates cellulose fibers in wood, and provides rigidity and strength to wood.

The use of lignin as a precursor for carbon fibers dates back to 1960s. Carbon fibers have been produced from a mixture of lignin and a polymer plasticizer dissolved in an alkaline solution [54]. Since then, carbon fibers have also been produced without a polymer additive. Different types of lignin used for the purpose include steam-exploded lignin [55], organosolv lignin [56,57], and hardwood kraft lignin [58]. More recent interest in lignin is derived from the fact that it possesses an aromatic structure, which facilitates conversion to carbon. However, lignin also possesses a three-dimensional molecular architecture, which makes it intractable in its unmodified state, and must be broken down into smaller-molecular-weight components. Softwood kraft lignin is commercially available, and has been chemically modified into an acetylated form that renders it melt processable. To date, lignin-based carbon fibers have displayed tensile strengths limited to about 1.6 GPa and tensile modulus limited to about 100 GPa, significantly lower than their MP-pitch or PAN-based counterparts [59].

In summary, the recent trend in the use of bio-based, sustainable materials has revived interest in the production of carbon fibers from lignin-derived precursors, which in turn are derived from wood and other biomass. Because lignin is abundantly available in the biosphere, it holds promise for producing cost-competitive carbon fibers, whose applications will extend to larger industrial and transportation sectors to achieve energy efficiency from the use of lightweight carbon-based composite materials.

ACKNOWLEDGMENTS

This material is based upon work partly supported by The Engineering Research Centers Program of National Science Foundation under NSF EEC-9731680, NSF 0823012, UTC/AFRL 08-S568-053-01-C1 / FA6550-05-D5807, and DOE-EPSCoR DE-FGO2-07ER46364. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Department of Energy, National Science Foundation, and other sponsors.

REFERENCES

Composite; March 2011.
graphitization in carbon fibers through exfoliation. Carbon 2004;
42:2567–72.
[12] Gupta A, Harrison IR. New aspects in the oxidative stabilization of
[13] Edie DD. The effect of processing on the structure and properties
[15] Grassie N, Hay JN, McNeill IC. Thermal coloration and insolubil-
[16] Takata T, Hiroi I, Taniyama M. Coloration in acrylonitrile poly-
polymers-III; thermal analysis of preheated polymers. Eur Polym J
[19] Fitzer E, Müller DJ. The influence of oxygen on the chemical
reactions during stabilization of PAN as carbon fiber precursor.
[22] Fitzer E, Müller DJ. Zur Bildung von gewinkelten leiterpolymeren
[23] Olivé GH, Olive S. Inter-versus intramolecular oligomerization of
Studies of the degradation of acrylonitrile/acylamide copolymers as
a function of composition and temperature. Carbon 1983;21:
[25] Chen SS, Herms J, Peebles LH, Uhlmann DR. Oxidative stabiliza-
[26] Gupta A, Harrison IR. New aspects in the oxidative stabilization of
[27] Grove D, Desai P, Abhiraman AS. Exploratory experiments in the
conversion of plasticized melt spun PAN-based precursors to
[28] Daumit GP, Ko YS, Slater CR, Venner JG, Young CC, Zwick MM.
[30] Paiva MC, Kotasthane P, Edie DD, Ogale AA. UV stabilization
route for melt-processible PAN-based carbon fibers. Polym Prepr
2003;44:651652.
[31] Naskar AK, Walker RA, Proulx S, Edie DD, Ogale AA. UV
assisted stabilization routes for carbon fiber precursors produced
from melt-processible polyacrylonitrile terpolymer. Carbon 2005;
produce carbon fiber via microwave assisted plasma processing.”
[33] Chae HG, Choi YH, Minus ML, Kumar S. Carbon nanotube
reinforced small diameter PAN based carbon fiber. Comp Sci
[34] Diefendorf RJ. Polymers for fibers and elastomers. Chapter 12.
pitch derived from isotropic anthracene pitch produced catalytically
The properties of co-oligomerized mesophase pitch from methyl-
naphthalene and naphthalene catalyzed by HF/BF3. Carbon
1998;36:369–75.
[37] Dumont M, Chollon G, Dourgies MA, Piailier B, Bourrat X,
Naslain R, Bruneel JL, Couzi M. Chemical, microstructural and
thermal analyses of a naphthalene-derived mesophase pitch. Carbon
2002;40:1475–86.
[38] Mochida IY, Korai Y, Ku CH, Watanabe F, Sakai Y. Chemistry of
synthesis, structure, preparation and application of aromatic-
[39] Kundu S, Ogale AA. Rheostructural studies on a synthetic
mesophase pitch during transient shear flow. Carbon 2006;44:
2224–35.
[40] Park YD, Mochida IY. A two-stage preparation of mesophase pitch
from the vacuum residue of FCC decant oil. Carbon 1989;27:
925–9.
[41] Fitzer E, Manocha LM. Carbon reinforcements and carbon /carbon
[42] Buckler JD, Edie DD. Carbon—carbon materials and composites.
[43] McHugh JJ, Edie DD. The orientation of mesophase pitch during
[44] Barnes AB, Dauche FM, Gallego NC, Fain CC, Thies MC. As-spin
orientation as an indication of graphitized properties of mesophase-
[45] Kundu S, Ogale AA. Microstructural effects on the dynamic
rheology of a discotic mesophase pitch. Rheol Acta 2007;46:
1211–22.
Microstructure of mesophase pitch-based carbon fiber and its
and physical properties of carbon fibers from coal tar mesophase
The properties of co-oligomerized mesophase pitch from methyl-
naphthalene and naphthalene catalyzed by HF/BF3. Carbon
1998;36:369–75.
pitch derived from isotropic anthracene pitch produced catalytically
The properties of co-oligomerized mesophase pitch from methyl-
naphthalene and naphthalene catalyzed by HF/BF3. Carbon
1998;36:369–75.


1. INTRODUCTION

Activated carbon fibers (ACFs) are porous carbons with a fiber shape and a well-defined porous structure which can be prepared with a high adsorption capacity. Although the ACFs are very promising materials, they do not still have a market as important as the activated carbons due to their difference in production costs. The main characteristics and advantages of the ACFs are the following [1–4]:

(i) They have both high apparent surface area and adsorption capacity. (ii) They have fiber shape with small diameters, typically between 10 and 40 μm, which is a very important characteristic for applications requiring higher packing density (i.e. gas storage) [3,5]. More recently, similar materials with much smaller diameters (in a nm scale) have been developed, as it will be commented later on. (iii) ACFs are light materials and can be easily woven into different fabrics (i.e. cloths, felts, etc.). (iv) The pore-size distribution of the ACFs can be narrow and uniform, being essentially microporous materials, although mesoporous ACFs can also be prepared. (v) The narrow diameter essentially eliminates mass transfer limitations, with the adsorption–desorption rates being very rapid.

Since the ACFs are fibrous materials that can be easily molded and woven, filters can be designed which do not have the settling and channeling problems of the conventional granular and powder activated carbons [1]. Due to their low hydrodynamic resistance, they can be used as thin cloths for the treatment of high flow of gases, which is very useful for control of gas-phase pollution [4].

The development of ACFs and activated carbon cloths is closely related to that of carbon fibers (CFs). This makes that the raw materials used for the preparation of ACFs be, chronologically, the same as for CFs. Thus, in 1966, viscose and acetate cloths were, like for CFs, the first materials used to obtain ACFs [6,7]. The low yield of the ACFs, and CFs, obtained from the above precursors, oriented the research toward the seeking of other raw materials for the preparation of cheaper CFs and ACFs and with higher yields. In this way, ACFs were prepared from 1970 using lignin (with the brand of Kayacarbon ALF), polyvinyl-chloride [8] (i.e. Saran polymer, already used to obtain activated carbons), and phenolic precursors [9]. The high yield and the good mechanical properties of the ACFs obtained make these precursors very useful for this application. In fact, Economy and Lin [10] developed ACFs from a phenol formaldehyde precursor which are commercialized since 1976 with the name of Novolak. In 1980, Kuray Chemical Co. Ltd. commercialized ACF from phenolic resin with the name of Kynol [11].

The preparation of PAN-based ACFs was initiated in 1976 by Toho Rayon Co. Ltd. [12] and the use of pitch to obtain ACFs started in 1985, and were commercialized by Osaka Gas Co. Ltd (Ad’all) [13]. Due to the low price of the pitch (petroleum and or coal tar) and high yield of the ACFs obtained, the manufacture of these pitch-based ACFs has increased considerably, nowadays being one of the main precursors of ACFs [1].

The research on ACFs is not different from what is usual for other activated carbon materials. It focuses on understanding the preparation process, on the characterization of the materials, and the analysis of their performance in given applications. A literature search on this topic gives us more
than eight hundred contributions, taking into account only the papers published in journals. The research on ACFs started more than thirty years ago regarding its preparation (as it has been detailed above). However, most of the work done concentrates mainly in the last twenty years and it is essentially focused on their characterization and applications. More recently, other forms of ACFs having diameters in a nm scale have been developed, such as activated carbon nanofibers (ACNFs) or activated carbon nanotubes (ACNTs) [5,14–20].

Thus, this chapter on ACFs will cover the most important aspects concerning the preparation routes and the most important applications of these materials (mainly ACFs and also ACNFs and ACNTs).

2. PREPARATION OF ACFs

Once the precursor (i.e. pitch, polymer, etc.) is transformed into a fiber shape, by a suitable spinning process and is carbonized after a proper stabilization stage, the activation of the resulting CFs is needed to increase its adsorption capacity.

The starting points for the activation of the CFs precursors are not different compared to those for conventional granular or powder carbon precursors. As it happens in all of them, the starting porosity of the carbon material to be activated has a relevant role in its subsequent activation. To prepare ACFs, the precursor and the method of preparation need to be conveniently selected. These two factors have great importance as they determine the final porous structure of the ACFs. For a given precursor, the main stage determining the porous structure is the method of activation.

The objective during the activation is both to increase the number of pores and to increase the size of the existing ones, so that the resulting porous carbon has a high adsorption capacity. The preparation of ACFs can be achieved by first selecting the appropriate precursor and controlling its pyrolysis step and then controlling their subsequent activation process by any of the following two methods: (i) using a reactant gas in the so-called physical activation and (ii) using a chemical agent in the so-called chemical activation.

The preparation of ACFs by physical activation includes a controlled gasification of the CFs at temperatures between 800 and 1000 °C with an oxidant gas, so that carbon atoms are being removed selectively. The removal of the outer and less ordered carbon atoms leads to the creation of micropores and/or the widening of their size, what results in an increase of its pore volume.

Thus, for a given precursor, the pore-size distribution in the ACFs depends on the preparation conditions (mainly temperature, time, and gas flow), the activating agent used, and the presence of catalysts. Some representative examples of the influence of the experimental conditions mentioned before can be found elsewhere [4,21–34]. In order to have an efficient activation process, the reaction must take place inside the carbon fibers, at least predominantly, compared with the reaction occurring outside. If only external reaction takes place, the obtained material does not develop porosity. However, if the reaction occurs inside the fiber, there is porosity development: the higher the amount of carbon removed, the higher the porosity development.

Carbon dioxide and steam are the activating agents most commonly used, which reactions with carbon are endothermic. These gases react with the carbon atoms in the precursor according to the following reactions:

\[ C + CO_2 \leftrightarrow 2CO \quad \Delta H = 159.0 \text{ kJ/mol} \]
\[ C + H_2O \leftrightarrow CO + H_2 \quad \Delta H = 118.5 \text{ kJ/mol} \]

Although the activation with carbon dioxide or steam produces essentially microporous ACFs, strong differences have been found between these two activating agents regarding the porous texture and the mechanical properties of the ACFs [21,32].

The addition of metals such as cobalt, silver, rare earth metals, or platinum either to the starting pitch (followed by spinning, stabilization, and carbonization) or to the carbon fiber, followed by gasification with steam, allows the preparation of ACFs with a significant content of mesoporosity [24,28–31,33,34].

The chemical activation process consists of mixing a carbonaceous precursor with a chemical activating agent, followed by a pyrolysis stage [35–39]. The material after this stage is rich in carbon content and presents a much ordered structure and, after the thermal treatment and the removal of the activating agent, has a well-developed porous structure. Different compounds can be used for the activation; among them, KOH, NaOH, H₃PO₄, and ZnCl₂ have been reported in the literature [26,27,35–39].

The chemical activation presents advantages over the physical one, that can be summarized as follows: (i) the chemical activation uses lower temperatures and pyrolysis time, (ii) it usually consists of one stage, (iii) the yields obtained are higher, (iv) it produces highly microporous activated carbons, and (v) it is a suitable method for applying it to materials with a high ash content [36–39]. On the other hand, the chemical activation presents disadvantages such as the need of a washing stage after the pyrolysis and the corrosiveness of the chemical agents used.

Although the work done on physical activation of carbon fibers is wide, the research on chemical activation of carbon fibers is scarce and is mainly restricted to the use of some alkaline hydroxides as activating agents [40,41]. The chemical activation must be done under well-controlled experimental conditions in order to avoid...
destruction of the fiber shape. The resulting ACFs are essentially microporous materials (i.e. pore size below 2 nm), although differences exist depending on the activating agent used and the starting CFs.

In the activation by metal hydroxides, the main variables affecting the final porous texture are hydroxide/carbon ratio, heating rate, temperature, and time of pyrolysis. Moreover, there are two additional parameters that have recently been reported [37–39]: nitrogen flow rate and the washing stage (washing with water or washing with hydrochloric acid), which have an important role in the porosity development.

An example of the porosity development upon activation of a given CF precursor (Donacarbo S-241, provided by Osaka Gas, Japan) can be seen in Fig. 1 and Table 1 [42]. The figure presents the N₂ adsorption isotherm of the ACFs obtained by KOH activation at 750 °C using three KOH/CFs ratios (Table 1 and Fig. 1a) and ACFs obtained by CO₂ activation at 825 °C at three activation times (Table 1 and Fig. 1b). It has to be pointed out that: (i) the starting CF has no adsorption of N₂ at 77 K due to well-known diffusion problems [43–47] and (ii) the CF precursor used develops very well its porosity upon activation. All the isotherms are of type I, typical of microporous solids in which most of the

![N₂ adsorption isotherms at 77 K of ACFs prepared by (a) KOH activation and (b) CO₂ activation.](image-url)
Pores have a size below 2 nm. At high activation yields, the porosity is mainly due to narrow micropores (<0.7 nm) as shown from CO₂ adsorption (Table 1), which implies that the resulting ACFs have narrow micropore-size distributions. At low activation yields, especially for CO₂ activation, the knee of the isotherms becomes wide, indicating the presence of supermicroporosity (between 0.7 and 2 nm) and some narrow mesoporosity (between 2 and 7 nm). Additionally, both Table 1 and Fig. 1 allow to see that KOH produces higher microporosity developments than CO₂ for a given activation yield.

Since the activation method used and the structure of the starting CF play an important role in the development of porosity, these two specific topics will be explained below in more detail.

### 2.1. Chemical versus Physical Activation

As seen in Fig. 1 and Table 1, there are important differences among physical and chemical activation regarding the porosity development and pore-size distribution. In general, higher porosity development is achieved by chemical activation with hydroxides and the pore-size distribution can be tailored by the selection of the appropriate activation conditions and chemical agent [38]. On the contrary, lower porosity developments are achieved by physical activation, being necessary higher burn-off degrees (lower yields) to achieve similar apparent surface areas than by chemical activation. Moreover, the pore-size distribution cannot be easily tailored for high porosity developments by physical activation. Figure 2 includes the pore-size distributions of ACFs prepared by physical activation with CO₂ at different activation degrees (Fig. 2a) and by chemical activation with KOH using different hydroxide/CF ratios (Fig. 2b). It can be confirmed that: (i) physical and chemical activation perform very well, (ii) in the case of the chemical activation, the porosity development increases with the hydroxide/CF ratio, and (iii) the samples prepared by physical activation show lower porosity developments and wider pore-size distributions than those prepared by KOH.

Table 2 presents an example of our comparative activation study carried out with CO₂, KOH, and NaOH. The Table compiles the characteristics data of three selected AFC samples having similar apparent BET surface area.
(between 2400 and 2500 m²/g) [38,41]. Although these three activating agents activate well the CFs used, the physical activation gives an activation yield which is too low from an industrial point of view (6% which is about eight times lower than the 47% for KOH). Additionally, SEM examination of samples of Table 2 shows that ACFs prepared by hydroxide activation preserve well the fiber morphology (fiber surface is not damaged) even at these high hydroxide/CF ratios (8:1 for KOH and 6:1 for NaOH). Contrarily, fiber surface damages can be observed upon CO₂ activation [41].

Combining different techniques, the distribution of porosity of single fibers of ACFs prepared by different methods has been analyzed [21,48–50], and important differences have been observed that provide very interesting information on the activation mechanism. Comparing CO₂ and steam activation results, we find that CO₂ activation takes place within the fibers, generating a quite homogeneous development of porosity and, in the case of steam activation, the development of porosity focuses on the outer parts of the fibers. This, obviously, results in important differences in the mechanical properties, being those of larger magnitude in the steam ACFs [21]. Additionally, a similar study with NaOH and KOH [51] shows, for both activating agents, that there is a high concentration of pores even in the inner regions of the fibers which means that they penetrate better than CO₂ and steam into the bulk of the fibers.

### 2.2. Effect of Carbon Fiber Crystallinity

The structural order of the carbon fibers has a strong relevance on the development of porosity that can be achieved by whatever activation method. Thus, materials with a low structural order, such as isotropic carbon fibers, can be easily activated, by either physical or chemical activation, producing ACFs with a good development of porosity, large apparent surface areas (above 3000 m²/g), and high porosity [5,21,41,52]. Some examples that reflect this statement can be those already presented in the previous sections. The situation is rather different when materials with higher structural order are used as precursors of ACFs. For example, high-performance carbon fibers (such as Hexcel Co. carbon fibers) cannot be activated by the so-called physical activation. The reactions of both CO₂ and steam with the CFs produce the removal of carbon atoms from the external surface of the fiber without the creation of a significant amount of porosity and, consequently, with a negligible increase in surface area [52]. The scenario is different when chemical activation with hydroxides is applied. In this case, the activation of high structural order carbon fibers (e.g. Hexcel or VGCFs) can be carried out by either KOH or NaOH (of course, with much more difficulties than with isotropic fibers). An acceptable porosity development (e.g. surface areas of about 900 m²/g) with activation yields close to 50 wt% can be obtained [52]. Additionally, in these fiber precursors, important differences have been observed during the activation process between NaOH and KOH. The activation with KOH is more efficient confirming its advantage over NaOH when materials having higher structural order have to be activated [20,38,52]. Obviously, this is a consequence of the activation mechanism participating in the hydroxide activation process.

During such activation process, a main redox reaction between the alkali-metal hydroxide and carbon takes place producing K or Na metals [53,54]. The metal formed in the reaction plays a secondary, but important, role in the activation due to its intercalation into the graphene layers of carbon materials having structural order [20,38]. Metal intercalation increases the graphene interlayer distance favoring, hence, the activation process and the porosity development. Such activation enhancement is much more effective for K than Na due to its easier intercalation.

A confirmation to the effect of CF crystallinity on chemical activation by hydroxides was obtained by microsmall angle X-ray scattering (μ-SAXS) [51]. The μ-SAXS experimental set-up makes it possible to analyze different regions of a single fiber across its diameter and to distinguish the structural features already existing in the raw fibers or being created during the activation process. The results showed that, depending on the precursor, the chemical activation process produces isotropic or anisotropic development of porosity. It was observed that chemically ACFs prepared from isotropic carbon fibers present an isotropic development of the porosity and that

<table>
<thead>
<tr>
<th>Sample</th>
<th>Act. temperature</th>
<th>S_{BET}/m²/g</th>
<th>V_{DR(N2)}/cm³/g</th>
<th>Yield/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOH 8:1</td>
<td>750 °C, 1 h</td>
<td>2420</td>
<td>0.94</td>
<td>47</td>
</tr>
<tr>
<td>NaOH 6:1</td>
<td>750 °C, 1 h</td>
<td>2541</td>
<td>0.96</td>
<td>37</td>
</tr>
<tr>
<td>CO₂</td>
<td>890 °C, 23 h</td>
<td>2487</td>
<td>0.86</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE 2 Activation Conditions, Porous Texture Data, and Yield of ACFs Prepared by KOH, NaOH, and CO₂ Activations**
a high micropore volume is developed not only in the external region of the fiber, but also in the core. On the other hand, in the case of anisotropic PAN-based carbon fibers, the existence of two regions with different structure was detected by $\mu$-SAXS measurements across the fiber diameter: an anisotropic external ring and a more isotropic fiber core [51]. The results showed that these two regions remain after chemical activation and that the activating agents are reaching the fiber core. It seems that the more isotropic the fiber core is, the easier will be the activation by NaOH, in relation to KOH activation.

In summary, as it happens with granular and powder carbons, a proper selection of the carbon fiber, the activation method, and the experimental conditions permits the preparation of ACFs with a tailored pore-size distribution, with the additional advantage of its fiber shape and small diameters which allow faster mass transfer rates compared to conventional activated carbons.

### 3. ACTIVATED CARBON NANOFIBERS (ACNFs) AND ACTIVATED MULTIWALLED CARBON NANOTUBES (AMWCNTs)

As commented previously, porous carbon materials can have different morphologies (e.g. powder, pellets, fibers, etc.). Activated carbon fibers (ACFs), a type of activated carbons with fibrous morphology and diameter around 10 $\mu$m [1], offer a number of advantages over conventional activated carbons due to their morphology and fiber diameter [4]. Much thinner fibrillar carbon materials are available such as carbon nanofibers [55–58], with diameters between 10 nm and <1 $\mu$m (CNFs) and multiwalled carbon nanotubes of about 10–20-nm diameter (MWCNTs).

There are different types of CNFs [56,59,60], most of them being prepared by catalytic chemical vapor deposition using hydrocarbon as carbon source [61,62]. These CNFs have interesting applications due to their nanosized diameters [56,59,60]. Unfortunately, they hardly can be activated because of their crystal ordering that makes, as commented above, difficult a suitable porosity development [5,52], as it also happens with the activation of MWCNTs [20].

Fortunately, other CNFs with lower crystal ordering have been developed using different preparation processes, for example, by a polymer blend technique [57,58,63] or using an electrosprinning method [64] or by CVD on the inner walls of the nanosized channels of an anodic aluminum oxide film, namely the template carbonization technique [65,66].

Although the activation process of both CNFs and MWCNTs has not been much studied, remarkable activation results have been reported [5,14–20] Additionally, interesting advantages over conventional ACFs due to their nanosized diameters (100–200 nm versus 10–20 $\mu$m) have been reported in areas such as in EDLC [14,15] or in hydrogen storage [5], as it will be presented later on.

Next, we will focus on our results obtained about the activation process of MWCNTs and CNFs (obtained by the polymer blend technique) with the aim of preparing high microporosity and high-surface area nanosized fibrillar materials.

### 3.1. CNF and MWCNT Preparation

The carbon nanofiber preparation was previously reported [5,14,57,63]. In summary, they are obtained by (i) blending and homogenizing a novolac-type phenol–formaldehyde polymer resin (carbon precursor polymer) and a high-density polyethylene (pyrolyzing polymer that produces no carbon residue), (ii) spinning them, allowing the micro-beads of phenolic resin in the polyethylene matrix be elongated to form long thin blended polymer fibers, and (iii) carbonizing these blended polymers under N$_2$ atmosphere to obtain bundles of several hundreds of carbon fibers with nanometer-size diameters. As an example, Fig. 3a shows the SEM image of the pristine CNFs [5]. It can be seen that it is formed by bundles (10–30 $\mu$m) of hundreds of thin carbon nanofibers with nanometric diameters (around 100–200 nm), much thinner than “conventional” carbon fiber diameters (around 10 $\mu$m). Additionally, as it was previously reported [5,14,15], these CNFs do not have a crystal, or well-oriented, carbon structure and hence they can easily be activated.

High-purity MWNTs with different degrees of graphitic organization were prepared by acetylene decomposition on a cobalt-based catalyst at different temperatures (450, 500, and 600 $^\circ$C) [20]. After synthesis, the catalyst precursor is eliminated by dissolution in concentrated HCl.

### 3.2. CNF and MWCNT Activations

The activation of the polymer-blended CNFs can be done either by chemical activation with hydroxides (NaOH or KOH [5,14]) or by physical activation (e.g. steam [15] or CO$_2$ [67]). However, the activation of MWCNTs can only be carried out by chemical activation with hydroxides [20] because physical activation usually destroys the nanotube morphology.

**Chemical hydroxide activation:** lentils of NaOH or KOH are mixed with the CNFs [5,14] or with MWCNTs [20]. Different activating agent/carbon material ratios can be used. The CNFs (or the MWCNTs) and the activating agent mixture are carbonized under a N$_2$ flow at a given temperature (usually close to 1000 K and 1 h of soaking time). Later on, the samples are washed with HCl (5 M) and
distilled water until they are free of chloride ions. After the washing step, the samples are dried and finally weighed to calculate the activation yields.

Figure 3b shows the SEM image of the ACNFs prepared with NaOH. It is important to point out that after activation by NaOH or by KOH, the fibrous morphology is maintained [5].

Physical activation: physical activation of these CNFs (not possible for CNTs) can be performed by steam [15] or by CO2 [67]. Both are good activating agents, allowing to maintain their fibrous morphology; however, for a given porosity development, the yields obtained with CO2 activation are higher than those of steam [67].

3.3. Activation Results

A good example of the ease activation of the CNFs obtained by the polymer blend technique is shown in Fig. 4a that compiles the N2 adsorption isotherms at 77 K of the pristine CNFs and the resulting ACNFs. The activation results, obtained using different KOH/CNFs ratios (0.5/1, 1/1, 2/1, 3/1, and 4/1), allow to point out that: (i) the pristine CNFs have an important initial microporosity that develops very well upon activation, (ii) the isotherms’ shape (all are of type I) indicate that all the samples are microporous, (iii) the hydroxide/CNF ratio has a strong influence on the porosity development, confirming previous results [37–39], and (iv) similar results (not shown here) are obtained when these CNFs are activated with NaOH [5].

An additional example of their ease activation, in this case by physical activation, can be seen in Fig. 4b and Table 3 [67]. The figure compiles the N2 adsorption isotherms at 77 K of the pristine CNFs and the resulting ACNFs prepared using CO2 activation at three activation times (12, 24, and 36 h) and hence different burn-off degrees. Table 3 summarizes the results obtained. Interestingly, the pristine CNFs already have a noticeable porosity that can further reach, upon activation, apparent BET surface areas higher than 2000 m²/g [5,14,67]; confirming the suitability of these CNFs for preparing ACNFs.

In the case of the MWCNT activation [20], it can be summarized that it cannot be done by physical activation but it can be carried out either with NaOH or with KOH. The activation results depend on the synthesis temperature of MWNTs, increasing as the synthesis temperature decreases, as does the structural order of their graphene layers. The highest BET surface area (1670 m²/g) is obtained for KOH activation using the most disordered MWNTs [20].

4. SOME EXAMPLES ON ACF APPLICATIONS

There are a large number of applications in which the performance of the ACF has been extensively analyzed. They range from conventional gas- and liquid-phase adsorption to antibacterial and energy storage uses. Anon, we will mention some of them and, later, we will explain in some more detail relevant and timely applications:

1. Pollutant removal in gas and liquid phases: the use of ACFs for pollutant removal in gas and liquid phases has been extensively studied since it is one of the main uses of activated carbons (ACs). It will be explained later in more detail.

2. Carbon molecular sieves (CMS): CFs and ACFs have been studied for the preparation and use as CMS for the separation of gases such as CO2/CH4 and N2/O2 [45,68–70]. It was shown that the narrow microporosity of the nonactivated CFs can be useful for this application and that the porosity of the ACFs can be tailored by cracking of hydrocarbons to develop good-quality CMS.
**Figure 4** N$_2$ adsorption isotherms at 77 K of ACNFs prepared by (a) KOH activation and (b) CO$_2$ activation.

**Table 3** Porous Texture Data and Yield of CNFs and ACNFs Prepared by CO$_2$ Activation

<table>
<thead>
<tr>
<th>Sample</th>
<th>$S_{BET}/m^2/g$</th>
<th>$V_{DR}(N_2)/cm^3/g$</th>
<th>$V_{DR}(CO_2)/cm^3/g$</th>
<th>Yield/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNF</td>
<td>520</td>
<td>0.22</td>
<td>0.26</td>
<td>—</td>
</tr>
<tr>
<td>CNF-12</td>
<td>1240</td>
<td>0.60</td>
<td>0.53</td>
<td>59</td>
</tr>
<tr>
<td>CNF-24</td>
<td>1910</td>
<td>0.87</td>
<td>0.59</td>
<td>30</td>
</tr>
<tr>
<td>CNF-36</td>
<td>2260</td>
<td>0.91</td>
<td>0.57</td>
<td>13</td>
</tr>
</tbody>
</table>
3. **Antibacterial use of ACFs**: related to water treatment, Oya and co-workers developed for the first time antibacterial ACFs consisting of Ag-containing samples [28–31].

4. **ACFs as catalyst and catalyst support**: there are interesting examples on the use of ACFs as catalyst and catalyst support [71–74]. The use of ACFs for this purpose is a consequence of the application of carbon materials in catalysis for decades [75]. The versatility of the carbon materials (i.e. porosity and surface chemistry), together with the special properties of the ACFs, makes these materials rather interesting and the number of papers on this subject has increased considerably since 1997. An interesting example is the oxidation of SO$_2$ to H$_2$SO$_4$ by O$_2$ and in the presence of water [73,74]. ACFs as catalyst supports of different metal catalysts were studied analyzing the influence of the catalyst dispersion and activity (i.e. refs. [71,72]).

5. **Gas storage (CH$_4$ and H$_2$)**: the storage of gases such as methane and hydrogen is of special interest for their use as fuel in conventional applications. It will be explained later in more detail.

6. **Supercapacitors**: the use of ACFs (mesoporous and microporous ACFs) as supercapacitors is also a subject of a strong research because of the applications of these devices in battery or fuel cell-based systems. It will also be explained later in more detail.

### 4.1. Pollutant Removal

The advantage of the ACFs compared to the conventional ACs for this purpose is mainly a consequence of their porous structure. As we have already mentioned before, the microporous ACFs essentially contain microporosity which is readily available to the adsorbate and the mass transfer limitations due to the diffusion within meso- and macroporosity are absent [1,4], making the adsorption process faster. Moreover, ACFs avoid the settling and channeling effects of the ACs [1,4]. Consequently, a large number of works on the use of ACFs for gas-phase adsorption can be found. As an example, we can mention the adsorption of gases such as SO$_2$ [76,77], NO$_x$ [78,79] and volatile organic compounds (VOC) [4,80]. The specific case of removal of benzene and toluene present in gas streams at low concentrations (200 ppm) has been little studied; however, high adsorption capacities have recently been obtained reaching 31 g of benzene and 53 g of toluene per 100 g of ACFs [81]. Liquid-phase adsorption on ACFs has similar advantages as gas-phase adsorption and their use for drinking water treatment and removal of organic compounds is well documented [4,80,82,83]. Essentially, the main advantage is the faster adsorption rate compared to the ACs, although the narrower pore-size distribution has also an important effect on the performance.

### 4.2. Gas Storage

The storage of gases such as hydrogen, methane (natural gas; NG), and carbon dioxide has a special interest from an environmental point of view. The former two having considerable advantages for their potential use as fuel in conventional applications and the latter for CO$_2$ capture (separation of carbon dioxide from other gas emissions) and CO$_2$ truck transport for its final storage, for example, in a geological reservoir.

In the case of natural gas and hydrogen, their uses as an automotive fuel present the advantage of their reduced emissions relative to gasoline. However, the main limitation is the problem of their storage on a limited onboard tank volume. Advanced storage units with sufficiently high (volumetric and gravimetric) energy storage capacity need to be developed, especially in the case of hydrogen. Among the possible alternatives for storing hydrogen [5,84–86], the use of high-pressure adsorption on microporous solids, as it also happens in the case of the adsorbed methane storage (ANG), seems to be a suitable way. Hence, such a method, which takes profit of the potential gas adsorption enhancement in micropores [87–92], is the subject of a strong research effort to develop high-tech microporous materials [38], such as microporous carbons [38,85], zeolites [93,94], MOFs, and COFs [95]. Among the available microporous materials, microporous carbons are among those with the best performance [38,96,97], with ACFs and ACNFs also being good candidates [5,26,67,85]. The advantage of the ACFs compared to AC is their essentially microporous character and higher packing density [3,26,98,99], although the higher price of the ACFs is detrimental.

**Adsorbed Natural Gas (ANG) storage.** Due to the favorable physical properties of methane in relation to hydrogen, its storage is easier and the research is much more advanced, reaching, at room temperature (RT) and pressures lower than 4 MPa, comparable storage values to that of CNG at 250 bar [100]. Nevertheless, for onboard storage applications, additional constraints are still placed on developing further the carbon adsorbents to increase the storage capacity. Currently, the use of carbon adsorbents, especially ACFs [3,98,99], has allowed to store considerable amounts of natural gas at RT and at pressures lower than 4 MPa [101,102], thus making its storage more attractive than hydrogen for practical use.

An overview of methane storage has been reported [99], analyzing a large variety of commercial and laboratory-prepared samples, including ACs, ACFs, activated carbon monoliths, and activated carbon cloths and felts. From this it is clear that to enhance methane adsorption capacity by unit volume of adsorbent, we have to control adsorbent properties such as porous texture, packing density, and pore-size distribution. Interestingly, the well-accepted
relationship between the gravimetric methane storage capacity (of a given AC) and its adsorption capacity (e.g. apparent BET surface) ends for samples having apparent BET surface areas above 2000 m²/g. In that case, the micropore-size distribution controls the storage of methane and, in terms of volumetric methane uptakes, ACFs have advantages. They have higher packing density than powdered activated carbons, reaching quite high volumetric methane uptake (166 v/v) and delivery (145 v/v) [3,98,99].

Hydrogen storage by high-pressure adsorption. For a given microporous adsorbent, two main alternatives can be used for storing hydrogen by high-pressure adsorption: (i) RT and very high pressures (from 10 MPa to 70 MPa) and/or (ii) low temperatures (cryogenic temperatures) at moderate pressures (<4 MPa). Of course, the type of adsorbent plays the most relevant role. Among the above-mentioned sorbent types (carbon-based materials, zeolites, metal-organic frameworks (MOFs and COFs), and others), carbon-based materials have received considerable attention [5,85,103–110].

In general, hydrogen storage at RT is, independently of the type of sorbent used, not so high (usually less than 3 wt % even at high pressures (10–30 MPa)). Up to 20 MPa, the most important factor is the narrow micropore volume (<0.7 nm) of the carbon sorbent, whereas at higher pressures (above 50 MPa) the hydrogen adsorption depends on the surface area and the total micropore volume [26,27,85,96,105]. The achievements are more promising at low temperature (e.g. 77 K, even at moderate pressures (1–4 MPa)) where the hydrogen adsorption capacity depends on both the micropore volume and the micropore size distribution [26,85,96,105]. Thus, values as high as 8 wt% and total volumetric hydrogen storage higher than the European target for onboard hydrogen storage systems (33 g H₂/L) have been reported [26,27,85,96,105] Hence, carbon materials prove to be good candidates for hydrogen storage with their advantages being, in relation to other microporous materials, packing density, low cost, wide diversities of bulk and pore structures, and good chemical stability.

As in the case of ANG storage, the density of the carbon adsorbent is important in terms of improving volumetric gas storage capacity. Additionally, two other variables are important: the development of high micropore volumes and the control of the pore-size distributions. Depending on the temperature and pressure used for storing hydrogen, one variable can be more important than the others [26,27,85,96,105].

In relation to ACNFs, it has recently been reported [5,67] that these nanosized activated carbons have interesting H₂ storage capacities at 77 K. Comparing H₂ storage on ACNFs and other microporous solids, it can be seen that, for a comparable adsorption porosity, the H₂ uptakes of the ACNFs are among the best ones [5]. Thus, ACNFs having a micropore volume of 0.69 cm³/g can adsorb 2.38 wt% of hydrogen at 77 K and 0.1 MPa. To achieve such value with other ACs and ACFs, they should have much higher micropore volumes (around 1 cm³/g) [85,96]. Furthermore, in a recent paper dealing with the synthesis of metal-organic frameworks (MOFs) and their hydrogen storage at 0.1 MPa and 77 K, a quite high hydrogen storage value has been reported (2.55 wt%) [111]. The authors indicate that such hydrogen value is among the highest reported for MOFs and they conclude that MOF samples are particularly suitable for gas storage. Comparing these values with those achieved by the ACNFs, one can see that they present similar hydrogen adsorption capacities, having their advantages such as easier preparation process, handling, and greater thermal and chemical stability. Additionally, in terms of a volumetric H₂ storage, due to their nanosized diameters they have much higher packing densities, and hence much higher H₂ storage capacity [5].

4.3. Supercapacitors

Electric double-layer capacitor (EDLC) is an increasingly popular energy storage device due to its energy and power density bridging the gap between batteries and classical electric capacitors. One of the applications of EDLC is as a second power system in electric vehicles. This application requires EDLC to have a small energy loss and a high energy density to improve the efficiency of electric vehicles. The energy storage mechanism in electric double-layer capacitors (EDLCs) is based on an electrostatic attraction between charges along the double layer formed at the electrode/electrolyte interface. Since this phenomenon is controlled by the surface area of the interface, activated carbons are the most widely used electrode materials for EDLCs [112–120]. The performance of porous carbons as electrical double-layer capacitors (EDLCs) is strongly dependent on a complex combination of properties which include carbon material properties, preparation method of the electrode, and electrolyte characteristics. Regarding the carbon material itself, porous texture (i.e. surface area, pore volume, and pore-size distribution), pore structure, and surface chemistry are the main properties that determine its performance. ACFs share the same properties as other porous carbons, which means that all of them can obviously be extended to this specific kind of materials. The use of ACFs as electrodes in supercapacitors has been reported elsewhere [121]. The advantage of ACFs compared to other materials is their pore structure which allows a fast mass transfer and, thus, a higher rate for the double-layer formation. This will be established at the end of this section.

Concerning porous texture, it is well known that, in general, the larger the surface area that an EDLC can
provide for adsorption of ions on electrodes, the more energy can be stored in the EDLC [115,116,119]. However, detailed analyses using porous carbon with tailored porosity have demonstrated that not all the pores are effective in the charge accumulation. For example, it has been shown that the very narrow micropores do not contribute to the total double-layer capacitance because of a molecular sieving effect [112–114,117,120,122]. Moreover, it has been observed that the efficiency of pore filling, i.e. of double-layer formation, is optimal when the pore size is around 0.7 nm in aqueous media and 0.8 nm in organic electrolyte [123,124]. Most of these studies have been done using low currents or low scan rates. However, from an application point of view, it is very important to know how the materials behave at high charge/discharge rates, since the accessibility of the porosity will depend also on the process rate. It has been observed that, for a porous electrode, the ohmic dissipation of energy varies down the pore (there is a continuous increase of electrolytic resistance down the pore from its entrance) [125]. Thus, it is very important to study the kinetics of the double-layer formation using porous materials to guarantee that they can be used in a high-power application.

Regarding surface chemistry, it has been shown [126–128] that the surface oxygen groups have, at least, two relevant contributions to the total capacitance of porous carbons. On the one hand, the surface oxygen groups (or the dangling bonds created after decomposition, in inert atmosphere or in a H2 atmosphere) determine the wettability of aqueous electrolyte; thus, their presence on the carbon surface is essential to take profit of the large double-layer contribution to the capacitance of the porous carbons which is associated to their high porosity. On the other hand, the surface oxygen groups have an additional contribution to the capacitance through faradic processes, especially in acidic conditions, which involve one or two electron transfer reactions. However, the surface oxygen functionalities of phenolic or carbonyl group can show pseudo-capacitance based on redox reaction even in organic electrolyte such as TEABF4/PC, since quinone derivatives are electrochemically active in aprotic organic media [129,130].

The pore structure plays a relevant role in the kinetics of the double-layer formation. In previous studies done with porous carbons about the kinetics of double-layer formation in aqueous electrolytes [131], a clear linear relationship was observed between the capacitance loss measured by impedance spectroscopy and the content in narrow microporosity in Na2SO4 electrolyte (Fig. 5); that is, the higher the percentage of narrow microporosity, the higher the capacitance loss. This means that from a kinetic point of view, porous carbon materials with a wide mean pore size have better performance, i.e. the porosity is more easily accessible for the electrolyte even at high charge/discharge rates. Quite different trends are obtained for the activated carbons and the activated carbon fibers. In the case of the activated carbons, the widening of the porosity (i.e. decrease of the percentage of narrow microporosity) does not produce a considerable improvement of the kinetic behavior and, after a significant widening of the porosity for the AC (29% of narrow porosity), still quite high values of capacitance loss are obtained (92%). However, in the case of the ACFs, the capacitance loss decreases importantly when the porosity gets wider, obtaining the lowest capacitance loss (77%) for an activated carbon fiber with a 74% of narrow porosity. Thus, for Na2SO4 electrolyte, ACFs seem to be more useful in high-power applications than activated carbons. The fact that a widening of the porosity produces better kinetic behavior in ACFs than in ACs in Na2SO4 medium is most probably due to the existence of a smaller tortuosity of networks of porosity in ACFs than in ACs, which makes the ACFs to have a shorter path length for the ions to move than in the case of ACs [4]. The important effect of the different path length of two carbon materials was also observed by Shiraishi et al. [15]. In that case, they studied the reversibility of an ion adsorbing/desorbing process and they compared an activated carbon nanofiber (ACNF) with a 100–200-nm fiber diameter with a conventional activated carbon fiber (ACF) with a fiber diameter of ~10 μm. They observed that the ACFs showed a significant irreversible adsorption of the 1-ethyl-3-methylimidazolium (EMIm+) cation in EMImBF4. However, the ACNFs effectively suppressed the irreversibility, suggesting the important effect of the shorter path length existing on the ACNFs on the ion adsorbing/desorbing process.

More recently, ACNFs prepared by hydroxide activation have also been applied in the EDLC field because of their relative high conductivity [14]. Their behavior was compared with conventional ACFs (fiber diameters of 10 μm, approximately). ACNFs clearly have enhanced their specific capacitance, attaining, for comparable surface.
areas than ACFs, values higher because of their nanometer-size diameters. Additionally, their capacitances were less dependent on the current density than conventional ACFs. These advantages were attributed to a faster adsorption/desorption rate in the micropores of the ACNFs, thanks to their smaller diameters [14].

5. CONCLUSIONS

- ACFs are fibrous microporous carbons that can easily be activated rendering adsorbents with high adsorption capacities and high-surface areas, as it also happens with classical activated carbons. Although their adsorption behaviors do not differ from other forms of microporous carbons, their fiber shape allows getting better performance in some applications. For this reason, ACFs can be considered, in relation to the classical activated carbons, as a new and more powerful generation of microporous carbons.

- The microporosity and the micropore size distribution of ACFs and ACNFs, responsible for most of their applications (i.e. VOC removal, gas separation, and methane and hydrogen storage, as it also happens with other forms of activated carbons), can be well controlled upon selection of the CF precursors, the type of activation, and the variables of the activation process.

- ACNFs much less studied than ACFs can present quite high porosity development and present remarkable interesting advantages over conventional ACFs due to their nanosized diameters (100–200 nm), for example, in areas such as in EDLC and in hydrogen storage. For this reason, ACNFs can be considered, in relation to the classical ACs and ACFs, as a new and more powerful generation of microporous carbons.

ACKNOWLEDGMENTS

The authors thank Spanish Ministry of Science and Innovation, FEDER and Plan E (Project CTQ2009-10813 and PLE2009-0021) and GV (PROMETEO/2009/047) for financial support.

REFERENCES


Activated Carbon Fibers


Carbon is a unique element in the sense that it possesses an electronic configuration that allows single, double, and triple bonds to be formed. Additionally, the extraordinary abilities of the chemical element carbon to combine with itself and with other chemical elements in different ways give rise to a rich diversity of structural forms of solid carbon and carbides [1–3]. This led to the development of different types of carbon having different applications in the early part of the twentieth century and, more recently, to a wide variety of high-performance materials such as carbon fibers, composites, fullerenes and nanocarbons [3–5]. Carbon and graphite in bulk form are commercially important materials finding applications in a variety of industries. The bulk polycrystalline graphite consists of agglomeration of smaller graphite crystallites into a three-dimensional mosaic. The bulk properties of this material depend on the size of the crystallites, their orientation with respect to each other within the mosaic structure, and intracrystallite and intercrystallite perfections. In fact, industrial carbons are a class of composite materials, consisting of a filler phase, coke and a binder phase, the matrixes both of the same element, carbon [2,6]. The bulk properties of the material can be altered by modifying crystallographic parameters of any of these phases. As a result, the bulk properties can be tailored to fit specific needs. However, utilization of the mechanical properties of the single-crystal graphite in manufactured carbons could not be more than 1–5%. The concept of using fibrous reinforcement to improve the strength of graphite dates from the early 1960s and coincides with the availability of carbon fibers and also with an upsurge in interest in the use of other fiber-reinforced composites materials [7], particularly glass fiber-reinforced thermosetting resins. The main drawbacks of polymer matrix composites (PMCs) include low operating temperatures, which can be overcome by using high-temperature matrix systems, metals, or ceramics. Here again, carbon surpasses all matrix systems since it has the highest working-temperature capability under inert environment. Thus, a new family of materials called carbon–carbon composite (C/Cs) was created [3,6,7]. Carbon–carbon composite (C/Cs) were initially developed, for the US Air Force space plane project, NASA's Apollo projects, and then for the US Space shuttle program, for the heat shields followed by the brake pads for fighter aircrafts. Their success in varied high-end applications generated confidence in reliable performance of these materials, and now almost all commercial aircrafts use C/C-based braking systems [3,8–11]. It was unique at that time in having the same material but of different forms for both fiber and matrix constituents of the composites. These materials have replaced graphite in many industrial applications as well. Since the time of development of C/Cs, a few other homocomposites such as silicon carbide–silicon carbide and alumina–alumina have emerged but C/Cs are by far the best known and the most mature example [12]. All these composites address the recognized need to toughen what are fundamentally brittle but refractory materials by fibrous reinforcement. The common objective of C/Cs is to capitalize on the many attractive properties of graphite and combine them with the mechanical properties associated with carbon fiber. Therefore, the interest to make use of excellent mechanical and thermal properties of carbon fibers in highly thermally stable matrix system and requirement of enhancing properties of bulk carbon materials resulted in development of carbon fiber-reinforced carbon matrix composites, or the so
called C/Cs – the group of composites that can have high specific strength and stiffness at elevated temperatures, high thermal conductivity, and excellent dimensional stability.

1. FIBROUS PREFORMS

C/Cs, as the name suggests, consist of highly ordered carbon fibers embedded in a carbon matrix. In general, the properties of C/Cs are attributed to three factors: the strength, content, and distribution of the carbon fibers, matrix, and their interface. The latter too, to a greater extent, is controlled by the energetic surface of the reinforcing fibers. Therefore, the type of the carbon fibers, the fiber content, and the orientation make great contribution to the end properties of C/Cs. Over the years, a great deal of research has gone into development of carbon fibers from the thermosetting polymers [Polyacrylonitrile (PAN) and rayon] as well as highly aromatic thermoplastic pitches through heat treatment at various temperatures in the range of 1200–3000 °C, resulting in carbon fibers with random to highly aligned crystalline structure and medium to ultra-high modulus and thermal conductivities [13–17]. This has been dealt in detail in chapter 2.8 in this book. The types of carbon fibers and their properties are summarized in Table 1. Unlike PMCs, the fabrication method of C/C composite products involves fabrication of a porous carbon-fiber preform (chopped random, unidirectional (UD), two-directional (2D), three-directional (3D) or multidirectional (MD)) having the desired shape and dimensions [6,18] of the end product and/or a porous carbon–carbon skeleton followed by densification using any of the basic two-densification techniques or their combination. The architecture of carbon fibers in the preforms is the backbone of the composites. It provides not only rigidity to the composites but also, in combination with fiber properties, it determines the properties of the end product. In other words, carbon fiber type, directions, and volume fraction in that particular direction are predetermined, dictated by the required end properties. Textile-structured designs play an important role in preparation of these preforms and exhibit balanced properties. A concept of the fibrous preforms is shown in Figure 1. More detailed fiber architectures are discussed by Frank Ko and others [18–21]. For most of the carbon–carbon applications, 2D composites are made, using layup techniques conventionally used in polymer composites. 2D structures are usually formed from several layers of woven carbon fabric (plain, twill, or satin weave), nonwoven felts with or without stitching, or needle punching in a third direction with carbon fibers or braids to give structural strength. For advanced specific applications requiring balanced thermal and thermomechanical properties, multidirectional reinforced (3D, 4D, 7D, etc.) structures are used (Figure 2).

Initially starting with prerigidized yarns to make 3D and MD structures, weaving procedures and looms have been developed to produce dry multidirectional constructions with fibers running in cartesian XYZ coordinates or 3D polar coordinate-weave constructions with fibers running in axial, circumferential, and radial directions or cylindrical structures in modified ways. Theoretically, maximum fiber-volume fractions (~90%) are obtained in UD structures followed by 2D structures (~75–78%) and MD structures (~40–65%) [6,19,22]. These types of fiber architecture possess a wide spectrum of interconnected pore dimensions, pore geometries, and pore distribution available for further infiltration with the carbon matrix and dense C/Cs.

2. PROCESSING OF C/Cs

Processing of C/Cs consists of building up of the carbon matrix around the carbon fibers. The conventional methods of melting or sintering used in ceramics technology cannot

<table>
<thead>
<tr>
<th>TABLE 1 Properties of Different Types of Commercial Carbon Fibers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precursor</strong></td>
</tr>
<tr>
<td>Diameter (μm)</td>
</tr>
<tr>
<td>Density (g/cc)</td>
</tr>
<tr>
<td>Tensile strength, GPa&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Tensile modulus, GPa&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Coefficient of thermal expansion along the fibers, x 10&lt;sup&gt;-6&lt;/sup&gt; K&lt;sup&gt;-1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Thermal conductivity, Wm&lt;sup&gt;−1&lt;/sup&gt; K&lt;sup&gt;−1&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
be applied to C/Cs since carbon or graphite melts under high pressure and temperature, i.e. to the order of 100 bars and 4000 K respectively. Like conventional carbon products, there are basically two routes to create the carbon matrix in C/Cs: (1) through gaseous route, i.e. thermal cracking of hydrocarbons, methane, propane, acetylene, etc. using chemical vapor deposition/infiltration (CVD/CVI) [22–24] and (2) through pyrolysis of carbonaceous precursors, i.e. a series of liquid impregnation and pyrolysis (LIP) steps using thermosetting resins (phenolics, Poly furfuryl alcohol (PFA), etc.) or thermoplastic pitches (coal tar pitch or petroleum pitch) as carbon precursors [6,22,24–26]. General scheme of manufacturing C/Cs is given in Figure 3. Both the processes have typical advantages in terms of achieving densities, desired microstructure, mechanical, thermal, and functional properties, and performance for desired application. A combination of the two processing routes is the best choice and is currently followed to produce carbon—carbon products.

2.1. Fabrication/Densification Techniques
2.1.1. CVI/Densification Technique
CVD/CVD processing involves infiltration of gaseous hydrocarbons into the porous preforms followed by cracking of these gases, resulting in densification of the porous carbon fiber preforms with the carbon matrix [22–29]. The fibrous preforms could be first rigidized with a small amount of thermosetting resin, followed by
pyrolysis to result in highly porous C/Cs with open and well-connected pores. The densification can be achieved through vapor-phase infiltration (or CVI) wherein the hydrocarbon gases such as methane, propane, etc. infiltrate into porous fibrous structure heated to a temperature of 1000–1400 °C and are made to crack therein.

In the case of the CVI, the feed gas undergoes a series of vapor-phase interaction both of homogeneous nature
as well as heterogeneous nature (gas/solid carbon, gas/ fibers) as it diffuses into the preform, forming intermediate species controlled by the reaction kinetics and process conditions, which further react to deposit carbon on all heated surfaces, thus coating all the carbon filaments taking their contour. Figure 4 summarizes the overall CVI process, rate controlling factors, and cracking mechanism, which are dependent on the reactor type, geometry, and temperature; hydrodynamic, the gas laminar flow, and pressure and the nature of the cracking hydrocarbon [24,30,31].

The overall deposition rate is highly sensitive to the surface characteristics, i.e. roughness, microstructure, especially orientation of graphitic planes, surface energy and surface area of the substrates, the fibers. It has been found that the active sites on a graphene surface and the plane edges are the preferential nucleation sites, which control the first step of kinetic deposition [32].

 Industrially, C/Cs are produced primarily using three techniques: (i) Isothermal—isobaric process, (ii) thermal-gradient process, and (iii) pressure-gradient process, sketched in Figure 5.
The isothermal technique is shown schematically in Figure 5a. In this process, the fibrous preform is placed within the susceptor of an inductionally heated furnace, and hydrocarbon gases are introduced into the susceptor to infiltrate and deposit carbon within the hot porous preform. The isothermal process is usually conducted at low temperatures (around 1000–1100 °C) and reduced pressure [6,22,23]. Principally, the gases should infiltrate into the pores and interfilamentary spaces and crack to make a uniform matrix. But these gases have tendency of cracking at the outer surface itself, which block the passages and cause close pores. As a result, the density of the composites is less at the center of the product. Processing at lower temperatures, pressures and gas concentration can improve densification. Practically, in order to minimize the density gradient through thickness of the component and to have dense composites, the process is interrupted after few weeks, surfaces are grinded, and the components are reinfiltated. Thus, the isothermal CVD/CVI process is a very slow process requiring several processing steps, and it takes months together to get a dense C/Cs, adding cost to the process. In variance with the slow and complex conventional isothermal-heating method, alternative methods of forced-flow/thermal-gradient processes have been developed for building up the carbon matrix around carbon fibers [33–35]. Carbon-bearing propylene, propane, or methane is forced under pressure through the perform while it is heated in an oven at 1000–1200 °C. In hollow preforms, the gases are forced to flow from inside to outside. Although carbon gets deposited at the first interacting surface, the pressure gradient lets it flow further, infiltrating right throughout the preform. These are also termed as “differential pressure process.” Forced-flow CVI process is used industrially to produce large quantities of carbon–carbon products for friction applications. Stack of components are impregnated simultaneously in a large-size furnaces.

A second procedure for CVI, known as thermal-gradient process, shown schematically in Figure 5b essentially comprises maintaining a temperature gradient in the substrate, with temperature being maximum at the center and decreasing to the surface. It forces vapor to crack first at the center, retaining flow through the preform, and gradually cracking to surface, ensuring even formation of the matrix. In the temperature-gradient process, inner surface of the preform is kept at higher temperature than the outer surface. Hydrocarbons diffuse through the cooler outer surface of the preform toward the hotter inner surface, thereby cracking on the inner surface. The temperature of the preform progresses outwardly as the densification increases from the inner to the outer surface. Vapor infiltration and carbon deposition are faster with these methods. This method is reported to deposit the carbon matrix up to 30% faster than conventional methods and allows thicker items to be produced [36,37]. However, this process is limited to densification of single product at a time. A modified thermal-gradient technique involves utilizing the conducting property of carbon fibers. The preform is placed in a short RF coil, and slowly drawing the preform through the coil while hydrocarbon gases flow at fast speed around the preform, the complete preform is heated and densified.

In variance to conventional CVD processes, modified CVD processes such as plasma-enhanced CVD process [38,39] has also been tried experimentally but has yet to be commercialized. Higher manufacturing expense associated with CVD/CVI processes limits the use of C/Cs.

Another novel process has been developed to obtain high conversion efficiency of hydrocarbon, the so-called film-boiling technique [24]. This process ensures uniform vapor infiltration, and it can be run under a wider range of operating conditions than the conventional CVD—the process is less dependent on precise heating, pressure, and timing conditions. In the future, this flexibility may even allow the addition of graphitization catalysts and oxidation preventers during production, thereby eliminating a separate treatment.

### 2.1.2. Liquid Impregnation and Pyrolysis Technique

The liquid impregnation and pyrolysis (LIP) technique is extension of the conventional filler/binder technique used for processing of industrial carbon products where carbon fibers
replace the fillers [6]. Porous fibrous structures are impregnated with liquid thermosetting resins (phenolic/polyfurfuryl alcohol) or thermoplastic pitches using evacuation and pressurization, followed by carbonization and high-temperature heat treatment (1000–2700 °C) in oxygen-free atmosphere. Schematic representation of this technique is included in the Figure 3. It essentially comprises formation of polymer composites using conventional polymer composites manufacturing techniques using carbonaceous resinous precursors. Mass loss due to volatilization of hydrogen, nitrogen, etc. and dimensional changes (shrinkage) during carbonization result in pores and cracks in the carbon matrix and a reduction in density. The density of carbon skeleton depends on the carbon yield of the resin precursor (Table 2) and pyrolysis conditions. Aromatic polymers with higher carbon yield have been developed to get skeleton carbon composites with high initial density [6,40,41]. Numerous cycles of evacuation and pressure impregnation of porous composites with carbonaceous liquids/pyrolysis have to be performed to fill in the pores and small cracks and to achieve the desired density. The part must then be reinjected, infiltrated with liquid carbonaceous precursors, and pyrolyzed several times (up to a dozen cycles) to fill in the small cracks and to achieve the desired density. The pyrolysis of the pitch/resin matrix can be done either under normal pressure or High Pressure Impregnation Carbonization (HIP). HIP results in high carbon yield up to 80% for pitches and hence carbon-composites skeletons are then sealed in a flexible thermosetting resins is controlled through addition of graphite powders or carbon black or pitch-derived cokes. Such modified matrix systems help in reducing the infiltration cycles and result in high densities with more isotropic carbon matrix [45]. The yield of thermoplastic pitches can be increased by reacting them with chemicals like sulfur or iodine. This also result in composites with higher densities at relatively lower impregnation cycles [46]. Another technique [47] uses variation on pre-impregnated or “prepreg” materials used to create a preform. Graphite yarns are coated with graphite precursor powders made from coke [47] and are then seared in a flexible thermoplastic sleeve to protect the powder coating during handling and manufacture. The treated yarn can then be woven into sheets or chopped into short fibers and applied to a mold. The laid-up form is then hot pressed to make the composite part. Yarns can also be processed into tubes, rods, cloth, thick textiles, UD sheets, and tapes. Such composites are three-phase materials comprising fibers and particulates as reinforcements and carbonaceous liquids as matrix precursors. Since filler carbons do not undergo further weight loss during pyrolysis, fewer densification steps are needed to manufacture dense composites. Better penetration of the matrix into fiber bundles ensures uniform properties in the composite and higher strength than conventional composites.

### TABLE 2 Carbonaceous Precursors and Their Carbon Yields and Pyrolysis Efficiencies

<table>
<thead>
<tr>
<th>Precursor</th>
<th>Carbon Yield (%)</th>
<th>Pyrolysis Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyacets</td>
<td>85</td>
<td>&gt;90</td>
</tr>
<tr>
<td>Polyphenylene</td>
<td>85</td>
<td>&gt;90</td>
</tr>
<tr>
<td>Polymides</td>
<td>60</td>
<td>85</td>
</tr>
<tr>
<td>Phenolic resins</td>
<td>55–65</td>
<td>75–80</td>
</tr>
<tr>
<td>Furan resins</td>
<td>55–65</td>
<td>75–80</td>
</tr>
<tr>
<td>Polybenzimidazole</td>
<td>74</td>
<td>95</td>
</tr>
<tr>
<td>Resorcinol furfuryaldehyde</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>Oxidized polystyrene</td>
<td>55</td>
<td>85</td>
</tr>
<tr>
<td>PAN</td>
<td>45</td>
<td>65</td>
</tr>
<tr>
<td>Polysulfone</td>
<td>48</td>
<td>66</td>
</tr>
<tr>
<td>Epoxidised phenol-formaldehyde</td>
<td>50</td>
<td>63</td>
</tr>
<tr>
<td>Epoxy resin</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

#### 2.1.3. Nano Reinforced C/Cs

With the advent of carbon nanomaterials, i.e. carbon nanofibers, carbon nanotubes (CNTs), graphene, etc, having super mechanical and thermal properties, the interest got generated to develop C/Cs with these carbon nanomaterials as third-phase material. These combinations result in improved thermal properties of the composites in transverse directions as well [48–50]. However, proper distribution of these nanomaterials in the resins/pitches is a major task.
2.2. Structural Aspects of Carbon—Carbon Composites

The prime factors of importance in achieving tough and high thermal conductive C/Cs are the proper choice of reinforcing carbon fibers and the carbon matrix since the microstructure of both the components have major influence on the mechanical and thermal properties of the composites. The processing routes and the choice of carbon precursor greatly influence the density, macrostructure (type, size, and quantities of defects, i.e. defects, pores, cracks, etc.), and the matrix microstructure (orientation of graphitic planes). Some of these striking structural features of C/Cs are listed in Table 3 and are shown in Scanning optical microscopy micrographs of the composites (Figure 6). The voids and macrocracks are normally less in pitch-derived C/Cs processed through HIP route than in those made by CVD route or phenolic resin-derived carbon route. The latter exhibit low densities. These features also control thermal transport mechanism in C/Cs. In order to have desired fiber/matrix bonding and to reduce cracking at the fiber/matrix interface, pyrolytic carbon coatings are also given on to the surface of the fibers.

The matrix microstructure (degree of graphitization) is also an important aspect governing the properties of C/Cs. From strength point of view, a semicrystalline or randomly oriented carbon matrix is desired, which can arrest crack propagation, whereas for thermal and electrical properties, highly graphitic matrix is a prerequisite. The development of crystalline carbon matrix originates from the CVD route or pitch route. Indeed, the structure and properties of pyrocarbons deposited through CVD route are related to the deposition conditions, which control the mechanism of carbon formation. Many studies have been carried out to understand the mechanism of carbon formation from different hydrocarbons by CVD process and to establish a relationship between the deposition conditions and the structure of the pyrocarbons [28—31]. Based on polarized

![Image](a)

![Image](b)

**FIGURE 6** Optical micrographs of C/Cs prepared by (a) pitch general route, (b) pitch—HIP route.

**TABLE 3** Micro/Macrostructural Features of C/Cs and Their Effect on Composites Properties

<table>
<thead>
<tr>
<th>Nature</th>
<th>Scale</th>
<th>Position</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Micromechanical cracking</td>
<td>* Fiber dia scale</td>
<td>* Cracking in matrix</td>
<td>* Load transfer among fibers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>* Fiber/matrix interface</td>
<td>* Transverse properties of fiber bundles</td>
</tr>
<tr>
<td>* Minimechanical cracking</td>
<td>* Cloth layer thickness scale</td>
<td>* Fiber bundle</td>
<td>* Porosity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>* Interface between fiber bundles</td>
<td>* Load transfer among fiber bundles and laminae</td>
</tr>
<tr>
<td></td>
<td></td>
<td>* Interyarn pocket of unreinforced matrix</td>
<td>* Major influence on mechanical and thermal properties</td>
</tr>
<tr>
<td></td>
<td></td>
<td>* Interface between fiber bundle and matrix pocket</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>* Matrix—matrix interface</td>
<td></td>
</tr>
</tbody>
</table>
light optical microscopy, pyrocarbon matrix in C/Cs is grouped into three main categories of microstructure: smooth laminar (SL), rough laminar (RL), and isotropic carbons, their formation controlled by the competing thermochemical reactions. As an example, for a fixed gas flow of methane, SL is formed at low temperature and high partial pressure of gas, whereas RL is deposited in the medium range and isotropic (ISO) at high temperature with rather a low pressure. Of course, boundaries between these structures are dependent on experimental conditions such as nature and porosity of the substrate, nature of the hydrocarbon precursor, and deposition temperature, pressure, etc. According to some studies, the ratio R (R = C2H2/C6H6) controls the formation of SL—RL—ISO microstructures. Lieberman and Pierson [28,29] attributed different microstructure types to relative quantities of gaseous acetylenic and aromatic species, i.e. C2H2/C6H6 molar ratio in cracking gas typically as SL < 5; 5 < RL < 70; ISO > 70. It has also been observed that SL and ISO microstructures form under a wide variety of conditions, while RL material apparently forms only under narrow range of conditions. These deposited structures are free of any large cracks. In the case of a thermal-gradient process, the temperature of the substrate felt is lower near the outer portion than the inner one. The C2H2/C6H6 molar ratio will also be low near the outside surface. Accordingly, one may expect a structural transition from the inside to the outside of the felt to occur in the order of ISO—RL—SL. As per another study, SL-type deposit is related to the use of aliphatics and the RL microstructure mainly to aromatics cracking. Quite often, a combination of microstructures (RL—SL) is quite often observed in a fibrous tow or in inter tow regions. Optical, X-ray Diffraction (XRD), and electron microscopic techniques have been used to characterize these types of carbons and their distinct properties. These are summarized in Table 4. The densities of various carbon structures increase with increasing anisotropy (dRL > dSL > dISO), consistent with crystallographic parameters. Table 5 gives comparison of the different processes used for making C/Cs.

The type of hydrocarbon has been observed to influence the texture of carbon in CVD carbon composites. Kimura et al. [33] have observed cone-like columnar structures in a propane-derived pressure-controlled carbon matrix, the dependence of which on processing parameters is shown in Figure 7.

### 2.2.1. Macroscopic Features

With respect to minimechanical features, C/Cs are a buildup of fiber bundles containing filaments, thin interfilament matrix, interbundle-matrix-rich regions, and macropores. The latter includes the voids formed as a result of imperfect fiber stacking or processing and the bubble pores due to evaluation of gaseous by-products during carbonization, dry zones, and shrinkage cracks, etc. [6,42,51,52]. Figure 8 is a schematic representation of different types of matrixes present in C/Cs [52]. Interfilament spacing in a bundle may also affect the microstructure of the matrix in that region.

In C/Cs made with pitches as matrix precursor by low-pressure processing, voids, bubble pores, and the matrix macrosequences are observed. Voids are features of the microstructure of C/Cs, which are mainly generated during the initial process of the prepreg and during the subsequent steps of carbonization. Polarized light microscopy shows distinct three types of voids in undensified composites: elongated cracks between fiber bundles in both lateral (parallel to the plane of the fiber bundles) and transverse (perpendicular to the plane of the fiber bundles) directions and isolated devolatization pores between or within the bundles (Figure 9). Optical micrographs of undensified C/Cs showed different types of voids. It showed elongated transverse cracks, which are produced by carbon matrix shrinkage during the carbonization treatment and by misalignment of the fibers during initial stacking of the coated fiber bundles. This crack runs parallel to the plane of the fiber bundles. It also showed lateral cracks, which can be devolatilization pores joined together by cracks of a single pore. These cracks run perpendicular to the plane of fiber bundles. The size of all these voids ranged from as low as 30 μm × 30 μm to as high as 1400 μm × 700 μm. Densification exhibits region of distinct primary and secondary carbon matrixes (Figure 9b).

The type of reinforcement and pitch-matrix precursor used is found to be influencing the structure of C/Cs (Figure 10). The high quinoline content pitch as the matrix displayed mosaic-type structure throughout the pyrolysis stage of the composite, while the low quinoline-content
### TABLE 5 Comparison of Characteristic Features and Properties of Carbon Matrixes from Vapor-Phase, Pitch, and Resin Precursors

<table>
<thead>
<tr>
<th>Type of Carbon Matrix</th>
<th>CVD</th>
<th>Pitch</th>
<th>Resin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density</strong></td>
<td>High, approx. 2000 kg m⁻³</td>
<td>High, increases with carbonization pressure</td>
<td>Low 1300–1600 kg m⁻³</td>
</tr>
<tr>
<td><strong>Carbon yield</strong></td>
<td>C directly deposited; no further thermal degradation</td>
<td>Varies according to pitch composition 50–80% w/w</td>
<td>About 50% for phenolics, increasing to 85% for polyphenylenes.</td>
</tr>
<tr>
<td><strong>Porosity</strong></td>
<td>Low, expect for isotropic form; laminar fissures</td>
<td>Macrosized gas entrapment pores plus shrinkage and thermal stress fissures.</td>
<td>High microporosity (pore diameter &lt;1.0 nm); becomes closed above 1000 °C. Macropores may be evident due to vapor evolution during the curing of resin.</td>
</tr>
<tr>
<td><strong>Microstructure (in the bulk state)</strong></td>
<td>Varies from isotropic to highly orientated laminar forms</td>
<td>Macromdomains (1–100 μm) showing preferred orientation developed from the mesophase state</td>
<td>Isotropic, except on the nanoscale, i.e. basic structural unit (BSUs) are randomly orientated.</td>
</tr>
<tr>
<td><strong>Orientational effects within fiber preforms</strong></td>
<td>Strong orientation of laminar matrix with fiber surfaces; orientation also on crack surfaces within other matrix types if subsequently subjected to CVI</td>
<td>Preferred orientation of lamellae with fiber surfaces increases as HTT develops the graphitic structure; modified by pressure pyrolysis</td>
<td>Preferred orientation at fiber surface but to a much lower extent than other precursors; increases with HTT.</td>
</tr>
<tr>
<td><strong>Graphitibility/crystallinity</strong></td>
<td>Laminar forms highly graphitizable; can have Lc and La &gt; 200 nm after 3000 °C HTT</td>
<td>Highly graphitizable Lc and La usually less than 100 nm</td>
<td>Normally nongraphitizing.</td>
</tr>
<tr>
<td><strong>Purity/composition control</strong></td>
<td>Controlled by gas composition; enables other elements to be incorporated into the deposit (e.g. Si, B)</td>
<td>Controlled by source; can be of high purity; not easy to incorporate other elements expect as powders</td>
<td>As for pitch.</td>
</tr>
<tr>
<td><strong>Reactivity to oxidizing gases</strong></td>
<td>Very low reactivity for highly oriented Pyrographite</td>
<td>Low reactivity decreases with HTT</td>
<td>Usually high reactivity due to micropore network; decreases with HTT but still relatively high.</td>
</tr>
<tr>
<td><strong>Thermal expansion</strong></td>
<td>Depends on preferred orientation; can be highly anisotropic approaching values for the crystal in the two major directions</td>
<td>Depends on domain orientation; expansion partly accommodated by lamellar cracks; 1–5 × 10⁻⁶ K⁻¹</td>
<td>Isotropic in bulk; -3 × 10⁻⁶ K⁻¹.</td>
</tr>
<tr>
<td><strong>Thermal and electrical conductivity</strong></td>
<td>Determined by referred orientation, approaching single-crystal graphite values</td>
<td>Depends on domain orientation, HTT, and internal porosity, increasing with HTT</td>
<td>Isotropic in bulk.</td>
</tr>
<tr>
<td><strong>Young’s modulus</strong></td>
<td>7–40 GPa depending on the structure</td>
<td>5–10 GPa depending on grain size, porosity, and degree of graphitization but up to 14 GPa for very fine-(1 μm) grain size materials</td>
<td>10–30 GPa.</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>Depends on microstructure and degree of graphitization; 10–500 mPa</td>
<td>Depends on porosity and pore geometry: 10–50 mPa for most polycrystalline carbon/graphites rising to 120 mPa for very fine-grain (1 μm) size materials</td>
<td>Approximately 8–150 mPa for glassy carbons, depending on porosity.</td>
</tr>
<tr>
<td><strong>Failure strain</strong></td>
<td>0.3–20% depending on structure; Higher values at highest deposition/HTT</td>
<td>Up to about 0.3% depending on grain size and degree of graphitization</td>
<td>Up to about 0.4% (largest values for glassy carbons).</td>
</tr>
</tbody>
</table>

Lc, stack height; La, stack width; HTT, heat treatment temperature; BSUs, basic structural units.
pitch and AR-mesophase pitch show flow-type structure. The matrixes in both bulk as well as interfilament show a well oriented lamellar-like structure. Although matrixes in all these composites exhibit good orientation of the carbon matrix, the extent of orientation is smaller in the case of high-strength PAN-based carbon fiber-reinforced composites than that seen in composites with intermediate modulus Dona F-180 pitch carbon fiber and ultrahigh modulus P-100 fiber-reinforced composites. C/Cs developed with AR-mesophase pitch, containing milled intermediate modulus-pitch fibers, exhibit disturbed structure of the matrix. In graphitized C/Cs with HT-PAN-based carbon fibers comparatively good bonding with the matrix is observed as against ultrahigh modulus-carbon fiber-reinforced composites.

2D C/Cs made with woven cloth as reinforcement exhibit additional significant microstructural features due to the yarns running in two directions (warp and fill), passing over and under each other and imparting a wavy nature to them. Although these composites are made by prepregging and compaction technique similar to UD composites, a significant portion of the composite remains as unreinforced matrix-rich region such as interyarn pockets, bundle cross-over region, interplay regions, etc. The configurations of these regions will depend on the weave type (plain, twill, or satin weave), stacking, compaction technique, etc. [52–55].

In 2D composites, especially in thick ones, the heating rate has to be chosen carefully to avoid delamination of the composites during heat treatment. Similar to UD and 2D composites, MD composites also contain shrinkage cracks and thermal stress cracks. Since the latter composites are composed of a rigid fiber architecture that restrains the overall shrinkage, cracks are likely to be more pronounced in them than in UD and 2D composites. Another striking feature of the MD carbon–carbon structures is the billowing of the fiber bundles, as shown in Figure 11 [56].
Thermosetting polymer route results in an amorphous carbon matrix. However, in composites made with high-strength carbon fibers or surface-treated carbon fibers due to strong fiber(matrix bonding at polymer stage and differential pyrolysis shrinkage in fibers and matrix, thermal stresses are generated at the fiber(matrix interface during carbonization of the composites, causing stress graphitization of the matrix in the composites. The graphitic structure starts developing at a temperature of 2000 °C (Figure 12). The interbundle matrix and the matrix far away from the fibers remain isotropic in nature. However, in the case of high fiber-volume content composites, the shrinkage stresses generated at the interfaces propagate further into the matrix, which on heat treatment to 2000 °C and above cause stress graphitization of the complete matrix [52,53,58]. This subject has been of great scientific interest, especially correlating fiber(matrix bonding and resulting orientation of graphite basic structural units in the matrix. The technological importance of this phenomenon has been to control the fiber(matrix bonding so that low-priced carbon fibers can be used for making C/Cs for mechanical and general purpose application.

FIGURE 9 Optical micrographs of C/Cs (a) skeletons and (b) densified composites.

FIGURE 10 Optical micrographs of the composites made with different types of fibers and composites.
Another prominent feature of the matrix microstructure, especially in the lamellar regions, is the extensive kinking of the layer planes. Kinking is typically observed in the lamellar matrix around fibers with uneven surfaces with the kinks conforming to the depression and asperities on the fiber surface. Kinked matrixes are also observed in composites made with excessively surface-treated carbon fibers.

In contrast to the microstructure of thermosetting-resin-derived carbon, that of thermoplastic-pitch-derived carbon is controlled by nematic-liquid crystal (mesophase) formation during liquid-phase pyrolysis of the pitches. The flow of the oriented molecules of the liquid crystals during their formation and subsequent deformation results in the introduction of disclination in the structure. During pyrolysis of the pitch precursor in the fabrication of C/Cs, the carbon filament forms a circular sheath several microns thick around each filament. The parallel alignment of the filaments and mesophase sheath force the disclination lines to be parallel to the fibers. However, the strength and type of disclination is determined by the local arrangement or
pattern of the arrays of filaments [59,60]. The sheath-like orientation of the pitch-carbon matrix has been observed in all types of C/Cs made with round-carbon fibers irrespective of PAN or pitch (Figure 13) [6,48]. However, due to the relatively less active surface in pitch-carbon fibers, the matrix debonds from the fibers and the majority of shrinking cracks are developed at the fiber/matrix interface, instead of in the bulk matrix.

3. PROPERTIES OF C/Cs

C/Cs, in true sense, cover a large range of materials with targeted properties. The properties of interest are strength and stiffness, fracture toughness, frictional properties, thermal conductivity, and resistance to oxidation at high temperatures. The operating mechanisms for these properties are quite different especially in such multiphase composite materials. The mechanical properties of the constituents and their volume fraction, bonding, and crack propagation mechanism control the mechanical properties of the composites, whereas the thermal properties are governed by thermal transport phenomena. Moreover, the constituents, both reinforcement and matrix, are likely to undergo a change in properties during processing as influenced by heat-treatment temperature, differential dimensional changes, thermal stresses, etc. All these factors influence the ultimate properties of the composites. It is difficult to cover completely all these aspects in this article. However, a general trend of the most fascinating properties of C/Cs are given below.

3.1. Mechanical Properties of C/Cs

The mechanical properties of advanced composites are generally related to the properties, volume fraction, and direction of the fibers relative to the direction of the applied stress. Considering these factors, the mechanical properties of PMCs or, in a broader sense, of ductile matrix composites are calculated by a simple rule of mixture. However, in C/Cs, the matrix has a failure strain of less than 0.5%, which is lower than that of the fibers [2,6]. Consequently, theories of reinforcement that have been developed for commercial composites with high-strain matrix systems and strong fiber/matrix bonding do not apply to C/Cs. Moreover, matrix structure (both at mini- and microscales) and fiber/matrix bonding in these composites differ depending on the type of reinforcements and processing routes. Accordingly, different principles will prevail for the mechanical properties of C/Cs processed through various routes. Furthermore, the data concerning mechanical properties, which are available in the literature, are based on the constituents and the processing condition adopted by individuals.

3.1.1. Strength of the Composite

The prediction of the strength matrix composites is rather complicated since it is governed by the mechanism responsible for crack initiation and crack propagation. However, in brittle matrix composites, with the matrix having a lower failure strain than the fibers, crack initiation takes place in the matrix at much lower stress levels, and the fiber strength is not fully utilized. Therefore, in order to have maximum utilization of the fiber strength, the failure strain of the matrix must be increased. Further, the strength and fracture of C/Cs are governed by Cook–Gorden [61] theory for strengthening of brittle solids, which states that if the ratio of the adhesive strength of the interface to the cohesive strength of the solid is in the right range, a large increase in strength and toughness of otherwise brittle material is achieved. Extensive work has been done in achieving the translation of highest possible fiber properties in C/Cs. Composites with strong fiber/matrix bonding fail catastrophically without fiber pullout, while those with controlled interface fail in the mixed tensile-cum-shear mode exhibiting high strength. A generalized view of the effect of processing temperature, the fracture mechanism, and the strength of different types of C/Cs is given in Figure 14.
Like with any composite material with a polymer matrix, in C/Cs too, the fiber layup architecture and size of the carbon control the ultimate strength of the composites. This is seen in Figure 15.

### 3.1.2. Mechanical Properties of Composites with Pyrolytic Carbon Matrix

In C/Cs with pyrolytic carbon, tensile strengths in the fiber direction of 1000–1400 MN/m² have been reported, which correspond to 80–90% of the fiber properties' translation [62,63]. At lower heat treatment temperature (HTT < 1500 °C), composites with high strength have been obtained, while the opposite trend is observed at HTT > 1500 °C.

The type of CVD technique chosen for composite fabrication has also been found to influence the mechanical properties. Composites with 50% fiber volume contents and made by the isothermal process are reported to exhibit much higher strength than those made with the temperature- or pressure-gradient technique. CVD-derived ISO and SL carbon matrices are strongly bonded to the reinforcing carbon fibers. As-deposited RL and SL carbons do not contain microcracks. However, after heat treatment of the composites, circumferential microcracks develop in the SL matrix as well as at the
interface between RL and SL due to anisotropic thermal contraction [63]. Both the strong fiber/matrix bonding and microcracks in the matrix result in low-strength C/Cs. Because of this, RL structures are preferred for making high-performance C/Cs. As-deposited RL composites possess a similar or somewhat higher flexural strength and modulus than SL composites. Upon heat treatment to graphitization temperature, SL matrixes exhibit a higher degree of graphitization with the others following in the order RL > ISO. With heat treatment temperature, there is only marginal change in properties of the latter composites. Kimura et al. [33] have studied the fracture behavior of CVD multiphase C/Cs. As seen in Figure 16, composites with ISO + Rough Columnar (RC) matrix fracture predominantly in tensile mode without fiber pullout. A crack initiated in RC carbon matrix propagates perpendicular to the fiber axis without any deflection owing to the strong interface between RC and ISO structures as well as between the carbon fiber and ISO matrix. These composites also exhibit a high compressive strength and the crack once initiated, encounters multiple deflections due to microcracks in the Smooth Columnar (SC) matrix. These composites possess a somewhat lower interlaminar shear strength but higher fracture energy. The conclusions derived from UD composites have been found to hold true also for CVD felt, 2D, and MD C/Cs [6,8-11,24].

Moreover, the strength of the composites made with high-modulus nonsurface-treated carbon fibers increases with further densification of the composites, but it decreases for those made with high-strength carbon fibers. Similar effects are observed when carbon fibers of the same high modulus but with different surface functionalities are used as reinforcement [6]. Contrasting differences are observed also in their fracture behaviors. Composites made with strong fiber/matrix bonding fail catastrophically with no fiber pullout (Figure 17), while others fail in mixed tensile-cum-shear mode with fiber pullout depending on the extent of bonding [42,52,57]. However, upon further heating of the composites to graphitization temperature, the strengths of the two composites exhibit different trends. Composite processing strong fiber/matrix bonding and low strength at the carbonized stage exhibit higher strengths at the graphitized stage and vice versa. Based on the experience with different fiber/matrix systems, a general trend can be derived as shown in Figure 17. The fracture behavior of the composites also changes with heat treatment. This is evident from stress-strain diagrams and scanning electron micrographs of the fractured composites. The change in the strength pattern on heat treatment is a consequence of the change in failure modes. In the temperature range of 1200—2200 °C, a mixed-mode failure prevails. The
heat treatment in this temperature range weakens the fiber/matrix interface resulting in crack blunting. This continues until an optimum level is reached where the interfacial bond strength is weak enough for deflection of the crack and strong enough to effectively utilize the stress-transfer capabilities of the matrix. As a net result, the composites fail in mixed mode with low fiber pullout. On further heating to 2500 °C and above, the graphitization of the interfilament matrix increases.

Multidimensional C/Cs exhibit pseudoplastic fracture behavior at room temperature with strains up to 5%, which is characteristically different to that of UD and 2D reinforced C/Cs. This difference attributes to the presence of reinforcing fibers in different direction and the continuous crack structure of the matrix whose geometry is dependent on the fiber array.

3.1.3. Sandwich Composites

Sandwich composites are mix-form composites made of predetermined alternating layers of UD and felt- or cloth-based prepregs. The purpose of making sandwich composites is to increase the transverse strength of C/Cs while retaining as much axial strength as possible, and the exact strength values, however, will depend on the type of reinforcement used and the ordering adopted for making the sandwich [6].

3.1.4. Fracture Toughness of C/Cs

In addition to strength and stiffness, another property of considerable importance from a structural point of view is toughness, i.e. resistance to crack propagation. Carbon as such, both in bulk as well as in fibrous form, is brittle with a fracture toughness (energy required to propagate a crack) in the range of 0–200 J/m². However, if carbon fibers and carbon matrix are combined to form composite materials with optimum fiber/matrix bonding, the composites will exhibit a higher toughness and work of fracture than carbon in monolithic form (Table 6). C/Cs exhibit good fatigue and creep resistance too. [64]

### TABLE 6 Fracture Toughness of Some C/Cs

<table>
<thead>
<tr>
<th>Material System</th>
<th>Vf%</th>
<th>Sample Size (mm)</th>
<th>K (Nm m(^{3/2}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Felt/pitch</td>
<td>25</td>
<td>1.8×5.0</td>
<td>83 ± 8</td>
</tr>
<tr>
<td>Felt/pitch</td>
<td>33</td>
<td>1.8×5.0</td>
<td>100 ± 8</td>
</tr>
<tr>
<td>Felt/CVD</td>
<td>9</td>
<td>9.5×9.5</td>
<td>34 ± 2</td>
</tr>
<tr>
<td>UD/pitch</td>
<td>50</td>
<td>4.8×4.8</td>
<td>15 ± 2</td>
</tr>
<tr>
<td>Fabric/pitch</td>
<td>35</td>
<td>2×10</td>
<td>15 ± 2</td>
</tr>
<tr>
<td>Angle ply/CVD</td>
<td>NA</td>
<td>NA</td>
<td>20 ± 3</td>
</tr>
</tbody>
</table>

#### 3.2. Mechanical Properties at High Temperature

Figure 18a and b shows typical cases of mechanical properties of various C/Cs at elevated test temperatures [6]. In all of the composites, the strength as well as the modulus at elevated temperatures is higher than that at room temperature. At high temperatures, several characteristic phenomena responsible for the mechanical properties of the composites take place simultaneously. Annealing of gross defects result in an improvement of mechanical properties. The closer of stress-distributed microcracks due to thermal expansion of fibers and matrix too change the fracture behavior of the composites [9,10]. As a combined result of all of these phenomena, the C/Cs exhibit increase in strength and stiffness by 10—60% above 1000 °C. The compressive strength has been observed to increase by as much as 100%. The tensile and comprehensive modulus of the composites decreases with temperature above 1600–2000 °C. The fall in the modulus can be explained on the basis of the symmetric potential curve of graphite crystals. With an increase in temperature, the average atomic distance is increased, resulting in a decrease in the modulus. Like strength, the work of fracture of C/Cs also increases with test temperature up to 1800–2000 °C. As mentioned above, the structural changes at high temperature influence the stress transfer mechanism in the composites.

Senet et al. [65] have studied the fracture behavior of 2D discontinuous carbon-fiber-felt-reinforcement phenolic resin and pitch precursor C/Cs at elevated temperature using both chevron-notch and straight-notch single-edge beam specimens. They observed that \( K_{IC} \) values of the chevron-notch specimens increased slightly with the test temperature through 1650 °C, while the straight-notch values remained almost constant throughout the temperature range up to 1650 °C. They suggest that fiber debonding, matrix damage, and fractional sliding during pullout are the critical constituents of an effective wake-zone process responsible for the toughening of such composites. Composites made with PAN-carbon fibers exhibit a much higher increase in fracture toughness than other composites [66].

The mechanical behavior of C/Cs at high temperature discussed above holds true for UD, bidirectional, and multidirectional reinforced C/Cs made with continuous fibers or felt and liquid- as well as gas-phase-derived carbon matrix. The absolute strength values may vary depending upon the configuration.

C/Cs are susceptible to oxidation at high temperature if not protected. Therefore, for the reliable performance of
a carbon–carbon component, it is essential to know the effect of oxidation on both the mechanical and thermal properties of the composite. Although this is an important aspect, not much information is available in the literature. Oxygen attacks preferentially the less-ordered carbon at matrix-rich region, bundle/bundle interfaces, or fiber/matrix interfaces. Accordingly, the fracture behavior and the mechanical and thermal properties will be altered. Thermal conductivity will be reduced due to the porosity generated in the composites. The change in mechanical properties are governed by the preferential attack by oxygen as stated above [65,67]. This will have a severe effect if the weight loss is due to attack on the fiber/matrix interface and is more than 5% [65], which will result in a drastic decrease in the strength of the composites. If oxidation takes place at low temperatures (400–500 °C) and is less than 2%, it may alter the fiber/matrix or matrix/matrix interfaces, resulting in change of fracture to mixed-mode type and, hence, an improvement of mechanical properties [68]. However, in practice, the application temperatures are much higher than 500 °C, and the oxidation weight loss is also higher than 5%. A general trend of the effect of air oxidation on the mechanical properties of C/Cs is shown in Figure 19 [65]. Similar trends are observed for CVI-based C/Cs [65,69].

CVD pyrocarbons-based C/Cs too exhibit a range of high fracture toughness (20–100 Nm m⁻¹/²) and good creep and fatigue resistance. When subjected to high temperature testing, C/Cs have been found to exhibit about 10–20% increase in mechanical properties at 2000 °C under inert atmosphere. However, in air, the properties drop down to 10–20%, depending on the temperature and time i.e. the weight loss. Carbon is known to be a stable and well-sought material for nuclear application; so are carbon–carbon composite C/Cs. Under neutron irradiation at low fluence level of 10²¹, these composites exhibit an increase in strength and fracture toughness by 20–30% and Young’s modulus by about 30% [9,70]

3.3. Thermal Properties of C/Cs

3.3.1. Thermal Conductivity
C/Cs being heterogeneous structure consisting of fibers, matrix, and pores, with the former two having varieties of microstructure, the estimation of thermal transport
properties of these composites becomes complex, as schematically represented in Figure 20.

Carbon fibers and the matrix may exhibit a variety of microstructures depending on the precursor and processing condition. Obviously, these microstructures will possess varying thermal properties. The thermal transport properties in the direction perpendicular to the fibers are difficult to predict. This is because the heat flow does not follow a simple path. The fibers and the pores serve to disrupt heat flow in the carbon matrix, thus decreasing thermal conductivity. The thermal conductivity data available in the literature generally pertain to products made by the manufacturers. Very few groups have carried out a systematic evaluation of the thermal conductivity of C/Cs. A general comparison of thermal conductivity of C/Cs with different fiber/matrix combination is given in Figure 21 [6]. In

---

**FIGURE 20** Schematic presentation of factors contributing to transport mechanism in C/Cs. (For color version of this figure, the reader is referred to the online version of this book.)

**FIGURE 21** General trends in thermal conductivity of different C/Cs. For color version of this figure, the reader is referred to the online version of this book.
contrast to thermosetting-resin-derived C/Cs, pitch-precursor-derived C/Cs exhibit very high thermal conductivity. This is also evident from Figure 21. The conductivity in the fiber direction increases markedly with the number of densification cycle. The absolute value is further increased on graphitization of the composites at 3000 °C [71,72]. Similar effects of the matrix precursor on the thermal conductivity of C/Cs have been observed in 2D composites [72].

A CVD-derived carbon matrix may exhibit a wide range of microstructures from isotropic to highly anisotropic RL structure. Accordingly, the thermal conductivity of C/Cs with a CVD matrix will also vary with the matrix microstructure and heat treatment temperature (Figures 22 and 23) [71–74]. The Thermal conductivity of heat-treated composites are found to be higher than that of composites, which have not been heat treated and is highest for RL structure. For most of the carbon–carbon products made from CVD carbon matrix, RL-type microstructure is preferred. In CVD-derived C/Cs too, irrespective of the matrix precursor, the thermal conductivity is observed to increase with increasing heat-treatment temperature and decrease with increasing measurement temperature. The decrease in thermal conductivity with increasing test temperature is observed in all C/Cs preheat treated to that temperature irrespective of the fiber type and architecture (Figure 24) [74]. The extent of variation depends on the matrix precursor and fiber/matrix combinations. The effect is maximum for very high conducting C/Cs.

FIGURE 22 Thermal conductivity of CVD C/Cs.

FIGURE 23 Thermal conductivity of carbon–carbon composite made with different pitch-based carbon fibers. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 24 Change in thermal conductivity of C/Cs with test temperature.
As discussed above, high thermal conductivity C/Cs can be produced by choosing the proper combination of reinforcing fibers and the matrix precursor. C/Cs with very high thermal conductivity in the order of 850 w/mK at room temperature have been developed using the ultrahigh-modulus pitch-based carbon fiber P-130 and mesophase pitch as matrix precursor [74] (Figure 25). By using ribbon-shaped mesophase pitch carbon fibers in pitch precursor carbon matrix, C/Cs have been developed with a high thermal conductivity in the directions parallel and perpendicular to the fiber direction [48,75,76]. The higher thermal conductivity of the composites transversing to the fiber direction suggests a considerable contribution by the matrix to the transverse thermal conductivity of ribbon-shaped carbon-fiber-reinforced carbon matrix composites. Current interest has been to use hybrid matrix derived from highly aromatic resins and pitches to develop high density, high conductivity C/Cs [77].

Using highly graphitic vapor-grown carbon fibers with resin/pitch-based carbon matrix followed by CVI, a thermal conductivity of 824 w/mK in the direction of the fibers has been reported. Though these composites exhibit highly anisotropic character and low conductivity in transverse directions, there is always scope of improvement by varying the fiber architecture and their addition in different forms and ways.

3.4. Thermal Expansion of Composites

Graphite is known to be highly anisotropic with pyrographite exhibiting a thermal expansion anisotropy of about 20. Accordingly, carbon fibers also exhibit a Coefficient of thermal expansion (CTE) range that depends on the precursor and heat treatment temperature. In C/Cs, thermal expansion in the direction of reinforcement is mainly controlled by that of the reinforcing fibers, whereas in the direction perpendicular to it, the type and content of matrix and porosity have a major influence. Though coefficient of thermal expansion of the composites is more dictated by the fiber orientation, typically, it is 1 ppm/°C in the fiber direction and 6—8 ppm/°C in a direction perpendicular to the fibers. Further, composites heat-treated to graphitization temperatures exhibit a lower thermal expansion in both directions. The ratio in the two directions increases with the crystalline order of the fibers. The CTE in 3D C/Cs is lower than in UD or 2D C/Cs [6]. Moreover, composites with a high fiber volume content and low porosity will exhibit a higher thermal expansion in the perpendicular direction than those with a low fiber volume content and a higher porosity.

3.5. Thermal Shock Resistance of C/Cs

Thermal shock resistant of a material is an additional required thermal property of material intended for structural application at high temperature, specifically under cyclic varying temperature condition.

Graphite, having a high thermal conductivity and a low CTE, exhibits very high thermal shock resistance. C/Cs, with a much higher tensile strength and thermal conductivity than polycrystalline graphite, exhibit a two to three times higher thermal shock-fracture toughness than polycrystalline graphite [78].

3.6. Frictional Properties of the Composites

Carbon-based materials are widely used in wear-related applications, e.g. bearings, seals, and electrical brushes.
The development of C/Cs has further widened the scope of application of carbon-based materials in wear-related applications from bearing seals and electrical brushes to brake pads for heavy duty vehicles such as military, supersonic, and civilian aircrafts to trucks and railways. The function of the brake is essentially controlled by its ability to absorb energy in the form of heat. Therefore, the thermal characteristics of candidate brake materials are of major importance—a potential high-performance brake material must have a high specific heat, a high melting point, and adequate mechanical properties at elevated temperature. Carbon fibers along their length are expected to have low wear and good frictional characteristics. Therefore, the early carbon—carbon disks were made with layers of carbon fabric or with random fibers arrayed parallel to the braking surface. C/Cs exhibit low coefficient of friction $\mu$ 0.3–0.5 in the fiber direction and 0.5–0.8 in perpendicular direction. Wear rates also follow similar trend (0.05–0.1 and 0.1–0.3 mm). The friction and wear mechanism of C/Cs on application of brakes is quite complex, and various factors like peak temperature and formation of debris and films on sliding surfaces further affect the coefficient of friction, etc. which are too vast to cover here. This results in the formation of ground-wear debris, containing flake-like turbostratic, less-ordered carbon than the base composite. The friction surfaces are characterized by dull, unpolished wear surface. The wear debris produced under these conditions has a higher degree of order than the taxiing debris. The particulate debris does not lower the coefficient of friction, whereas film-type debris acts as a solid lubricant, lowers the coefficient of friction, and reduces the wear rate [79–83]. Resin-derived CVD and pitch-derived carbon matrixes heated to same high temperature not only posses different graphitic contents and crystalline ordering, leading to variation in thermal conductivity, but also vary in structural characteristics, i.e. fiber/matrix bounding. Accordingly, during a braking operation these composites are heated up to different temperatures, which are highest for resin-derived low-conducting carbon and lower for CVD and pitch-derived highly ordered, high thermal-conducting carbon matrixes. The average friction coefficient is to increase with decreasing initial velocity. The friction coefficient decreases with increasing pressure. According to Kimura et al. [84], the coefficient decreases with increasing Young’s modulus of the composites. This implies that the composites should be heat-treated at higher temperatures than usual. The frictional characteristics are also found to be influenced by the environment. In a wet environment, the rise in friction surface temperature is low and so is the coefficient of friction. Not only the carbon fiber and composite characteristics but also the inorganic additive, if added to the matrix of the composites, influences the friction characteristics of the ultimate composites.

In commercial carbon—carbon brake pads, carbon fiber-fabric plies are used with fiber tows inserted in the third direction and pitch/CVD route for densification. Figure 26 shows comprehensive data on tribological properties of C/Cs [6]. However, the choice of the constituents and processing parameters is propriety of the manufacturers.
4. OXIDATION PROTECTION OF C/Cs

Carbon, depending on graphitic contents, begins to oxidize in air at temperature around 400 °C, the rate of oxidation increasing rapidly with temperature. The rate-controlling factors for oxidation of C/Cs at high temperatures (600–700 °C) are the diffusion of the oxygen through the boundary layers surrounding the components and outward diffusion of reaction products (CO or CO₂). For long-time application of these composites at elevated temperature under normal environment, it is essential that these composites, in the true sense, all components of the composites: fibers, matrix, and interface, be protected against oxidation. Therefore, studies on oxidation protection of C/Cs is as important as the development of the composites themselves. Therefore, oxidation protection is an important aspect of carbon–carbon technology. Studies on oxidation of C/Cs and their protection started concurrently with the development of these composites. Many types of coatings have been developed for protection of C/Cs. Various approaches that have been described in the open literature for protection of C/Cs are summarized in Table 7. Oxidation protection systems for C/Cs are based on (i) to block the carbon active sites and slow down the oxidation rate through modification of the matrix by addition of oxidation inhibitors (like P, B, Si, Zr or their compounds) and/or (ii) application of diffusion barrier coatings and overcoats to prevent oxygen ingress and carbon egress. However, most of the current systems of oxidation protection use several of these methods in a compatible or complimentary manner. Needless to say, the type of protection system required is dictated by the end application of the C/C.

The rate of carbon gasification at low temperature can be effectively retarded by decreasing the porosity and blocking or poisoning the active sites by elements, which have an inhibiting effect on oxidation, such as boron, phosphorus, and silicon. Phosphorus compounds alone or in combination with halogens are found to have very strong inhibiting effects. The inhibition effect of halogens on the oxidation behavior of carbon appears to be confined to temperature below about 900 °C. At higher temperatures, the halogens atoms are desorbed from the carbon surface, and their protective effect is thereby lost [84–86].

Of all the materials used as oxidation inhibitors for C/Cs, boron in one form or another, i.e. either elemental boron or boron compounds, have been found to be the most appropriate materials for oxidation protection at temperatures less than 1500 °C. Phosphates in combination with other inhibitors such as silica, silicon carbide, and boron oxide have been found to reduce the oxidation of carbon at 1500 °C to one tenth. Mixtures of metallic borides such as those of hafnium, zirconium, elemental boron, silicon carbide, and other glass-forming particulates are also often added as inhibitors in the matrixes during the formation of C/Cs. On exposure to oxygen, these additives form mobile borates or borosilicate glasses and protect the C/Cs from oxidation.

Silicon-based ceramic coatings, due to their suitable thermophysical properties, have been quite attractive for carbon–carbon oxidation protection at temperature exceeding 1000 °C for extended periods. Various techniques have been developed to coat C/C with silicon-based materials [87]. However, excessive cracking, especially during cooling, may become a path for the ingress of oxygen. Therefore, multilayer approaches have been adopted wherein the glass layers between the outer layer and the substrates heal the cracks and protects the carbon fibers and the matrix. Numerous patent reports describe multilayer coatings consisting of Mo, Zr, Ta, Cr, Al₂O₃, silicon, boron, etc. as such or in the form of their inorganic compounds [88]. Figure 27 shows a general scheme of such a multilayer coating system. A permutation/combination of the elements and their compounds mentioned in Table 8 can be selected.

5. APPLICATIONS OF C/Cs

As is evident from the previous sections, C/Cs of the desired shape with properties required for particular application can be produced by meticulously choosing the type, architecture, and amount of carbon fiber and matrix precursor, and the processing conditions. Therefore, C/Cs are considered to be a class of materials with a wide spectrum of properties and applications. Moreover, the ongoing development of high-performance carbon fibers, high char-yielding synthetic resins, and pitches continuously add to the spectrum of carbon–carbon properties and products. It is with these milestones, C/Cs, which were developed in 1958 and nurtured under the US Air Force space plane program, Dyna-Soar, and NASA's Apollo projects to meet the anticipated needs of the emerging space shuttle program, are regarded nowadays as a logical addition to the

<p>| TABLE 7 Methods for Oxidation Protection of C/Cs |</p>
<table>
<thead>
<tr>
<th>Inhibitors and Sealants</th>
<th>Barrier Coatings and Overcoats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halogens</td>
<td>Noble metals</td>
</tr>
<tr>
<td>Phosphorus compounds</td>
<td>Borides</td>
</tr>
<tr>
<td>Boron compounds</td>
<td>Carbides/nitrides</td>
</tr>
<tr>
<td>Polysiloxanes</td>
<td>Silicides</td>
</tr>
<tr>
<td>Silicon</td>
<td>Intermetallic compounds</td>
</tr>
<tr>
<td>Borate and silicate glasses</td>
<td>Engel/beewer compounds</td>
</tr>
</tbody>
</table>
member of fine-grained carbon and graphites. C/Cs offer a large potential as high-performance engineering material. Therefore, in addition to the special defense, aircraft, and spacecraft applications, a steady interest is also growing in civil market segments. Based on the worldwide consumption, a general trend of carbon—carbon consumption can be discerned as shown in Figure 28. In terms of mass consumption and money, the main applications of C/Cs today are still in military, space, and aircraft industry, while those in the field of general mechanical engineering are picking up steadily. In the following sections, general applications of C/Cs are arranged according to the various specific properties of the composites. There is great potential for the wider application of these materials.

5.1. C/Cs as Brake Discs
Concerted research and development efforts in the design and manufacture of carbon—carbon brake systems led to

<table>
<thead>
<tr>
<th>Element</th>
<th>Compound</th>
<th>Major Function/Protection Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>Pyrolytic graphite</td>
<td>Substrate adhesion/protection from chemical attack</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>Phosphate</td>
<td>Low-temperature oxidation protection</td>
</tr>
<tr>
<td>Boron</td>
<td>Alone</td>
<td>Substrate adhesion</td>
</tr>
<tr>
<td>Boron</td>
<td>Boron carbide</td>
<td>Substrate adhesion</td>
</tr>
<tr>
<td>Boron</td>
<td>Boron nitride</td>
<td>Oxidation-inhibiting matrix addition</td>
</tr>
<tr>
<td>Boron</td>
<td>B₂O₃</td>
<td>Sealants</td>
</tr>
<tr>
<td>Boron</td>
<td>Borates</td>
<td>Oxidation inhibition</td>
</tr>
<tr>
<td>Silicon</td>
<td>Silicon carbide</td>
<td>Substrate adhesion</td>
</tr>
<tr>
<td>Silicon</td>
<td>Silicon nitride</td>
<td>Oxidation protection</td>
</tr>
<tr>
<td>Silicon</td>
<td>SiO₂</td>
<td>Oxygen barrier</td>
</tr>
<tr>
<td>Iridium</td>
<td>Alone</td>
<td>Carbon-egression barrier</td>
</tr>
</tbody>
</table>

FIGURE 27 Generalized schematics of multilayer coatings for oxidation protection of C/Cs.

FIGURE 28 General market trends for C/Cs.
the widespread use of commercial C/Cs aircraft wheel brakes from military aircrafts [89] to civilian and commercial aircrafts with improved performance. The potential for growth in carbon–carbon brakes market is further underlined by their uses in racing cars, heavy-duty surface transport systems, and high-speed rail system [90]. For surface transport systems, the temperature required for brake operation is quite low (400–600 °C), and therefore, relatively cheap designs and normal processing methods of 2D C/Cs can be utilized for such applications. Figure 29 shows a photograph of carbon–carbon disk for an automotive brake.

5.2. Carbon–Carbon Materials for the Aerospace Industry

The aerospace field continues to be one of the primary areas of application of C/Cs. The classic examples are the thermal protection system of space shuttles, rocket nozzles, exit cones, nose tips of reentry vehicle, etc. These components provide thermal protection to the instrumentation from the searing heat of reentry and maintain structural integrity over multiple missions, especially under condition of subzero temperatures in the outer space to close to 1700 °C during entry back into the atmosphere. C/Cs for such applications are required to process high thermal conductivity and high specific heat so that the component operates as a heat sink, absorbing the heat flux without any problem. Advanced C/Cs with silicon carbide and ceramic multilayer coating for oxidation protection are utilized in the European space plane, Hermes and National Aerospace Plane (NASP), with anticipated temperature of 2750 °C on the nose cap and approximately 1950 °C on the wing and tail-leading edges. In Rocket motors with a solid propellant system, the throat and exit cone are made of C/Cs. Although these materials are only for short time in operation, the property requirements are extremely stringent.

5.3. Aeroengine and Turbine Components

The choice of C/Cs in jet engine rotors and stators offers the possibility of operating at temperature well in excess of 500–600 °C higher than those used in conventional engines with high-temperature alloys. The basis of this development is the utilization of the high tensile strength of the fiber materials, which allows the compensation of the centrifugal stress on the blades.

5.4. C/Cs as Refractory Materials

With carbon-fiber reinforcement, the mechanical properties of these carbon refractories can be further improved, thereby opening up new fields of industrial applications.

5.5. High-Temperature Fasteners

The strength and stiffness at high temperature of C/Cs guarantee high self-fastening stability at high
temperature. Therefore, carbon–carbon bolts, screws, nuts, and washers (Figure 30) are used where high temperature and severe chemical condition are present [91,92].

5.6. C/Cs in the Glass Industry
Carbon and Graphite are not wetted by molten glass. Therefore, carbon–carbon materials are used as non-wetting crucibles for molten metals and in various parts of glass-container-forming machines [6]. They are used in channeling systems to carry the gobbets of molten glass, as molds for crystal glass products, and as an asbestos replacement for hot-end glass contact elements for moving hot glassware articles (Figures 31 and 32).

5.7. Emerging Applications
The application and market for C/Cs are continually developing for different sectors. The latest applications of C/C structural materials are in advanced thermal management systems. These materials have potential for mechanical engineering applications in hot-load-bearing structure and corrosive surroundings, as substrates for electronics applications.

The high electrical conductivity of airborne graphite particles creates an unhealthy environment for electrical equipment near machining areas. This aspect will be always kept in mind while processing, machining, and selecting application of C/Cs for general industrial applications.

FIGURE 30 Carbon–carbon nuts and bolts.

FIGURE 31 Carbon–carbon heating elements.

FIGURE 32 Carbon–carbon channeling for carrying gobbets of molten glass (a,c) and molds for crystal and lead crystal drinking glasses (b).


REFERENCES

[74] Burchell TD, Oku T. Materials property data for fusion reactor plasma facing C/Cs, preprints (personal communication)
[90] Klein AJ. Mater process 1986;130:64.
[91] Carbon fiber reinforced composites Product Brochure, Schunk Kohlenstoff, Germany.
1. INTRODUCTION

Fuel-cell power technology is considered to be one of the most promising new power-generation technologies owing to its various merits, such as high power efficiency and excellent compatibility with the environment [1]. Fuel cells can be broadly classified into phosphoric acid fuel cells (PAFCs), molten carbonate fuel cells, polymer electrolyte fuel cells (PEFCs), and solid oxide fuel cells. Car manufacturers and electronics companies are competing to develop PEFCs because of their advantages such as small size and weight, high power density for normal temperature operation, and excellent start-up capability.

Carbon is an essential component material in PEFCs. This chapter, while focusing on the separator, electrocatalyst, and proton conductor, reviews how the research on carbon materials is involved in improving the performance of PEFCs.

2. OPERATING PRINCIPLES AND CHARACTERISTICS OF PEFCs

A PEFC comprising a proton-conductive polymer electrolyte membrane of thickness 50–180 μm gives an all solid-state structure consisting of gas diffusion electrodes made from carbon supporting platinum on both sides of the membrane (Figure 1). Hydrogen supplied to the anode side is oxidized, and produced H\(^+\) moves through the membrane to the cathode side. This H\(^+\) is used in the reaction in which water is formed by the reduction of oxygen supplied to the cathode side. As the electrons generated at the anode flow through the external circuit to arrive at the cathode, an electric current can be retrieved. The reactions at the anode and cathode, and the overall reaction, are given as follows.

\[
\text{Anode: } H_2 \rightarrow 2H^+ + 2e^- \\
\text{Cathode: } \frac{1}{2}O_2 + 2H^+ + 2e^- \rightarrow H_2O \\
\text{Overall reaction: } H_2 + \frac{1}{2}O_2 \rightarrow H_2O
\]

Pure hydrogen or hydrogen reformed from methanol or natural gas is used as fuel. In addition, the direct methanol fuel cell (DMFC) is considered where methanol is supplied directly to the anode instead of hydrogen to produce power.

The most typical polymer electrolyte membrane is the perfluorosulfonic-acid-type ion exchange membrane such as Nafion®, which has a molecular structure based on a main polytetrafluoroethylene chain with side chains containing sulfonic acid. PEFCs operate at temperatures below ca. 90 °C. The high power density of PEFCs is largely attributable to the high ionic conductivity of the polymer electrolyte membrane at these temperatures. However, because the polymer electrolyte membrane is a strong acid, the materials that can be used in the vicinity of the electrolyte are limited. A polymer electrolyte membrane containing water has acidity comparable to a solution with 10 wt% sulfuric acid. The anode side is a hydrogen ambient atmosphere, while the cathode side is
air. Because of humidification and water produced by the reaction, the reactive gas contains saturated water vapor. In addition, because the incorporation of impurity ions in the polymer electrolyte membrane lowers the performance, the use of acid-soluble materials should be avoided. With these strong constraints, fine particles of precious metals such as platinum are used for the electrocatalyst, whereas for the catalyst supports, current collectors, and separators, carbon materials with high electrical conductivity and high chemical stability are used.

3. SEPARATORS

The PEFC stack mainly consists of the membrane-electrode assembly (MEA) and separators, as shown in Figure 2. There are grooves on the surface of the separators that allow for the supply of the reactive gases of hydrogen and oxygen. This member is called a separator because of its role as a partition; thus, the reactive gases supplied to neighboring single cells in the stack do not mix. It is sometimes called an interconnector because of its function of electrically connecting neighboring cells in series. It is also called a bipolar plate or bulkhead plate.

(1) Separator Classification

Separators can broadly be classified into metallic and carbonaceous. When considering cost, weight, corrosion resistance, and machinability, the metallic separator is inferior to the carbonaceous separator.

Given the carbonaceous separator, glasslike carbon was the material of the first commercialized separators. It was used as a separator for the PAFC. Although its gas impermeability was highly sufficient, there were difficulties in the electrical conductivity and machinability. They are expensive and not currently used. Consequently, separators made of resin-impregnated isotropic graphite were used. They are still used as single cells in research. Molded carbon resin products are now the mainstream of the separator material in the PEFC stack. This material can be divided by classification of the forming method (compression molding and injection molding) or a variety of resins (thermosetting resin and thermostable resin). In any case, the separator costs approximately over 90% of entire PEFC costs, and it is thus desirable to develop high-performance separators that can be mass produced at low cost.

(2) Molded Carbon Resin Separators

For the serpentine flow field of the reactive gas engraved on the separator, both the groove width and depth are about 1 mm. Furthermore, the gas inlet and outlet construction and the flow channel for the cooling water must be set up. After ensuring the machining accuracy of these fine structures, it must withstand repeated expansion and contraction due to the difference between the operating temperature of ca.90 °C and room temperature, the swelling shrinking stress of the electrolyte membrane attributable to operation stoppages and a certain amount of expected vibration when mounted on a vehicle.
Based on the study of the manufacture of the molded carbon resin separator that overcomes such conditions, the following methods are currently recommended: the ‘compression molding using thermosetting resin’ and the ‘injection molding using thermoplastic resin’. The former has the disadvantage of time requirements as high as 10 min to produce a single separator. On the other hand, the latter has disadvantages such as high electrical resistance and inadequate gas permeability, although it is possible to produce a single separator in about 1 min.

Incidentally, the details of the resin used in molded carbon resin separators are not disclosed by the manufacturer.

(3) Carbon Coating for Metal Separators

Metallic separators are effective for slimline modeling and are excellent in terms of conductivity and reliability. Accordingly, carbon coating has been investigated in order to increase the corrosion resistance under the operating conditions of a fuel cell. As methods of carbon coating, the following have been tested: press molding onto a metallic substrate of materials similar to those used in the above-mentioned molded carbon resin separator [2], carbonizing an applied polymer by spin coating, amorphous carbon coating by plasma sputtering [3], and so on. In all cases, the coating is carried out after removal of the oxide film on the metal. The evaluation findings for the corrosion resistance and conductivity of the finished product are favorable, although it is still in the prototype stage.

4. CURRENT COLLECTORS

To make electrical contact between the above-mentioned separator and the electrocatalyst described later, carbon fiber has been used as the current collector. Carbon fibers are broadly categorized as carbon paper and carbon cloth, and are selected depending on the user’s choice. For example, from the E-TEK Division of PEMEAS USA Inc. Fuel Cell Technologies, which sells fuel cell parts, various products of carbon paper or cloth treated with the desired degree of water repellent can be purchased. Academic interest related to carbon materials is scarce with respect to the dependence of current collectors on the power-generation characteristics; therefore, it will not be discussed further.

5. ELECTROCATALYST

(1) The Role of the Carbon Support

The electrode reaction takes place at the so-called three-phase interface (platinum—electrolyte—reactive gas). In the case where the electrolyte is a solid-state membrane, the reaction location is limited to the contact interface between the platinum electrocatalyst and the electrolyte. Therefore, the performance of PEFCs is affected by not only the amount and surface area of the platinum used as catalyst but also the availability of the platinum actually taking part in the reaction. For example, in order to increase the availability of the expensive platinum, there...
is a need to decrease platinum particles that cannot contribute to the reaction because of no contact with the electrolyte or reactive gas. By choosing a carbon support, the supported platinum amount, surface area, and availability can be controlled. Therefore, it is no exaggeration to say that the PEFC performance is determined by the carbon support. If a superior carbon support can be developed with characteristics that can activate the electrocatalyst, it is possible to reduce the amount of platinum that should be used. PEFCs can become cheaper when the amount of platinum used is small, and the problem of depletion of platinum resources can be resolved if its use becomes widespread.

Because of the high specific surface area of carbon black, the platinum particles can be supported on carbon black by using various methods, such as the impregnation method, preparation using colloidal particles, or preparation by ion exchange. With carbon black as a support, the microparticulation of platinum is made possible. Accordingly, the activity per unit mass of platinum can be increased. In addition, given the diffusion of the reactive gas within the electrode or the distance of ionic conduction, it is better for the electrocatalyst layer to be thinner. In such a case, the supported density of platinum must be increased with maintaining the highly dispersed state. For example, the loaded amount and particle diameter of platinum supported on carbon black, Vulcan® XC-72, as well as the platinum surface area are summarized in Table 1 [4]. Here, XC-72 is the carbon black which is widely used in the research of electrocatalysts. There have been many reports on carbon supports with better performance compared to the reference standard XC-72. Because the specific surface area of ca. 250 m²/g is not so high, Ketjen Black with an increased specific surface area is sometimes used. In Figure 3 [5], it is shown that the mass activity (MA) normalized by weight of platinum is dependent on the specific surface area of the support, and this shows the effect of increased surface area of platinum as the particle diameter became smaller.

The basic idea behind the carbon support is roughly the same, whether it is the anode or the cathode. However, since the chemical species of the supported fine particles of precious metal are sometimes different, and, naturally, the electrode reactions, oxidation, and reduction are completely different, these are explained separately below.

(2) Cathode

With respect to the fine particles of platinum, which act as the catalyst, the increase in the MA of platinum with the decrease in the particle diameter was referred to above. In contrast, there are many reported cases that the activity per surface area of platinum (specific activity) decreases as the particle size decreases. Consequently, there would be an optimum value for the platinum particle diameter considering the activity for both per unit weight and per unit surface area of platinum, but to clarify the effect of particle size would require the collection of further data. In addition, the particle diameters 3–10 nm are sufficiently accessible when the platinum is supported 30–70% using carbon black as a support. Again, catalytic activity is determined by the particle diameter and surface area of the platinum catalyst, and the platinum particle diameter and surface area are highly dependent on the carbon black support.

Electrocatalytic activity is also affected by the reactive site and mass transfer inside the porous gas diffusion electrode that is determined by the form of the carbon black. First, as shown in Figure 4, for the case where dibutylphthalate (DBP) absorption is small, the cell voltage decreases [5]. The pores inside the electrode have the function of diffusing oxygen and expelling water, but they are absent when DBP absorption is small and the performance has deteriorated. However, from performance assessments of cells using various kinds of carbon black
(Figure 5), it is shown that primary-pore-rich carbon black is not suitable as a support [6], for polymer electrolyte cannot intrude into the pores. In this manner, the effect of the structure of carbon black on the electrocatalytic activity cannot be explained easily, and further studies are awaited.

(3) Anode

There is no problem when pure hydrogen is used as fuel, but when hydrogen is obtained by reforming methanol or natural gas, the residual carbon monoxide (CO) is strongly adsorbed onto the platinum catalyst surface, and this leads to a large drop in catalytic activity. To prevent this poisoning, the method of adding metals of group VIII, such as ruthenium, has been widely used.

The particle diameter and specific surface area of the platinum catalyst, which contains ruthenium, are similarly dependent on specific surface area of the support and show a trend similar to platinum alone. It is therefore advantageous for the specific surface area of the support to be large for the electrode oxidation activity of pure hydrogen. In addition, results of a previous study [5] show that it is advantageous for CO poisoning tolerance to increase the specific surface area of the carbon black support, which allows the platinum–ruthenium specific surface area to also increase, while there are also opinions that if the specific surface area of the support is high, the platinum–ruthenium alloying cannot be done well, and the poisoning tolerance is not sufficient. It is therefore necessary to comprehensively weigh up these properties to determine the desired carbon black.

(4) Durability

In PAFCs, aggregation of the platinum catalyst particles occurs because of long periods of power generation; hence, a drop in performance is said to occur. This happens probably because of the comparatively high operating temperature of ca. 200 °C. Although PEFCs are operated at most below 100 °C, similar catalyst particle growth has occurred [5], and catalyst particle growth is detected particularly at the cathode side. Furthermore, at the anode side, the degradation of CO poisoning tolerance has also been observed. However, the details of the relationship between durability and the carbon black support have not yet been clarified, and, at present, research is proceeding [7].

(5) Development of a New Carbon Support

As mentioned above, carbon black is used as a support for the electrocatalyst used in PEFCs, but research is also underway using other carbon materials aimed at trying to make performance improvements. Carbon nano-horns discovered by Iijima et al. [8] have a conical structure with a single layer of graphene with a diameter of 2—4 nm and a length ca. 50 nm. This carbon nano-horn is obtained in the form of numerous spherical aggregates, with the platinum particle becoming as tiny as about 2 nm when the platinum is loaded on the aggregates. It has also been found that a uniform dispersion on the carbon nano-horn aggregate is possible; therefore, a superior electrocatalyst is being developed for DMFC use.

On the other hand, many studies of new electrocatalyst supports have reported using carbon nanofibers manufactured by the arc discharge, chemical vapour deposition (CVD), and template methods [9—22]. All of these reports conclude that when carbon nanofiber is used as a support, a superior electrocatalyst is obtained. For example, Joo et al. [13] showed that the electrocatalyst with nano-carbon prepared by the template method provides a good oxygen reduction MA (Figure 6). One reason for this is that the supported platinum particles are small and have uniform particle diameter in the same way as the above-mentioned carbon nano-horns. Furthermore, other studies demonstrate...
other reasons why carbon nanofiber is a superior electrocatalyst support: (1) the high electrical conductivity of carbon nanofiber [9,14]; (2) the presence of only few impurities, such as sulfur, compared to carbon black [9]; (3) the fact that the morphology of supported platinum is different for carbon black [9,15]; and (4) some kind of interaction between platinum and nano-carbon [14,15]. Detailed explanations could be considered during a future task.

Despite having a high specific surface area similar to carbon black, active carbon is not used as a catalyst support for PEFC. This is due to the low electrical conductivity of carbon. Furthermore, even if the platinum is supported in the interior of the micropores and sub-micropores of the active carbon, ions cannot reach there via the polymer electrolyte. Maruyama et al. have reported that if trifluoromethane sulfonic acid is adsorbed onto the micropores, it facilitates ion transfer and hence catalytic activity increases (Figure 7) [23].

Many papers have elucidated the carbon support dependence on the catalytic activity. When the catalytic activity was compared for the vapor-phase oxidation of hydrogen, methanol, and formic acid, marked dependence on the carbon support was noted [24]. In the case of electrochemical oxidation of methanol with platinum—ruthenium particles supported on mesocarbon microbeads, a higher activity than an electrocatalyst with carbon black support was reported [25,26]. There are also reports that hard carbon spheres [27], graphite intercalation compounds [28], and graphene sheets obtained by reducing oxidized graphite [29] are excellent supports.

There are only few cases where the causes of these support effects have been investigated. However, in reports relating to the electron spin resonance (ESR) measurement [30] and the analysis of the electrical double layer [31], the degree of electron transfer between platinum and carbon is different, depending on the type of carbon. Consequently, the electron density on platinum is also different.

Additionally, through the pretreatment of carbon used as a support, attempts are being made to improve the electrocatalytic activity. Activity improvements for methanol oxidation were reported when acid functional groups were added to the surface of carbon fiber as catalyst support [32]. Activity improvement of platinum has also been noted by adding fluorine functional groups [33].

(6) Carbon with Its Own Electrocatalytic Activity

An important research field for PEFCs is the development of alternative materials to platinum. Instead of enhancing the electrocatalytic activity by choosing the kind of carbon supports and reducing the amount of platinum which have been stated so far, this research aims to develop other active substances without using platinum. In the past, the use of every kind of metal complex and metal oxide has been attempted. Regarding carbon materials, heat treatment products of cobalt macrocyclic complex were found to have a catalytic activity in the oxygen reduction reaction [34], and research is underway with carbonized products of similar complexes.

Of the central metal, the use of group VIII transition metals, such as iron, cobalt, or nickel, is very common. For the cobalt complex, carbonized products of phthalocyanine, porphyrin, or other types of complex have been reported. For the onset of oxygen reduction, both cobalt metal and nitrogen are important factors.

For the iron complex ligands, there are many reported cases of not just the phthalocyanine and porphyrin macrocyclic ligand but also ligands containing the phenanthroline [35] or cyan group [36]. Interestingly, the ligands in these metal complexes contain nitrogen atoms as well. Furthermore, with regard to the heat treatment temperature during carbonization, it is experimentally known that catalytic activity of treated samples is high in the region from 600—900°C. It has been anticipated that, due to the heat treatment, some kind of arrangement of iron and nitrogen atoms will form an active site. If the heat
treatment temperature is too high, aggregation of the iron takes place and the activity will probably drop. Furthermore, Maruyama et al. showed that, in light of the fact that the carbonized products of hemoglobin and catalase possess high catalytic activity in oxygen reduction, the carbonized products of red blood cells are also promising as an electrocatalyst in oxygen reduction [37]. These can be classified in the category of complexes that contain iron and nitrogen atoms. It is also interesting from the standpoint of waste utilization during meat manufacture.

In recent years, it has been reported that some carbon materials with nitrogen atoms possess high oxygen reduction activity even in the absence of metal atoms. When nitrogen is introduced to the surface of carbon black with the ammonia treatment at ca. 600°C [38], oxygen reduction activity appears in the carbon itself. Further improvement in oxygen reduction activity has been reported by doping with boron and nitrogen [39].

Ozaki et al. assumed that the onset of oxygen reduction activity is dependent on the shell structure of the carbon surface. Unusual shell structures are produced when macrocyclic complexes of iron, cobalt, and nickel are carbonized, and this structure is assumed to be effective in oxygen reduction activity. This activity seems to be significant if nitrogen and boron atoms exist at an edge site. The results of the power-generation performance assessment of an MEA made without using platinum are shown in Figure 8 [40]. The performance is low compared to the case where platinum is used, but it is understood that reasonable values have been obtained.

6. PROTON CONDUCTOR

(1) Fullerene-based Proton Conductor

In recent years, it has been understood that the fullerene derivatives behave as proton conductors, and research directed at their application for PEFCs has been started. Proton conduction of fullerene derivatives was reported for the first time in 2001 [41]. C_{60}(OH)_{n} (n = 10–12) is pressed at 500 MPa to manufacture a pellet of 15-mm diameter, giving the impedance characteristic of Figure 9. From the facts that the sample electrical resistance value was 10^{14} Ω cm and the time constant in the high-frequency region was approximately 10^{-3} s, proton conduction can be concluded. In the absence of moisture, the proton conductivity of C_{60}(OH)_{n} is 7 × 10^{-6} S/cm (295 K), which shows that the estimated proton dissociation constant pK_{a} is close to the value for phenol. On the other hand, C_{60}(OH)_{24} hardly shows any proton conduction due to its weak electron affinity.

In order to ensure that protons pass through as a form of H^{+} and that H_{2}O^{+} is not involved, the time dependence of the proton conductivity was observed using samples in dry conditions. From the fact that the proton conductivity was constant, as shown in Figure 10, H^{+} is proved to be responsible for proton conduction. From the temperature dependence of proton conduction shown in Figure 11, activation energy of 0.33 eV was obtained. This value is an order of magnitude higher than the proton conductivity for Nafion, which predicts that the basis of proton conduction is hopping conduction.

Powder X-ray diffraction shows that the crystal structure of C_{60}(OH)_{n} (n = 10–12) is similar to that of C_{60}. In other words, the OH group is considered to exist in the gap
where the neighboring C_{60} framework is not connected. With such a crystal structure, it is not easy to explain the hopping conduction, and we are eager to anticipate the findings of further research.

As we know that there is proton conductivity in fullerene hydroxide, the hunt for fullerene derivatives possessing even higher conductivity has started. For example, Li et al. showed that the proton conductivity of C_{60} > C[PO(OH)]_{2} (in the formula, > means a C cross-linking bond between adjacent carbon atoms in C_{60}), under the conditions of relative humidity (RH) at 95% and at 25 °C, was 0.01 S/cm [49]. As shown in Figure 12, the proton conductivity is strongly dependent on the RH. Under dry conditions, C_{60}(OSO_{3}H)_{x}(OH)_{y} (x ~ 6, y ~ 6) and C_{60}(OSO_{2}O)_{z} →(OSO_{3}H)_{2}(OH)_{z} (Z < 1) show high proton conductivity, and especially the former is reported to have a proton conductivity of 0.01 S/cm [42].

(2) Fullerene–Nafion Composites

One trend in attempting to use fullerene-based proton conductors as proton conduction membrane is fullerene composites with Nafion. Tasaki et al. formed the composite by immersing Nafion 117 in C_{60} toluene solution or C_{60}(OH)_{12} tetrahydrofuran (THF) solution, and reported its characteristics [43]. As shown in Table 2, the water content of the composite is greater than simple Nafion, and the water retention characteristics were found to be excellent (Figure 13). Because of this retained moisture, the proton conductivity was improved in conditions where the RH is low, as shown in Figure 14.

The interaction between fullerene derivatives and Nafion is not yet fully understood, but by computer simulation it has been possible to confirm the affinity between the sulfonic group in Nafion and the OH group in PHF: C_{60}(OH)_{12} [43]. The simulation concluded that the fullerene derivative is present in the proton channel of Nafion. The existence of fullerene derivatives with proton conductivity even under dry conditions in the proton channel of Nafion gives a high proton conductivity of the composite in conditions of low RH.

While the affinity between fullerene derivatives and Nafion was discussed, the aggregate of the fullerene derivative was observed with an optical microscope in the fullerene derivative–Nafion composite [43,44]. To avoid the aggregation, a special fullerene derivative C_{60}[CH_{2}C_{6}H_{4}(OCH_{2}CH_{2}O)_{3}OCH_{3}]_{m}, which gives an extremely improved dispersion characteristic, was used [44]. A Nafion composite with either HC_{60}(CN)_{3} or C_{60}(TEO)_{5}, which has the cyano or ethylene oxide group, gives improved dispersion characteristic and proton conductivity [45]. Some of the results are shown in Table 3.

As mentioned above, it was found that the proton conductivity for the fullerene derivative–Nafion composite was high even under conditions of low RH, but its application for the fuel cell has not been reported.

<table>
<thead>
<tr>
<th>Water uptake</th>
<th>Nafion 117</th>
<th>C60/Nafion117</th>
<th>PHF/Nafion 117</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet</td>
<td>25.5</td>
<td>27.4</td>
<td>27.1</td>
</tr>
<tr>
<td>Dry</td>
<td>4.3</td>
<td>5.5</td>
<td>4.7</td>
</tr>
</tbody>
</table>
When preparing the MEA, Shioyama et al. added C$_{60}$(OH)$_{10}$ or C$_{60}$F$_{36}$ to the Nafion ionomer in order to increase the performance at the electrode reactive site on the interface between the electrocatalyst and the Nafion 117 membrane[46]. As shown in Figure 15 and Table 4, an improvement in power-generation characteristics was acknowledged in all cases. Here, the effect of C$_{60}$(OH)$_{10}$ was meager, for the sites where C$_{60}$(OH)$_{10}$ exist are restricted to the vicinity of the electrocatalyst, and was not able to affect the proton conductivity of the whole membrane of Nafion 117.

On the other hand, there is a marked improvement in power-generation characteristics with the addition of C$_{60}$F$_{36}$. As proton conduction could not be expected with the addition of C$_{60}$F$_{36}$, the improvement in power-generation characteristics leads to the conclusion that there is a structural change in the proton channel. It is known that Nafion comprises a perfluoro backbone and the side chains a sulfonic group. As a result, the rod-like structures of the sulfonic group pointing outwards are taken off when Nafion is dispersed with a polar solvent such as water or alcohol. When C$_{60}$F$_{36}$ is added to this Nafion ionomer solution, the C$_{60}$F$_{36}$ molecule is found to exist not in ethanol or in the vicinity of the sulfonic group but rather in the Nafion aggregate, which has been revealed by analysis using photochemical methods[47]. These results can arise from the fact that C$_{60}$F$_{36}$ is not able to interact with the [Ru(bpy)$_3$]$^{2+}$ which exists in the vicinity of the sulfonic group as shown in Figure 16. In other words, in the case where the MEA is made by adding C$_{60}$F$_{36}$ to Nafion ethanol solution, an interaction between Nafion perfluoro backbone and C$_{60}$F$_{36}$ is thought to have affected the Nafion proton channel structure in the vicinity of the electrocatalyst and to have improved the power-generation characteristics. The details of the change in the proton channel structure are not understood, and it merits further research.
Fullerene-based MEA without Nafion

In the preceding section, proton conduction of fullerene derivative–Nafion composites was reviewed, whereas fullerene-based MEA without Nafion was attempted in fuel cell generation.

Hirakimoto et al. have synthesized a proton conductor where seven \( \text{C}_2\text{F}_4\) \( \text{O} \) \( \text{C}_2\text{F}_4\) \( \text{SO}_3\text{H} \) groups are appended to each \( \text{C}60 \), and the neighboring fullerene derivatives are bonded to each other by \( ((\text{CF}_2)_8)_n \). The MEA with this fullerene derivative gives good performance as shown in Figure 17 [48]. This performance is the same as when using Nafion as a proton conductor. With the low amount of methanol crossover, which is inferred from Figure 18 in view, future developments can be expected.

### 7. CONCLUSION

This chapter reviewed carbon materials involved in studies of the dissemination of the practical use of PEFCs. The association between PEFCs and carbon materials is profound. Developments are expected in a complementary manner between these two study fields in the future.

### REFERENCES


1. ELECTRIC DOUBLE-LAYER CAPACITORS

Electric double-layer capacitors (EDLCs) using porous carbon electrodes have a long history [1], with the first patent appearing in 1957. In the 1970s, EDLCs were commercialized in memory-backup devices and the market was established with small-size devices having capacitances approximately up to the 10 F range. Conversely, medium- and large-size EDLC cells with a capacitance of 100–1000 F are anticipated for future applications such as power assistance in hybrid electric vehicles, and load leveling for pulsed current change in power generators using renewable energy, in which the storage device is required to work with a large and instantaneously fluctuating electric current flow. In comparison with secondary batteries, which work via electrochemical reactions between electrode materials and electrolyte, EDLCs are superior in such applications due to intrinsic characteristics such as high power output and long durability.

2. GENERAL ASPECTS OF CARBONS FOR EDLCs

EDLCs are electric energy storage devices in which the electric double layer between a polarizable electrode surface and an electrolyte stores electric charge. Since the amount of electric charge stored in the electric double layer increases with increasing electrode surface area, activated carbons with high porosity are employed for the realization of EDLCs. However, overall performance of EDLCs is governed not only by the surface area of the activated carbon, but also by the pore structure, size, and distribution. The surface area of activated carbon is usually determined by the measurement of nitrogen gas adsorption isotherms and influenced mainly by micropores in the materials, which are classified as less than 2 nm in diameter. Conversely, the capacitance has been found to depend mainly on the surface area of carbon electrodes, even though high surface area carbons do not always give a large capacitance on carbons derived from the same precursor. This is attributed mainly to the fact that the micropores are not useful in forming the electric double layer of solvated ions, and do not offer sufficient accessibility to the electrolyte. Since the disadvantage of micropores is also notable under high-rate charge–discharge conditions, the incorporation of mesopores (2 ~ 50 nm in diameter) with smaller resistance for ion transfer is considered to be necessary. The control of pore-size distribution in electrode materials is also important for the energy and power density of EDLCs, since these characteristics are strongly dependent on the properties of the electrode material, and there is usually a trade-off relationship. The pore structure of carbons in electrode materials has a definitive and complicated influence on the performance of EDLCs and is still under discussion [2,3].

The specific surface area of carbon materials is essentially finite. This fact implies that the improvement of specific capacitance by means of the enrollment of specific surface area has a finite limitation. Accordingly, the introduction of redox (Faradic) reactions at the
electrode surface to store electric charge has been widely attempted to enhance the specific capacitance of carbon materials. Since the capacitance due to these redox reactions differs from the electric double-layer capacitance, this kind of charge storage is called pseudocapacitance. The redox reactions that contribute to pseudocapacitance are a variety of electrochemical reactions and generally result in a certain degree of deterioration of durability.

The energy density of a capacitor cell is determined by the product of electrode capacitance and the square of the working voltage, which is restricted by decomposition of electrolyte solutions or electrode materials. Electrolytes are aqueous and nonaqueous solutions of different types of acids, bases, or salts. The working voltage of aqueous solutions is usually restricted to under ~1 V, due to the decomposition of water. On the other hand, commonly used organic solutions have a tolerance of around 2.5 V. This difference in the working voltage window results in about a 6-fold larger energy density for nonaqueous electrolytes compared with aqueous electrolytes. Recently, EDLCs using organic electrolytes have attracted great attention from industrial and fundamental points of view, because of the potential application in high-energy density devices [4,5]. In fact, most recent capacitor cells with high capacity have been realized by adoption of an organic electrolyte. Nonaqueous EDLCs, however, have several barriers to widespread commercialization, including cost, and, in some cases, safety of the components. On the other hand, the aqueous electrolyte for EDLCs is mainly dilute sulfuric acid, which has a long history of use in lead—acid secondary batteries, and is superior from the viewpoints of low cost, safety, and physical properties. Aqueous electrolytes, such as dilute sulfuric acid solutions, have higher electrical conductivity than the commonly used organic electrolytes. This is an important parameter in achieving a high-rate charge—discharge performance. The operating temperature range of aqueous electrolyte solutions is also suitable for capacitor cells, when compared with the most organic solutions (except ionic liquids). Therefore, the development of new types of carbon electrodes, which can be used in aqueous electrolytes while attaining high energy density, is eagerly awaited.

Single-walled carbon nanotubes (SWCNTs) are one of the most anticipated materials for EDLC electrode materials because of their large surface area, monotonic pore-size distribution, and high electrical conductivity [6]. However, most SWCNT products are bundled, resulting in lower surface area than the ideal single carbon nanotube. In recent years, the so-called super growth method of SWCNT production (SG-SWCNT) was developed, producing high-purity and bundle-free SWCNTs at mass scale. The advantages of using SG-SWCNTs as EDLC electrodes have been reported by Tanaike et al. [7].

As mentioned above, the investigation of novel carbon materials is directly connected to the improvement of EDLC performance, since carbon is essential for capacitor electrode materials. The following section presents recent developments in carbon materials for EDLC electrodes, in which the enhancement of various characteristics was achieved by various techniques such as chemical and structural control, and hybridization with heteroatoms. Recently, capacitors using these novel carbon materials, having Faradic reactions and novel mechanisms to enhance the total performance, have become known as supercapacitors.

3. TEMPLATE METHOD FOR PORE STRUCTURE CONTROL OF ELECTRODE CARBONS

A template method for the synthesis of various carbons has been developed for control of morphology and pore structure of the resulting products. For capacitor electrodes, the carbon materials should have sufficiently high electrical conductivity and well-developed pore structure. For this purpose, various template materials, such as mesoporous silica [8–12], zeolite [13], colloidal silica [14], and organic nanocomposite [15], have been examined for the carbonization of different carbon precursors. Among these template materials, mesoporous silica has been studied the most extensively. Hyeon et al. first developed porous carbons from mesoporous silica templates and reported their application as electrode materials for supercapacitors in an early paper [16]. The synthesis process comprises the incorporation of carbon precursor from vapor or liquid/solution phase into the narrow spaces throughout the template structure, followed by carbonization at elevated temperature under inert atmosphere. The template structure is then eliminated by using strong chemicals, such as hydrofluoric acid. Mesoporous carbons produced from a mesoporous silica template have attracted attention from various international groups, due to the homogeneous and monodisperse mesoporous structure. The resulting materials appear to be suitable for use as capacitor electrode materials.

Although there are different types of mesoporous silica, two kinds of structure have been revealed to be most suitable as templates to produce porous carbons, namely, MCM-48 with three-dimensional continuous mesopores, and SBA-15 with a hexagonal arrangement of rod-like mesopores. The mesoporous carbons derived from these templates have a relatively high surface area from about 700 to 2000 m²/g, with highly developed mesopores. By choosing the template and controlling the synthesis conditions, the porous character of the product carbons can be changed, as shown in Table 1. The specific surface area and the pore-size distribution in the resulting carbons are strongly affected by the precursor compounds when
SBA-15 and MCM-48 are used. The largest specific area in the resulting carbons was achieved by a type of zeolite template, reaching to ~4000 m$^2$/g.

In most cases, the electric double-layer capacitance of carbon materials is closely related with the BET (Brunauer–Emmett–Teller theory) surface area. This is also true of the mesoporous carbons; Figure 1 shows the strong correlation of the gravimetric capacitance of porous carbons with the specific surface area. In other words, all the carbons have the areal specific capacitance in the range from 0.1 to 0.2 F/m$^2$. Since the areal specific capacitances of mesoporous carbons apparently coincide with those of usual activated carbons, the advantage of using mesoporous carbons from the template method is ambiguous from the viewpoint of volumetric capacitance, due to the higher porosity of mesoporous carbon.

Conversely, mesoporous carbons are considered to be suitable for high-rate charge–discharge conditions (high current density), because the developed mesopores with homogeneous and narrow pore-size distribution allow fast transfer of solvated ions in the electrolyte solution. Since the formation of the electric double layer is essentially accompanied by ion transfer from the bulk solution to the solid–solution interface, larger pores are favorable for this phenomena due to smaller ion-transfer resistance. If the carbon material possesses a large amount of meso- or macropores, the porosity increases and the volumetric capacitance decreases. Therefore, the balance of pore-size distribution is certainly an important factor in obtaining optimized capacitive behavior.

Microporous carbons formed using a zeolite template have a narrow distribution of micropore size, and extraordinarily large surface areas as high as 4000 m$^2$/g. However, significantly higher rate performance, even in organic electrolytes (in which the solvated ions have large diameter), has been reported. The periodicity of micropores with monotonic diameter is considered to be the reason for the small resistance of ion transfer at high current density polarization.

### 4. NITROGEN-ENRICHED CARBONS

Some layered silicate materials, such as clay and mica, are known to accommodate small nitrogen-containing organic solvent molecules in their chemical structures. These can be carbonized in nano-space, in between the adjacent silicate layers. The resulting carbons have a morphology reflecting the template of the layered silicate materials, resulting in very thin films that are transparent in electron micrographs, and that are circular with a diameter of a few tens of nanometers. Since the intercalation of solvent molecules inside the layered structure is mainly due to interaction of the nitrogen-containing molecules with cations between the silicate layers, organic molecules such as quinoline or pyridine are preferably employed. The carbons obtained from this process contain substitutional nitrogen heteroatoms originating from the organic precursor molecules, and are generally referred to as nitrogen-enriched carbons. Nitrogen-enriched carbons derived from expandable fluorinated mica have a specific surface area of about 80–100 m$^2$/g and a gravimetric capacitance of about 100–160 F/g in sulfuric acid electrolyte (Figure 2) [17]. This results in the areal specific capacitances ranging from 1.2 to 2.2 F/m$^2$, which are 10-fold larger than those of conventional

![Graph showing the relation between the specific surface area and the gravimetric capacitance of mesoporous carbons and activated carbons in a three-electrode cell with 1 mol/dm$^3$ H$_2$SO$_4$ electrolyte.](image-url)
activated carbons. Such large capacitance values per unit surface area are not explainable by simple charge accumulation in the electric double layer, and are considered to be due to the pseudocapacitance related to nitrogen functionality in the carbon skeleton.

Stimulated by such nitrogen-enriched carbons derived from mica templates, a variety of carbon materials synthesized from nitrogen-containing organic precursors have been extensively studied for EDLC electrodes [18–22]. One representative material is carbon synthesized from polymerized melamine (C₃H₆N₆), using a mica template. This material contains more than 20 wt% nitrogen after heat treatment at 800 °C [23]. Electrochemical testing in sulfuric acid electrolyte indicated that the melamine-derived nitrogen-enriched carbons have 150 ~ 200 F/g gravimetric capacitance without an apparent redox reaction, even though the materials have only 15 ~ 400 m²/g specific surface area [19]. The simple calculation of areal specific capacitance using the above results gives extraordinary values which are probably not realistic.

Nitrogen atoms in carbon materials have various types of chemical bonds, which can be elucidated by X-ray photoelectron spectroscopy (XPS). Figure 3 shows that nitrogen in melamine-derived carbons exists mainly in the pyridinic, quaternary, and oxidized states, and that the population of those chemical states changes with heat treatment temperature. By comparing materials heat-treated from 850 to 1000 °C, quaternary bonding (substitutional nitrogen bonded to three neighboring carbon atoms) increases with increasing temperature, and pyridinic bonding (nitrogen located at the outside edges of carbon rings and bonded to two neighboring carbon atoms) decreases with increasing temperature. The role of nitrogen in carbon materials for electrochemical capacitors is considered to be the fast faradic reaction between cations and lone pair electrons in the pyridinic state. However, the nitrogen-enriched carbons usually do not show any redox reaction in electrochemical measurements such as cyclic voltammograms (CV). The high capacitance of these materials has been observed not only in acidic media but also in basic solutions [24]. Although the mechanism behind large gravimetric capacitance in carbons with low surface area derived from melamine resin is thought to be some kind of electrochemical reaction, cyclability has been confirmed up to few million cycles while maintaining the initial capacitance value.

Other materials in this category are carbonized melamine foam, and mesoporous nitrogen-enriched carbons derived from mesoporous silica templates. The carbonization of melamine foam (used as commercially available cleaning sponges) results in the formation of a unique carbon foam which has resiliency and a similar shape to the original sponge [25]. The gravimetric capacitance of carbonized foam in sulfuric acid electrolyte, measured by galvanostatic charge–discharge cycling, reached 240 and 220 F/g at 0.2 A/g and 1.0 A/g discharge current density, respectively (Figure 4).

By using mesoporous silica templates and nitrogen-containing organic compounds, nitrogen-enriched mesoporous carbons have been produced to accomplish high capacitance in tandem with nitrogen-associated pseudocapacitance, and highly developed pore structure.
[26]. Figure 5 shows scanning electron microscope (SEM) and transmission electron microscope (TEM) images of nitrogen-enriched mesoporous carbons synthesized from quinoline-polymerized pitch with an SBA-15 template. In the micrographs, a series of straight, aligned mesopores are clearly observed. The wall of the mesopore structure is composed of nitrogen-enriched carbon containing ~6 wt% nitrogen, and possessing a significant amount of micropores, revealed by the adsorption isotherm measurement. The gravimetric capacitance of related materials is listed in Table 2. Clearly, nitrogen-containing mesoporous carbons show higher capacitances than nitrogen-free materials using the same template. The higher capacitance of nitrogen-enriched mesoporous carbons was also confirmed in organic electrolyte (1 mol/dm³ Et₄NBF₄ in propylene carbonate) in which specific capacitances of 100–140 F/g were obtained, exceeding that of activated carbons used in commercial capacitor devices (70 ~ 90 F/g). Similar enhancement by nitrogen enrichment in carbon materials was also recognized for microporous carbons.

**TABLE 2 Specific Surface Area and Gravimetric Capacitance of Nitrogen-Enriched Mesoporous Carbons**

<table>
<thead>
<tr>
<th>Template material</th>
<th>Carbon precursor</th>
<th>BET surface area (m²/g)</th>
<th>Gravimetric capacitance * (F/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBA-15</td>
<td>Nitrogen-containing quinoline pitch</td>
<td>615</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>Nitrogen-free pitch</td>
<td>595</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Nitrogen-free polymer</td>
<td>1233</td>
<td>132</td>
</tr>
<tr>
<td>MCM-48</td>
<td>Nitrogen-containing quinoline pitch (A)</td>
<td>1710</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td>Nitrogen-containing quinoline pitch (B)</td>
<td>1296</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>Nitrogen-free polymer</td>
<td>1171</td>
<td>131</td>
</tr>
</tbody>
</table>

*1 mol/dm³ H₂SO₄ electrolyte, current density 200 mA/g
derived from zeolite templates using an acetonitrile precursor [27].

5. EXFOLIATED CARBON FIBERS

In the case of activated carbon for EDLC electrodes, the enhancement of specific capacitance was pursued mainly by increasing the specific surface area and controlling the pore-size distribution. Modification of surface functionality, introduction of heteroatoms, and the addition of metal compounds have also been extensively studied. Most of these studies, however, were performed on carbons with low crystallinity because of the moderate preparation temperature. There were no experiments on carbons with high crystallinity, for example, graphite. This is because it is very difficult to increase the specific surface area of graphite, which is considered to be essential for capacitor electrodes. Recently, exfoliated carbon fibers (ExCFs) which possess both high crystallinity and large surface area were developed, and their characteristic performance as capacitor electrodes was determined [28, 29].

Carbon fibers (mesophase-pitch- or PAN-based) were used as starting materials. Their intercalation compounds were synthesized by electrolysis in 13 mol/dm$^3$ HNO$_3$. The intercalated carbon fibers thus obtained were rapidly inserted into a tubular furnace at 1000 °C for 5 s for exfoliation. The detailed exfoliation behavior of carbon fibers is reported in several papers [30–32].

X-Ray diffraction patterns of pristine, electrolyzed, and exfoliated carbon fibers are shown in Figure 6. Conventional activated carbon with low crystallinity is also shown for comparison. The pristine carbon fiber has high crystallinity, showing a sharp diffraction peak with a d spacing of 0.337 nm (Figure 6a). After electro-oxidation in HNO$_3$, the diffraction peaks of the graphite structure disappeared and a broad peak at a d-value of 0.78 nm appeared (Figure 6b). According to the chemical analysis of this material, the product of this treatment coincided with graphite oxide. Subsequently, graphite oxide was completely decomposed after heat treatment at 1000 °C (Figure 6c) and the carbon structure was recovered, although the carbon layer stacking along the c-axis is smaller than that in pristine carbon fibers, because of the exfoliation process.

SEM observations revealed that individual fibers were converted to bundles of thin filaments by exfoliation, reflecting the fibril texture along the fiber axis. The nitrogen adsorption isotherm for ExCFs at 77 K showed a well-developed mesoporous character, with about 300 m$^2$/g BET surface area. Mercury porosimetry indicated that ExCFs have total pore volume of about 20 cm$^3$/g, which can be reasonably attributed to meso- and macropore regions among the thin filaments.

Extremely large capacitance values reaching 555 and 450 F/g were obtained by CV and charge–discharge measurements, respectively, when 18 mol/dm$^3$ H$_2$SO$_4$ electrolyte was used. The CV curve slightly deviated from the ideal rectangular shape, but no redox peak was observed. The galvanostatic experiment at low current density gave a triangular charge–discharge curve with linear change in potential, reflecting the CV curve, without peaks. Galvanostatic cycling with high current density (500 mA/g) in 18 mol/dm$^3$ H$_2$SO$_4$ was measured to be stable for more than 7000 charge–discharge cycles, although the specific capacitance was almost half that observed in the former experiment because of the high current density. When 1 mol/dm$^3$ electrolyte was used, the capacitance was 160 and 117 F/g measured by CV and the charge–discharge curve, respectively.

Figure 7 shows the dependence of areal specific capacitance for ExCFs and activated carbon fibers (ACFs), obtained from the 50th discharge curve of galvanostatic cycling, plotted against sulfuric acid electrolyte concentration. The capacitance value for ExCFs increased gradually with increasing concentration up to 10 mol/dm$^3$, then suddenly jumped to an extremely large value in the higher concentration region, reaching about 1.4 F/m$^2$ (450 F/g). However, the capacitance for ACFs does not show any trend with changing electrolyte concentration, with an approximately constant, small value of about 0.1 F/m$^2$. Since the capacitance of ExCFs in concentrated sulfuric acid is considerably large compared with the reported value of conventional carbon materials, the value for ExCFs obtained here cannot be explained by only the electric double layer. The pseudocapacitance also has to be taken into account.

![Figure 6](image_url)
Exfoliated natural graphite (ExNG) with the highest crystallinity showed a similar trend with electrolyte concentration, but abnormally large capacitances were obtained in 12 and 18 mol/dm$^3$ solutions, and a distinct intercalation reaction of H$_2$SO$_4$ molecules into graphite interlayer was confirmed.

Considering the above observation, the enhancement of the specific capacitance of ExCFs with concentrated electrolyte is not associated with the formation of stage structure caused by the intercalation reaction, although there may be some strong interaction between the electrode and the H$_2$SO$_4$ molecules. The large capacitance of ExCF in concentrated electrolyte implies that the charge transfer reaction (Faradic reaction) occurs at the electrode surface, since the capacitance values were unreasonably large and the intercalation of H$_2$SO$_4$ is obvious. The lack of stage structure formation in ExCF may due to lower crystallinity and a more distorted structure compared to graphite. The behavior of ExCF is better than ExNG, since the formation of stage structures introduces a volume change in the electrode, resulting in electrode collapse at short cycling times.

6. CARBON-COATED TRANSITION-METAL CARBIDES

Several types of transition-metal oxides are known to work as capacitor electrode materials, since the metal undergoes a change in oxidation state and then a hydrated state over a wide range working window. The apparent CV curves from these compounds show a pseudo-rectangular box shape resulting in the origin of the term pseudocapacitance. The most well-known example is ruthenium oxide, which has the highest capacitance in acidic media. Ruthenium oxide, however, is not considered suitable for use in commercial device applications because of its extremely high cost.

Inagaki et al. developed a simple process to coat carbon onto various metallic particles and compounds, and studied their functionality in different applications [33]. A similar procedure was applied for tungsten and molybdenum compounds to exploit their large change in oxidation state, as a replacement for ruthenium [34,35]. It was found that the thus-obtained carbon-coated tungsten and molybdenum carbides were easily converted to their corresponding oxides in sulfuric acid electrolyte, and that they showed stable capacitive behavior.

The starting materials used were hydroxyl propyl cellulose (HPC) as the carbon precursor, and reagent-grade K$_2$WO$_4$ and K$_2$MoO$_4$ powders as the precursors for tungsten and molybdenum carbide, respectively. The HPC was dissolved in water at a concentration of 10 mol/dm$^3$, and the carbide precursor (either K$_2$WO$_4$ or K$_2$MoO$_4$) was dissolved into this HPC solution in a mass ratio of 1/1. The mixed precursor solutions thus prepared became gels at room temperature, were dried at 100 °C, and then heat-treated at a temperature between 800 and 1050 °C in Ar gas flow. After heat treatment, the obtained powders were washed by 1 mol/dm$^3$ H$_2$SO$_4$ and then distilled water, in order to remove any potassium metal which might remain.

Figure 8 shows the change in X-ray diffraction pattern with heat treatment temperature for mixtures of HPC and K$_2$WO$_4$. Above 850 °C, two types of tungsten carbide were observed (W$_2$C and WC), as well as metallic W. With increasing temperature, the relative intensities of the
diffraction peaks for WC increased, suggesting that the HPC carbon precursor decomposes to carbonaceous materials, which coat K₂WO₄ particles. K₂WO₄ particles covered by the carbonaceous layer probably then decompose to amorphous tungsten oxide below 850 °C. This amorphous tungsten oxide reacts with the coated carbonaceous materials to form W₂C, and finally WC. From TEM observation, opaque particles of 30−50 nm in size with irregular shape were observed, and reasonably assumed to be WC. They appear to be coated by a carbon layer, assumed to be porous. In the case of K₂MoO₄, the formation of Mo₂C was confirmed above 800 °C.

Figure 9 shows the CV curve of carbon-coated tungsten components—a mixture of WC, W₂C, and W in 1 mol/dm³ sulfuric acid electrolyte. No apparent redox peaks were observed, and the gravimetric capacitance was calculated to be 104 F/g from this voltammogram. It was revealed by chemical analysis that only 13 wt% of carbon was included in the composite material, indicating the important contribution of tungsten components on the capacitive behavior. The X-ray powder pattern of the composite materials after one voltammetry cycle showed diffraction lines from tungsten oxyhydrate (H₀.₁₀WO₃ $\cdot$ 1.₀₆H₂O), and all tungsten components found in the raw composite had disappeared. This implies that the first polarization of tungsten carbide in sulfuric acid causes an electrochemical reaction and the formation of oxyhydrate from the parent carbide. It is known that tungsten oxyhydrate undergoes a change in oxidation state associated with the hydration composition along with the polarization potential. Therefore, these changes might be responsible for the pseudocapacitive behavior in sulfuric acid electrolytes. Since the tungsten compounds were coated with carbon and remained inside the carbon layer, the electrode fabricated from this composite material showed stable charge−discharge cycling for several hundred cycles. Similar behavior was also found for carbon-coated molybdenum carbide particles.

7. CARBON XEROGELS WITH CONDUCTIVE POLYMER NANO-COATING

Poly(p-fluorophenylthiophene) (PFPT), a conductive polymer, is one of the most promising candidates for electrode materials because it has a large charge storage capacity for both p- and n-doping, and a large potential difference between these two processes [36]. However, the rate performance of polymer-based capacitors is inferior to that of EDLC, because the rate of diffusion of ions incorporated in the polymer is much slower than that of accumulation of ions in an electric double layer. In order to improve the rate performance, it is necessary to shorten the diffusion path of ions by forming the polymers into thin layers, on porous substrates with high electrical conductivity.

Carbon xerogels (CXGs) are derived via sol−gel polycondensation, subcritical drying, and high-temperature heat treatment. They have properties making them suitable for use as electrode materials, including high electrical conductivity, porosity, and corrosion resistivity. These properties are also desirable in substrates for PFPT coating by polymerization. In order to study PFPT/CXG composite electrodes, the relationship between electrochemical properties and layer thickness was first investigated on a PFPT layer formed on a nonporous carbon film, with the aim of determining the optimal layer thickness. Then, carbon xerogels were coated with a PFPT layer having appropriate thickness, and the electrochemical properties of the resulting electrodes were characterized.

The monomer p-fluorophenyl thiophene (FPT) was synthesized by a nickel(II)-catalyzed coupling reaction, according to the literature (Figure 10) [37]. By using electrochemical polymerization of dissolved FPT in a propylene carbonate (PC) solution of 1 mol/dm³ tetraethylammonium tetrafluoroborate (Et₄NBF₄), a PFPT layer with uniform thickness was formed on the carbon film. The thickness of the PFPT layer increased in proportion to the

![FIGURE 9](image9.png) Cyclic voltammogram of the carbon-coated tungsten carbide fine particles in a three-electrode cell with 1 mol/dm³ H₂SO₄ electrolyte at 1 mV/s of scan rate.

![FIGURE 10](image10.png) Synthesis scheme of p-fluorophenyl thiophene (p-FPT) monomer.
polymerization charge per unit surface area \( Q_{\text{poly}} \), and reached about 8 \( \mu \text{m} \) at \( Q_{\text{poly}} = 0.28 \text{ mAh/cm}^2 \) (Figure 11).

In the cyclic voltammogram for the PFPT layer, two peak pairs, assigned to p- and n-doping and de-doping, appear in the potential ranges of 0.2 to 0.8 V and -2.1 to -1.6 V, respectively. The capacity of p- and n- de-doping at various sweep rates is shown in Figure 12, as a function of \( Q_{\text{poly}} \). In both p- and n-doping, the maximum capacitance exceeds 210 F/g. Above \( Q_{\text{poly}} = 0.01 \text{ mAh/cm}^2 \), the capacity decreases with increasing \( Q_{\text{poly}} \), and the decrease is made larger with increasing sweep rate. This is because ionic species cannot be inserted inside the PFPT layer, due to the insufficiently slow diffusion of ionic species in the polymer matrix. Conversely, below \( Q_{\text{poly}} = 0.01 \text{ mAh/cm}^2 \), the thickness of the PFPT layer is so small that the ionic species can be inserted across the whole layer even under fast potential sweeping. Thus, the capacity is determined by the structure of the PFPT layer. The change of capacity with \( Q_{\text{poly}} \) suggests that there is a structural difference between PFPT layers with different thickness.

In order to investigate the change in reaction mechanism during polymerization, the stoichiometric number \( N \) was calculated from the neutralization charge, the polymerization charge, and the molecular weight of FPT. As shown in Figure 13, the value of \( N \) is considerably larger than the ideal case of 2 [38] at \( Q_{\text{poly}} < 0.01 \text{ mAh/cm}^2 \). This indicates that some preferential side reaction takes place. It is therefore considered that the small capacity at \( Q_{\text{poly}} < 0.01 \text{ mAh/cm}^2 \) results from the existence of structural defects introduced by side reactions. From the results shown above, it is concluded that a layer thickness of about 200 nm, achieved by polymerization with \( Q_{\text{poly}} = 0.01 \text{ mAh/cm}^2 \), is optimal from the viewpoint of maximizing capacity.

The pores in the CXG substrate have to be large enough so as not to destroy contact with the external surface even after the PFPT coating is applied. At the same time, the walls of CXG should have the minimum volume in order to maximize the PFPT content in the coated composite material, while maintaining high electrical conductivity. By taking into account the estimated densities of the CXG skeleton and the PFPT (1.80 and 0.94 g/cm\(^3\), respectively) the thickness of the CXG walls must be below 100 nm to give 90 wt% PFPT in the coated composite.

CXGs were prepared from aqueous solutions of resorcinol, formaldehyde, and sodium carbonate, using...
a procedure described in the literature [39]. The solution was gelated by a two-step heat treatment at 20 °C for 7 days then 50 °C for 4 days. In order to minimize shrinkage during drying, water in the wet gel was first exchanged with acetone, and then acetone was removed from the gel by evaporation. The organic xerogels were carbonized by heating at up to 1500 °C under nitrogen gas flow. By optimizing the preparation conditions (such as gelation period), the above-mentioned structural parameters required of the pore diameter and wall thickness were achieved. Figure 14a indicates that the thus-obtained CXG has a uniform distribution of primary carbon spheres, which are connected in a three-dimensional network. The size of intra-particle mesopores increased with gelation period, exceeding the target size of 400 nm after 7 days (Table 3).

To form PFPT on CXG, the PC solution saturated with FPT was first impregnated into the CXG electrode under reduced pressure. Then, the electrode was placed in a 0.5 mol/dm$^3$ Et$_4$NBF$_4$ PC solution, and a constant potential of 1.05 V was applied. The current decreased as polymerization progressed, due to the reduction of the amount of residual monomer. After the current reached 1 μA, the PFPT on the CXG was neutralized by sweeping the potential to 0 V. Figure 14b shows that the PFPT uniformly coated the primary carbon spheres in CXG, with

![FIGURE 13](image13.png)

**FIGURE 13** Change in stoichiometric number of electron during the electrolytic polymerization of p-FPT with polymerization charge density.

![FIGURE 14](image14.png)

**FIGURE 14** SEM images of carbon xerogel electrodes (a) before and (b) after applying PFPT coating.

<table>
<thead>
<tr>
<th>TABLE 3 Preparation Conditions and Characteristics of Pore Structure in Carbon Xerogels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>

$t_{20}$: Gelation period at 20 °C, $\rho_{bulk}$: Bulk density, $\rho_{app}$: Apparent density, $V_{total}$: Total pore volume, $S_{ext}$: External surface area, $D_{ave}$: Average pore diameter
a thickness of about 100 nm. The cyclic voltammograms of PFPT-coated CXG gave the same electrochemical response for p- and n-doping and de-doping as PFPT on a carbon film, over a wide range of scan rates (Figure 15).

8. SUMMARY

Carbon materials with various structural characteristics have been developed and employed as capacitor electrode materials. The strategy for the enhancement of capacitance is also based on a broad array of mechanisms. The variation in structural diversity of carbon materials and combination with heteroatoms and other materials will enable them to be applied successfully in supercapacitors for future applications.

REFERENCES


Chapter 3.1

Silicon Carbide and Other Carbides: From Stars to the Advanced Ceramics

Branko Matović* and Toyohiko Yano**
*Vinca Institute of Nuclear Sciences, University of Belgrade, 11001 Belgrade, Serbia, Former Affiliate: Research Laboratory of Nuclear Reactor, Tokyo Institute of Technology, **Research Laboratory for Nuclear Reactors, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

Chapter Outline
1. Introduction 225
2. Carbides in Nature 226
3. Transition Metal Carbides 228
4. Covalent Carbides 233
5. Synthesis 235
6. Extreme Environment Application 240
7. Importance of Natural and Synthetic Carbides 241
References 241

1. INTRODUCTION

Carbides are a class of ceramic materials that have drawn great interest to the scientific community due to their exceptional profile of mechanical, physical, chemical, and even biological properties. Therefore, carbides are the basis of advanced ceramics with numerous applications and with a future prospect. Although they have been known for over one hundred years, and being important materials in many technological fields, such as tool industry, mechanical engineering, engine industry, optic, electronic, catalytic, nuclear technology, and chemical industry, most of their applications are recent. Although an extensive compilation of the carbides exists in textbooks, almost no one treats the natural carbides. The poorly known data are discussed and pointed. Therefore, in this chapter, carbides are described and discussed in the context of their natural occurrences as well as their same synthetic phases. Since native carbides are refractory compounds, special attention is given to the carbides that possess high-melting point, particularly on their synthesis, structure, and some high-temperature properties.

From the chemistry point of view, carbides are compounds in which carbon is combined with less electronegative elements of a metal or a semimetal. Carbon is the chemical element (symbol C) with atomic number 6. It belongs to group 14 on the periodic table, and, thus, it is nonmetallic. Carbon is tetravalent, enabling 4 electrons available to form covalent chemical bonds. Since carbon is the one of the most abundant chemical elements in the universe by mass (after hydrogen, helium, and oxygen) as well as in the Earth’s crust, it forms more compounds than any other element. Among them, carbides are of great significance as engineering materials with numerous applications. Because of numerous carbon compounds, sometimes it becomes extremely cumbersome to define to which group the corresponding compound belongs. By convention, the term carbide is only used for compounds formed by carbon with other elements of lower or nearly equal electronegativity.

They may be classified on the basis of chemical composition, chemical bond, method of manufacture, physical form, or according to their applications. However, the interrelated atomic features play an essential role in the formation of carbides, i.e., the difference of the electronegativities (ΔEN) of carbon and the metal or the semimetal, the atomic size of the respective elements, and their corresponding bonding type. According these criteria, several classes of carbides are usually distinguished: (i) salt like, with high ΔEN and ionic behavior, (ii) covalent compounds, with small, almost vanishing ΔEN and strong covalent bonding, (iii) interstitial compounds, and (iv) “intermediate” transition metal carbides with intermediate ΔEN and metallic properties.

1.1. Salt-like Carbides (SCs)

Salt-like carbides are formed by the elements of groups I, II, and III. Because of the highly electropositive elements of alkali, alkaline-earth metals, and group III metals,
carbon can be assumed to be anion and thus the corresponding bonding is anionic. According to the type of the carbon anion, salt-like carbides can be subdivided into isolated C anions, and C₂ and C₃ dumbbells, namely methanides with C⁴⁻ anions (Be₂C and Al₄C₃), acetylides with C₂⁻ anions (Na₂C₂ and CaC₂), and allylenides with C₃⁻ anions (Li₄C₃ and Mg₂C₃).

1.1.1. Methanides
Methanides group or methides get the name because of methane gas created during their reaction with water. The most important methanides are beryllium carbide (Be₂C) and aluminum carbide (Al₄C₃). However, only Be₂C meets the criteria as high-melting-point carbide:

\[
\text{C}^4^- + 4\text{H}^+ \rightarrow \text{CH}_4 \quad (1)
\]

\[
\text{Al}_4\text{C}_3 + 12\text{H}_2\text{O} \rightarrow 3\text{CH}_4(g) + 4\text{Al(OH)}_3 \quad (2)
\]

1.1.2. Acetylides
Acetylides are carbides which can be described as salt-like derivates of C₂⁻. Anion C₂⁻ has a triple bond between two carbon atoms. Alkali metals and alkaline-earth metals form acetylides as well as lanthanoids with formula M₂C₃. Actinide compounds, with stoichiometry MC₂ and M₂C₃, also belong to this subgroup. They react with water, giving acetylene and the corresponding hydroxide:

\[
\text{CaC}_2 + 2\text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2(g) + \text{Ca(OH)}_2 \quad (3)
\]

1.1.3. Sesquicarbides
Sesquicarbides possess anion C₃⁻, which yields methylacetylene (propyne), CH₃CCH, and hydroxide on hydrolysis:

\[
\text{Mg}_2\text{C}_3 + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{C} = \text{CH} + \text{Mg(OH)}_2 \quad (4)
\]

1.2. Covalent Carbides
There are only two elements (Si and B) that are most similar to carbon in size and electronegativity and therefore form carbides with a high degree of covalent bonding (SiC and B₄C). The fundamental structural unit in SiC is a covalently bonded primary co-ordination tetrahedron (either SiC₄ or CSi₄) and in B₄C is boron icosahedrons (B₁₂) bridged by carbon atoms.

1.3. Interstitial Carbides (IC)
Interstitial carbides are formed from relatively large transition metals (metal atom radius is greater than approximately 0.135 nm) of the groups IV, V, and VI and 4th, 5th, and 6th periods, which act as a host lattice for the small carbon atom that occupies the interstices of the close-packed metal atoms. They are often described as interstitial carbides. According to the mode of metal atom arrangement, interstitial carbides possess 1:1 or 2:1 stoichiometry. When the metal atoms are cubic close packed (ccp), all octahedral interstices are filled by carbon, resulting in 1:1 stoichiometry, such as rock salt structure. On the other hand, when the metal atoms are hexagonal close packed (hcp), the octahedral interstices lie directly opposite each other side of the layer of metal atoms and only one is filled by carbon achieving 2:1 stoichiometry.

1.4. Intermediate Transition (IM) Carbides
Intermediate carbides are formed from transition elements of groups VII and VIII. The size of metal ion of the intermediate transition carbide is smaller than the critical 0.135 nm and accommodation of carbon atom in interstitial positions is impossible without distortion of the crystal lattice. They are characterized by multiple stoichiometries and from the view of structure they are not interstitial but are more complex.

2. CARBIDES IN NATURE
2.1. Carbides and Natural Occurrence
It is thought that the total amount of carbon in the atmosphere, oceans, and other near-surface reservoirs is negligible compared to that stored in the Earth’s mantle (highly viscous layer between the crust and outer core of the Earth) [1,2]. However, the mode of carbon storage in the mantle is unknown. The observation of micro-bubbles on dislocations in minerals from mantle xenoliths (an individual foreign crystal included within an igneous body) indicates that carbon could be soluble in silicates at high pressure [3,4]. Carbon solubility in silicates is crucial for understanding the dynamics of the carbon exchange between the mantle and near-surface reservoirs. Even the solubility of the order of a few hundred parts per million (ppm) by weight would be sufficient to incorporate all carbon in the upper mantle [5]. Since carbon is abundant in the Earth and, therefore, carbides’ occurrence in nature is not surprising.

The first natural occurrence of carbides was mineral moissanite (SiC), which was found in 1893 as a small component of the Canyon Diablo meteorite in Arizona by Dr. Ferdinand Henri Moisson [6], after whom the material was named in 1905.

Until the 1950’s, no other source, apart from meteorites, had been encountered. Later, moissanite was found as
inclusion in kimberlite rock and lamproites that are formed deep within the mantle from a diamond mine in Yakutia [7–10], Shandong Province, China [11], Green River Formation in Wyoming [12], West Kimberly Province, western Australia [13], Monastery, South Africa [14], Sloan, USA kimberlites [15], and Argyle, Australia [16].

Discoveries have shown that moissanite occurs naturally as inclusions in diamonds, xenoliths, and ultramafic rocks such as kimberlite and lamproite [17,18]. In addition, the other SiC polymorphs α- and β-SiC are found in kimberlite from Fuxian, China [19]. The discovery of a cluster of SiC entrapped in diamond indicates diamond–SiC paragenesis. As SiC can be synthesized only at high temperatures, and because of its association with diamond-bearing kimberlite, a source of region of Earth’s mantle is inferred. Thus, SiC is probably generated in a deep Earth crust and its exhumation in many rocks comes later with geological activities [20].

SiC occurs naturally and that is a widespread, albeit rare, phase in diamond-bearing rocks.

Recently, several natural carbides were found in Loubusha ophiolite in southern Tibet [21]. Most of them are carbides of transition elements (Table 1).

From metallurgy it has been shown that W–Co–C phases could dissolve a substantial amount of metals such as V, Cr, and Ta, which are known to positively influence the microstructure of hard metals [22]. This is the nature approach. The minerals listed in Table 2 show the high degree of their solubility. Thus, carbide minerals are more solid solutions than individual carbides.

It is considered that transition carbides are xenocrysts from the mantle, transported to shallow depths by a rising plume and then captured by the melts from which the chromite magma is crystallized [23]. The other evidence of the carbide minerals (Table 2) is discussed in the following.

### 2.2. Stellar Sources of Carbon and Carbides

While rare on the Earth, carbides are common in space, especially silicon carbide. Carbides have been found as stardust around carbon-rich stars [37]. Their evidence comes also from pristine (unaltered) meteorites. They are isotopically different from the fine-grained matrix of primitive meteorites, such as hondrites, whose differences from the surrounding meteorite suggest that they are older than the solar system and usually called presolar grains [38].

Crystallinity in these presolar compounds ranges from that of micrometer-sized silicon carbide (SiC) crystals down to that of diamond and unlayered graphene crystals with fewer than 100 atoms. These grains were probably

<table>
<thead>
<tr>
<th>Mineral</th>
<th>Cohenite (Fe,Ni,Co)₃C</th>
<th>Isovite (Cr,Fe)₂₃C₆</th>
<th>Haxonite (Fe,Ni)₂₃C₆</th>
<th>Tonghaite Cr₃C₂</th>
<th>Yarlongite (Cr,Fe,Ni)₉C₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Composition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6.56</td>
<td>5.77</td>
<td>5.70</td>
<td>13.34</td>
<td>9.22</td>
</tr>
<tr>
<td>Fe</td>
<td>54.92</td>
<td>26.82</td>
<td>89.50</td>
<td>–</td>
<td>40.60</td>
</tr>
<tr>
<td>Ni</td>
<td>26.86</td>
<td>–</td>
<td>4.91</td>
<td>–</td>
<td>8.54</td>
</tr>
<tr>
<td>Cr</td>
<td>–</td>
<td>67.41</td>
<td>–</td>
<td>86.66</td>
<td>–</td>
</tr>
<tr>
<td>Co</td>
<td>9.66</td>
<td>–</td>
<td>0.18</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Σ</td>
<td>100</td>
<td>100</td>
<td>100.29</td>
<td>100</td>
<td>99.74</td>
</tr>
</tbody>
</table>

**TABLE 1** Some of Natural Transition Element (Fe, Cr, Co, Ni) Carbides [21]
formed in supernovae or the stellar outflows of red giant stars, and then incorporated into the molecular cloud from which the solar nebula separated to form our solar system. Presolar grains identified so far consist of refractory minerals, which survived the collapse of the solar nebula, and the subsequent formation of planetesimals (solid objects form out of dust grains that collide and stick to form larger bodies).

There are different types of presolar grains. The most abundant is silicon carbide, which usually appears in submicrometer- to micrometer-sized grains. Presolar SiC occurs as a single-polytype grain or polytype intergrowths. The atomic structures observed contain the two lowest order polytypes: hexagonal 2H and cubic 3C with varying degrees of stacking fault disorder as well as 1-dimensionally disordered SiC grains. In comparison, terrestrial laboratory-synthesized SiC is known to form over two hundred different polytypes. This limited polytype distribution in presolar SiC can be explained by condensation of SiC at a relatively low total pressure in circumstellar shells [39].

Presolar grains consisting of the following carbide minerals have so far been identified: SiC, TiC, RuC, Fe3C, ZrC, and MoC [40–42]. Crystals of Ti-, Zr-, and Mo-bearing carbides are found within the spheres which served as nucleation centers for the graphite growth. These carbides have remained cloaked in a protective covering of graphite since their formation and therefore provide direct information on the chemical and physical properties of carbon-bearing circumstellar condensate [40].

Energy dispersive X-ray spectroscopy (EDS) analysis of the crystals shows that they are generally carbides of the refractory elements Ti, Zr, and Mo, although some crystals rich in Ni, Fe, and Ru were also detected. The most abundant element is Ti, which appears mostly as pure Ti carbide. Although the variability in element ratio points to a significant degree of solid solution carbides, which is in good agreement with numerous papers from the materials science studies. The atomic ratios relative to the dominant element Ti are enhanced by a factor of 10 to several thousands compared with material from solar origin, which clearly depicted their presolar origins as stellar condensates. Enrichments of Zr and Mo by a factor of 10–100 are predicted to occur in the envelopes of an AGB (asymptotic giant branch stars appear as a red giant star of low or intermediate mass in a late phase of stellar evolution with a central and inert core of carbon and oxygen and a shell where helium is undergoing fusion to form carbon—helium burning, another shell where hydrogen is undergoing fusion forming helium—hydrogen burning) star through s-process nucleosynthesis in their He-burning shells. Such enrichments are astronomically observed in carbon stars that have s-process enrichments of Zr by a factor of 10–100 relative to Ti [43].

Equilibrium thermodynamics give us a possibility to infer the relative condensation sequence of the various carbides observed in the meteorites. The types of minerals condensing from a gas with solar abundances of all the elements depend critically on the C/O ratio. In the case of C/O < 1, the formed phases are oxides and silicates, and opposite, when C/O > 1, most of oxygen is tied up in CO, allowing graphite and carbides to condense. The studies that explored the parameter space of pressure, temperature, and C/O ratio show the order of condensation: ZrC condensed first, followed by MoC or Mo2C and TiC [44]. This is consistent with the relative condensation temperatures of these carbides, i.e., ZrC, MoC, and TiC.

Abundant carbide—magnetite assemblages occur in the matrix of chondrite-type meteorites. Carbides, cohenite (Fe,Ni)23C6, and haxonite (Fe, Ni)32C6 depict compositional variations between different meteorites and appreciable ranges within meteorites. Metal associated with carbides and magnetite consists of a high amount of Ni and Co. Textural observations indicate that carbide—magnetite assemblages formed by replacement of metal-sulfide nodules. The high Co contents of residual kamacite (natural alloys of iron and nickel) in association with carbides indicate that Co is not incorporated into carbides (i.e., Fe/Co is much higher in the carbides than in kamacite). In Fe carbides rich with Ni, Krut et al. [45] suggested that carbide—magnetite assemblages are formed by hydro-thermal alteration of metallic Fe in metal-triolite (iron sulfide, FeS) nodules by a C–O–H-bearing fluid on their parent bodies. This alteration resulted in the carbodization of Fe–Ni metal by CO gas:

\[15\text{Fe}(s) + 4\text{CO(g)} = 4\text{Fe}_3\text{C(s)} + \text{Fe}_3\text{O}_4(s)\]  

\[3\text{Fe}(s) + 2\text{CO(g)} = \text{Fe}_3\text{C(s)} + \text{CO}_2(g)\]  

while the magnetite is formed by the oxidation by H2O gas:

\[3\text{Fe}(s) + 4\text{H}_2\text{O(g)} = \text{Fe}_3\text{O}_4(s) + 4\text{H}_2\]  

The C–O–H bearing fluids are derived from ice, adsorbed gases, or hydrated minerals, while CO is the result of the reaction of carbon compounds (e.g., hydrocarbons) with water vapor or magnetite.

Iron carbide (cohenite) has been reported as inclusions in diamond, either isolated [34] or intergrown with graphite [46].

3. TRANSITION METAL CARBIDES

3.1. Parameters Influencing the Carbide Structure and Properties

The important parameters that can explain the variety and similarity of carbides are given in Table 3. Although the Hf carbide is not found in nature, it is considered that Hf is
always present in Zr compounds. In addition, it has a potential for structural components that are able to serve in extreme high-temperature conditions.

The electronegativity considerations suggest simple metal(Me) → C electron donation. Electropositive alkali and alkaline-earth metals form ionic-type compounds with negatively charged nonmetals. These carbides are characterized by lack of electrical conductivity, transparency, and hydrolysis. On the other hand, these properties are not found in the alloys of carbon atoms with transition metals. The interstitial carbides are excellent electrical conductors and hydrolytically stable. Thus, simple electronegativity charge transfer concepts do not work when carbon atoms are interstitially dissolved in metal lattices. The interstitial carbides are excellent electrical conductors and hydrolytically stable. Thus, simple electronegativity charge transfer concepts do not work when carbon atoms are interstitially dissolved in metal lattices. It is pointed out that the amount and direction of charge transfer depend on the property considered [47]. When considering band occupation, electron transfer is from C → Me, but this reverses when considering integrated charges around atoms. However, some trends exist. For example, the difference in electronegativity between carbon and the elements forming carbide is more pronounced for the elements from the groups IV, V, and VI of the periodic table. In general, a large difference in electronegativity leads to interstitial carbides (ICs), while a smaller difference favors covalent carbides (CCs) or intermediate carbides (IMs).

In interstitial carbides, carbon atoms take the interstitial space within the framework of the metal atoms and the spacing between carbon atoms is too large for atomic interaction; there is no carbon-to-carbon bond resulting from overlapping electron shells. The overall bonding energy is the sum of metal-to-metal and metal-to-carbon bonds. Pierson [48] reviewed the atomic bonding of interstitial carbides and he gave some of the major differences between the host metal and the carbide:

1. Carbides are hard and brittle, while the host metals are malleable and much softer.
2. Carbides have a high bond strength, which exceeds by far the strength characteristics of the host metal.
3. The melting point of carbides is generally much higher than that of the parent metal.
4. A switch to a more stable structure occurs in every case when going from the metal to the carbide.

The best qualitative view of the bond’s energy can be obtained from the comparison of the melting points of carbides. The higher melting point of carbide means the higher Me–C bond strength.

The carbides from group IV have much higher melting points than their host metals. The difference is less pronounced in group V, while the metals have higher

<table>
<thead>
<tr>
<th>Element</th>
<th>Atomic radius (nm)</th>
<th>Carbon/metal atomic ratio</th>
<th>Electronegativity</th>
<th>Type of carbide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>0.078</td>
<td>—</td>
<td>2.5</td>
<td>—</td>
</tr>
<tr>
<td>Silicon</td>
<td>0.117</td>
<td>0.666</td>
<td>1.8</td>
<td>CC</td>
</tr>
<tr>
<td>Titanium</td>
<td>0.1467</td>
<td>0.526</td>
<td>1.5</td>
<td>IC</td>
</tr>
<tr>
<td>Zirconium</td>
<td>0.1597</td>
<td>0.483</td>
<td>1.4</td>
<td>IC</td>
</tr>
<tr>
<td>Hafnium</td>
<td>0.1585</td>
<td>0.486</td>
<td>1.3</td>
<td>IC</td>
</tr>
<tr>
<td>Vanadium</td>
<td>0.1338</td>
<td>0.576</td>
<td>1.6</td>
<td>IC</td>
</tr>
<tr>
<td>Niobium</td>
<td>0.1456</td>
<td>0.530</td>
<td>1.6</td>
<td>IC</td>
</tr>
<tr>
<td>Tantalum</td>
<td>0.1457</td>
<td>0.529</td>
<td>1.5</td>
<td>IC</td>
</tr>
<tr>
<td>Chromium</td>
<td>0.1267</td>
<td>0.609</td>
<td>1.6</td>
<td>IM</td>
</tr>
<tr>
<td>Molybdenum</td>
<td>0.1386</td>
<td>0.556</td>
<td>1.8</td>
<td>IC</td>
</tr>
<tr>
<td>Tungsten</td>
<td>0.1394</td>
<td>0.553</td>
<td>1.7</td>
<td>IC</td>
</tr>
<tr>
<td>Manganese</td>
<td>0.1261</td>
<td>0.618</td>
<td>1.5</td>
<td>IM</td>
</tr>
<tr>
<td>Iron</td>
<td>0.1260</td>
<td>0.619</td>
<td>1.8</td>
<td>IM</td>
</tr>
<tr>
<td>Cobalt</td>
<td>0.1252</td>
<td>0.623</td>
<td>1.8</td>
<td>IM</td>
</tr>
<tr>
<td>Nickel</td>
<td>0.1244</td>
<td>0.627</td>
<td>1.8</td>
<td>IM</td>
</tr>
</tbody>
</table>

melting points than the corresponding carbides in group VI. A crude interpretation of this trend is that the maxima are associated with the formation of a half-filled d-shell and, to achieve this, carbon donates one electron to the metal. According to the melting points of interstitial carbides and their host metals [49], Storms [50] has established the following order (Figure 1):

![FIGURE 1 Melting points of interstitial carbides and their host metals of the groups IV–VI.](image)

3.2. Crystallochemistry

The crystallochemistry of the interstitial carbides is similar to that of pure metals. They often possess a simple crystal structure, with the transition metal atoms forming lattices of face-centered cubic (fcc, or, the other term, cubic close-packed, ccp), hexagonal close-packed (hcp), simple hexagonal (hex), or body-centered cubic (bcc) structures. Their crystal structure can also be explained by the Engel–Brewer valence bond theory [56]. According to this theory, the crystal structure is determined by the number of outer-shell sp electrons per atom (\(e/a\)). These are itinerant and control the range order. For up to 1.5 \(e/a\), the bcc structure is obtained. A value of 1.7–2.1 \(e/a\) produces the hcp structure, and for 2.5–3 \(e/a\) attains the ccp structure. The carbon atoms enter into the interstitial sites between metal atoms showing the origin of the name for this class of materials. The available sites can be octahedral in fcc and hcp, or trigonal prismatic in hex lattices. The carbon atoms in these alloys occupy the octahedral interstitial sites. Thus, the crystal structure is determined by the geometric factor, which is based on an empirical rule given by Hågg. The rule states that simple structures are formed when the ratio of the radii of nonmetal to metal (\(r_n/r_m\)) is less than 0.59. This criterion fits the metals of the early transition elements (group IV–VI), with chromium being the exception (Table 1). Carbon/chromium atomic ratio is 0.61 and, strictly speaking, belongs to the intermediate class of carbides. However, the physical and chemical properties of chromium carbide are similar to the other transition carbides.

The structure of transition carbides can be described by means of a model for the simple ionic structure by two coordination numbers — one for each type of ion. They possess the rock salt structure (NaCl) or calcium fluoride structure (CaF\(_2\)). In the NaCl structure (Figure 2), the number of cations and anions are equal, and both the coordination numbers are six, so the structure is (6,6).

![FIGURE 2 Crystal structure of the interstitial carbides.](image)
The other structure type (CaF₂) is (8,4), meaning that eight anion neighbors surround each cation, and each anion by four cations.

Their elementary cell differs only due to either (i) atom radius of metal or (ii) stoichiometry of carbon atom in the structure. Many binary transition metal carbides, especially carbides from group IV elements, exist over a broad range of compositions with an upper limit of the carbon to metal ratio near 1. The chemical formula is usually written as MeC₁₋ₓ, where Me represents a transition metal, C is carbon, and x is the C/Me ratio. Ideal stoichiometry does not exist in these compounds. The values of x range from approximately 0.5 to 0.97. Thus, these materials are carbon-deficient, nonstoichiometric compounds with a wide range of vacancy concentrations [57]. Thus, interstitial carbides can be considered to be essentially nonstoichiometric materials in which stoichiometry is rarely complete. This is particularly often in the case of the Ti–C system. Figure 3 shows an example of the group IV transition metal–carbon systems. They are generally very similar to each other concerning their properties: extensive homogeneity ranges and high congruent melting point. In addition, they possess NaCl-type cubic basis crystal structure. Their structures thus consist of identically stacked, closely packed {111} planes of the metal atoms alternating with {111} planes containing the C atoms. Monocarbides from group V have the same crystal structure, as well as comparable melting points (Figure 4).

However, they contain the structurally identical subcarbide phases Me₂C and Me₄C₃₋ₓ. Also, the phases Me₂C have low- and high-temperature modifications, which differ only by arrangement of the carbon atoms on the interstitial positions [58]. The compositions of the phase Me₄C₃₋ₓ deviate from the ideal crystallochemical C/Me ratio 3/4 [51]. The other carbides with different carbon content such as Nb₃C₂, Ta₃C₂, Nb₄C₃, Ta₄C₃, V₆C₅, and V₈C₇ have been observed [59]. Deviation of stoichiometry is the largest for V₄C₃₋ₓ in which carbon content is below 40 at%.

The group VI of transition metal–carbon system, the stoichiometric monocarbides WC and MoC are hexagonal. Also, cubic monocarbides MeC₁₋ₓ in the Mo–C and W–C system exist but at high temperatures and with stoichiometric compositions. In addition, the other phases with different stoichiometries are observed (W₂C, Mo₂C, and Mo₃C₂). The Cr–C system differs from the other transition metal–carbon systems (Figure 5). This system is characterized by the absence of a monocarbide phase. The other phases are Cr₃C₂, Cr₇C₃, and Cr₂₃C₆. In the phase Cr₂₃C₆, each carbon has 8 Cr neighbors in the form of a quadratic antiprism and this structure is basically different from all other early transition metal carbides.

Intermediate transition carbides in the Mn–C and Fe–C have similarities with the Cr–C system with respect to occurring phases (Figure 6). The phase Mn₃C is isotypic

FIGURE 5  Example of phase diagram of the Group VI transition metal–carbon systems. Chromium–carbon system.
with $\text{Fe}_3\text{C}$; also, $\text{Mn}_{23}\text{C}_6$ and $\text{Mn}_{7}\text{C}_3$ are isotypic with the chromium carbide phases of the same compositions. In general, these phases are not stable. Their phase diagrams are of eutectic type. They are obtained in rapidly quenched Ni–C and Co–C alloys.

### 3.3. Properties

The carbides of the transition metals have extremely high melting points and sometimes they are collectively referred to as refractory carbides. Also, these carbides are extremely hard, retaining high hardness at very high temperature and therefore finding industrial use in cutting tools and wear-resistant parts. In addition, they possess good thermal conductivity as well as good thermal shock resistance. Although the Fe–Ni–Co–C systems are not high-temperature systems as carbides of the group IV–VI, they are the more important systems in the metallurgy of steels. As a solid solution they can persist for a very long time at temperatures up to 700°C. Their decomposition is extremely slow. WC is also used for fabrication of as cemented carbide tools for cutting steel. The addition of other carbides such as TiC, TaC, and NbC improves the working temperature by increasing the oxidation resistance of the tool. Sintered carbides of Co, Mo, and W are known as superalloys, which are used in high-temperature applications such as rocket nozzles and jet engine parts.

### 4. COVALENT CARBIDES

As mentioned in Section 1, there are only two covalent carbides: boron carbide ($\text{B}_4\text{C}$) and silicon carbide ($\text{SiC}$). However, a low abundance of boron in both the solar system and the Earth's crust is a reason why the $\text{B}_4\text{C}$ is not found naturally. On the other hand, occurrence of $\text{SiC}$ is widespread.

Silicon and carbon atoms are located one above the other in the periodic table having similar electronic structure ($\text{C} - 2s^2 2p^2$, $\text{Si} - 3s^2 3p^2$); both form four $sp^3$ hybridized orbital with a tetrahedral symmetry and the angle between the edges form $109^\circ 28'$ identical to the structure of diamond (Figure 7). The difference in the electronegativity of atoms C and Si ($2.5 - 1.8 = 0.7$) indicates that the ionic character of the relationship is $\leq10\%$. The bond strength of $\text{SiC}$ is about 300 kJ/mol, in diamond it is 356 kJ/mol.

Silicon atom is only slightly larger than carbon atom. Some general characteristics of SiC are:

1. It is a nonmetallic compound,
2. its elemental constituents have low atomic weight,
3. it has a low density,
4. its bonding is essentially covalent,
5. it is an extremely hard and very strong material,
6. it fully meets the refractory criteria (high melting point, and thermal and chemical stability), and
7. it has semiconductor properties.
Silicon carbide is the only stable compound in the silicon–carbon system [60]. It is binary phase with a composition of 70.05 wt% Si and 29.05 wt% C (Figure 8). SiC does not have a melting point, it readily decomposes at 2545 °C into carbon and a silicon-rich melt in a closed system at a total pressure of 0.1 MPa. This is the highest temperature at which SiC is formed. In the open system, SiC starts to decompose at approximately 2300 °C with generated species such as Si, SiC2, and Si2C. Also, SiC and Si degenerate into an eutectic at 1404 °C and 0.02 at% C. The solubility of carbon in liquid silicon is about 13 at% C at the peritectic temperature.

Silicon carbide exhibits a pronounced tendency to crystallize in a multitude of different modifications named polytypes [61]. All polytypes consist of close-packed layers of carbon (C) and silicon (Si) atoms, where the C atoms are situated above the centers of triangles of Si atoms and underneath the Si atoms belonging to the next layer (Figure 9). The distance between the neighboring silicon or carbon atoms (a) is approximately 0.307 nm for all polytypes. The carbon atom is positioned at the center of mass of the tetrahedron outlined by the four neighboring Si atoms so that the distance between the C atoms to each of the Si atoms is the same. Geometrical considerations give that this distance, C–Si, is approximately equal to 0.189 nm and the distance between the carbon atoms is approximately 0.307 nm [62]. The distance between two silicon planes is approximately 0.252 nm.

The short-range structure of SiC consists of tetrahedra connected together in a two-dimensional space with lattice spaces arbitrarily occupied by Si or C atoms resulting in a planar SiC4 or CSi4 configuration. These planes are arranged in a parallel or antiparallel manner for the long-range structure of SiC.

![Figure 7](image1.png) **FIGURE 7** The tetrahedral bonding of a carbon atom to four nearest silicon neighbors.

![Figure 8](image2.png) **FIGURE 8** Equilibrium phase diagram of the Si–C system.
Usually, the ABC notation, known as Ramsdel notation, is used as a more descriptive of the position of atomic layers (Figure 9). For example, the number 3 in 3C–SiC refers to the three-bilayer periodicity of the stacking ABC and the letter C denotes the cubic symmetry of the crystal. The stacking sequence ABAB shows the wurtzite crystal structure for 2H–SiC reflecting its two-bilayer stacking periodicity and hexagonal symmetry. This periodicity doubles and triples in 4H and 6H polytypes. All the rhombohedral polytypes are labeled by R (15R–SiC and 21R–SiC) (Table 4).

The parallel arrangement makes an ABCABC sequence of either Si or C layers, which has the cubic, zinc blende structure. According to a definition, the zinc blende structure is an fcc package with half of tetrahedral sites filled. Since SiC is a covalent compound it can be said that the SiC structure consists of two identical interpenetrating close packings, one of Si and the other of C. This structure is denoted as β-SiC. The antiparallel arrangement makes an ABAB sequence of layers, which has a hexagonal, wurtzite structure. It is denoted as α-SiC and it has several polytypes. The difference between the polytypes is the stacking order between the succeeding planes of the tetrahedra. As far as mechanical properties are concerned, the polytypes do not play an important role and from this point of view only α-SiC and β-SiC are important. The typical polytypes are shown in Figure 10.

The cubic polytype 3C-SiC is assumed to be more stable than the hexagonal polytypes up to 2100 °C [65]. The 2H polytype is the unstable one. In presolar grains only these two SiC polytypes occurred. This is explained by low pressure in the circumstellar shells during SiC formation [39]. Also, it is noticed that under some conditions 6H–SiC can transform to 3C–SiC. This does not commonly occur at high temperatures. During the hot pressing of 6H–SiC and TiN mixture in a flowing Ar atmosphere, all the hexagonal SiC is converted to cubic SiC above 2200 °C, although 3C polytype is not stable at this temperature. It is found that N has an essential influence on 6H–SiC transformation [66].

5. SYNTHESIS

Natural sources of transition metal carbides as well as silicon carbide are extremely rare and are a mineralogical curiosity that has no significance as raw materials. All carbide-based ceramics are derived from the synthetic materials, exclusively. The majority of carbide ceramics is made from powders and therefore depends, to a greater extent, on the quality of the starting powders. All the carbidies can be prepared by heating almost any source of corresponding metal and carbon at high temperatures [67]. Several different synthetic routes for obtaining carbide powders are available; some of them are mentioned below:

1. solid combustion synthesis by direct combination of the elements at elevated temperatures,
2. carbothermal reduction reaction of metal oxide compounds,
3. vapor phase synthesis,
4. pyrolyzes of metal-organic compounds,
5. sol–gel route,
6. laser-induced reactions, and
7. plasmachemical synthesis.

The synthesis of pure carbide powder is one of the important factors for obtaining dense ceramics. Although the quality and performance are important, the cost of mass
production is the key factor in the commercialization of carbide ceramics. With this point of view, many of the above-mentioned fabrication methods would never meet the commercialization.

5.1. Solid Combustion Synthesis

In the solid combustion synthesis of carbides, the mixed powders are compacted beforehand to increase an intimate contact between the individual metal and carbon particles. The reactant mixture is degassed and ignited in a vacuum or in an inert atmosphere [68]. Some examples of this route are:

\[
\begin{align*}
\text{Si} + \text{C} &\rightarrow \text{SiC}, \quad T = 1400^\circ \text{C}, \quad (8) \\
\text{W} + \text{C} &\rightarrow \text{WC}, \quad T = 1500^\circ \text{C}, \quad (9) \\
\text{Ta} + \text{C} &\rightarrow \text{TaC}, \quad T = 2200 - 2400^\circ \text{C}. \quad (10)
\end{align*}
\]

Combustion synthesis can be conducted in a self-propagating mode (SHS, self-propagating high-temperature synthesis). Compared with the conventional ceramic processing, the main advantages of SHS are:

1. it needs less energy than conventional methods,
2. it takes a short reaction time resulting in less processing cost and time saving,
3. shaped and unshaped products can be obtained in a single step, and
4. expensive and complicated operating equipments are not necessary.

The SHS reactions can be characterized by the adiabatic combustion temperature, \(T_{ad}\) [69]. The combustion temperature can be calculated by assuming that the enthalpy of the reaction heats up the product and no loss of energy occurs to the surrounding environment. As a rule of thumb, if \(T_{ad}\) is less than 1200 \(^\circ\)C, combustion does not occur. In case \(T_{ad}\) is higher than 2200 \(^\circ\)C, self-propagating combustion occurs. In the range between 1200 \(^\circ\)C and 2200 \(^\circ\)C, a combustion wave cannot propagate but it is possible by special techniques such as preheating of the

<table>
<thead>
<tr>
<th>Polytype</th>
<th>No. of space group</th>
<th>Lattice constant (a_0) (nm)</th>
<th>Lattice constant (c_0) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 C</td>
<td>216</td>
<td>0.43596</td>
<td>0.43596</td>
</tr>
<tr>
<td>2 H</td>
<td>186</td>
<td>0.30730</td>
<td>0.50480</td>
</tr>
<tr>
<td>4 H</td>
<td>186</td>
<td>0.30730</td>
<td>1.0053</td>
</tr>
<tr>
<td>6 H</td>
<td>186</td>
<td>0.30730</td>
<td>1.5110</td>
</tr>
<tr>
<td>8 H</td>
<td>186</td>
<td>0.30730</td>
<td>2.0147</td>
</tr>
<tr>
<td>19 H</td>
<td>156</td>
<td>0.30730</td>
<td>4.7849</td>
</tr>
<tr>
<td>21 H</td>
<td>156</td>
<td>0.30730</td>
<td>5.2870</td>
</tr>
<tr>
<td>27 H</td>
<td>156</td>
<td>0.30730</td>
<td>6.7996</td>
</tr>
<tr>
<td>36 H</td>
<td>156</td>
<td>0.30730</td>
<td>9.0650</td>
</tr>
<tr>
<td>15 R</td>
<td>160</td>
<td>0.30730</td>
<td>3.7700</td>
</tr>
<tr>
<td>21 R</td>
<td>160</td>
<td>0.30730</td>
<td>5.2890</td>
</tr>
<tr>
<td>24 R</td>
<td>160</td>
<td>0.30730</td>
<td>6.0490</td>
</tr>
<tr>
<td>27 R</td>
<td>160</td>
<td>0.30730</td>
<td>6.7996</td>
</tr>
<tr>
<td>33 R</td>
<td>160</td>
<td>0.30730</td>
<td>8.3110</td>
</tr>
<tr>
<td>45 R</td>
<td>160</td>
<td>0.30730</td>
<td>11.333</td>
</tr>
<tr>
<td>51 R</td>
<td>160</td>
<td>0.30730</td>
<td>12.843</td>
</tr>
<tr>
<td>105 R</td>
<td>160</td>
<td>0.30730</td>
<td>26.439</td>
</tr>
<tr>
<td>141 R</td>
<td>160</td>
<td>0.30730</td>
<td>35.505</td>
</tr>
<tr>
<td>189 R</td>
<td>160</td>
<td>0.30730</td>
<td>47.628</td>
</tr>
<tr>
<td>393 R</td>
<td>160</td>
<td>0.30730</td>
<td>98.760</td>
</tr>
</tbody>
</table>
reactants, or by introducing mechanical energy in the system. For example, the reaction between Si and C that results in SiC is not self-sustaining because of $T_{\text{ad}}$ (1527 °C) and the ignition temperature is 1300 °C. In the case of Ti and C, which give TiC, reaction is self-sustaining ($T_{\text{ad}} = 2937$ °C) and ignition temperature is 1027 °C.

Also, the SHS reaction can be accomplished by introducing a mechanical energy in the system by ball milling of metal powders and carbon. During the milling process, powder particles are repeatedly fragmented and cold-welded, which results in the intimate mixing of the reactant particles on the nano scale. The milling process also leads to the reduction in the crystallite size and the accumulation of defect in the powder particles, which introduces an additional energy to the reactant system and thus effectively lowers the activation barrier for

FIGURE 10 Crystal structure of typical SiC polytypes.
the reaction. For some highly exothermic reaction systems, spontaneous compound formation can occur by mechanically induced self-propagating reactions, the mechanism of which is analogous to that of the thermally ignited self-propagating high-temperature synthesis. Mechanochemical synthesis is usually carried out under room temperature conditions and therefore it is particularly attractive for the synthesis of high-melting-point carbides [70]. For example, TiC as well as β-SiC can be synthesized by a high-energy ball milling at room temperature without thermal treatment [71].

In general, combustion synthesis is difficult to control due to the high reaction rate. Some improvement of reaction control is possible by the addition of diluents, which do not take part in the reaction, but increase the thermal mass of the system and lower $T_{ad}$. The other way is to increase the particle size of the reactants resulting in the decrease of the combustion waves.

5.2. Carbothermal Reduction

Commercially, most of the carbides are obtained by the Acheson process, which is also known as the carbothermal reduction reaction. The production process can be carried out in a variety of ways in various types of reactors (electric arc furnaces, rotary tube reactors, moving bed furnaces, fluidized bed reactors, etc.). However, the basic design of the first furnace (Acheson) remained unchanged, despite better efficiency. Generally, graphite electrodes connected to a graphite core are laid in a mixture of carbon and metal oxide. When an electric current is passed through the graphite core, the reaction takes place resulting in the formation of designated carbide. Also, it is the most promising candidate for obtaining a large variety for all types of nonoxide products with important technical uses. This reaction involves reduction of oxygenated materials, such as silica ($\text{SiO}_2$) usually by mixing with a reducing agent (carbon) in excess at a temperature higher than 1600 °C for several hours under an inert atmosphere. As carbon sources can be used graphite, charcoal, carbon black, pyrolyzed organic polymers, however, the most common carbon source is the final residue from the crude oil refining known as the petroleum coke.

This procedure offers the possibility of an economically attractive production route from naturally occurring materials that possess high specific surface area, high metal content, and low price [72]. The carboreduction of oxides is the most developed production method for SiC. Many authors have studied the formation of SiC powders from the raw materials, such as high-purity quartz sand [73]. Sepiolite, diatomaceous earth, and mounting leader have been used recently [74–76].

During the thermal treatment of the silica/carbon mixture, Si or SiO is liberated in gaseous form, which further reacts with excess carbon to form SiC according to the following reaction:

$$\text{SiO}_2 + 3\text{C} \rightarrow \text{SiC} + 2\text{CO} \quad (11)$$

The formation of the final product is more complex than the above equation. There are many intermediate stages, since the formation of SiC requires a series of solid—solid, solid—gas, and gas—gas reactions [77]. It is usually accepted that SiC forms as a result of four main sub-reactions (Eqns 12–15), which provide mass transport via vapor phases:

$$\text{SiO}_2(s) + \text{C}(s) \rightarrow \text{SiO}(g) + \text{CO}(g) \quad (12)$$

$$\text{SiO}_2 + \text{CO}(g) \rightarrow \text{SiO}(g) + \text{CO}_2(g) \quad (13)$$

$$2\text{C}(s) + \text{SiO}(g) \rightarrow \text{SiC}(s) + \text{CO}(g) \quad (14)$$

$$\text{C}(s) + \text{CO}_2(g) \rightarrow 2\text{CO}(g) \quad (15)$$

Silica initially reacts at the contact point with carbon particles in the solid-state reaction liberating CO and SiO gases (Eqn 12). The reaction between CO and remaining SiO results in the formation of additional SiO and CO gases (Eqn 13). SiC is formed by the reaction of gaseous SiO directly at the surface of the carbon particles (Eqn 14). The intermediate gaseous phase CO$_2$ is re-converted to CO by Eqn 15.

Typical examples for the carbothermal reduction and the corresponding lowest occurring temperatures at atmospheric pressure are given below [69]:

$$\text{TiO}_2 + 3\text{C} \rightarrow \text{TiC} + 2\text{CO} \quad 1300^\circ \text{C} \quad (16)$$

$$2\text{MoO}_3 + 7\text{C} \rightarrow \text{Mo}_2\text{C} + 6\text{CO} \quad 500^\circ \text{C} \quad (17)$$

$$\text{WO}_3 + 4\text{C} \rightarrow \text{WC} + \text{WC} + 3\text{CO} \quad 700^\circ \text{C} \quad (18)$$

The conversion of TiO$_2$ to TiC takes place via the intermediate phases such as Ti$_3$O$_5$, Ti$_2$O$_3$, and TiO. Also, the reduction of the other oxides of transition metals proceeds via the intermediates.

Generally, carbides are prepared by the reduction of oxides with carbon; only Mo$_2$C and WC are manufactured by the reaction of the metal powders with graphite or carbon black [60].

Thermite reaction is a special type of combustion synthesis involving the reduction of a metallic compound by strongly reducing metals, such as Mg and Al. The more desirable reducing agent is easily leachable with hydrochloric acid. Reaction proceeds in two steps. The first involves the reduction of an oxide to an element and, the second, the reduced element reacts with another element to form carbide. For example, reducing silica by Mg, followed
by a reaction between Si and C for obtaining SiC by thermite-type reaction [78]:

$$\text{SiO}_2 + 2\text{Mg} + \text{C} \rightarrow \text{SiC} + 2\text{MgO} \quad (19)$$

A metal oxide is formed as a by-product. For the synthesis of pure carbide powders, the oxide has to be chemically leached.

Another example is obtaining chromium carbide by reducing Cr$_2$O$_3$ with Al in the presence of carbon:

$$3\text{Cr}_2\text{O}_3 + 6\text{Al} + 4\text{C} \rightarrow 2\text{Cr}_2\text{C}_3 + 3\text{Al}_2\text{O}_3 \quad (20)$$

This principle has a practical usage in the synthesis of WC powders from a naturally occurring tungsten mineral wolframite (FeWO$_4$):

$$\text{FeWO}_4 + \text{CaC}_2 + 2\text{Al} \rightarrow \text{WC} + \text{FeO} + \text{Al}_2\text{O}_3 + \text{C} \quad (21)$$

In principle, this technique is used when the cost of elemental powder is too high (Ta, Hf), or when the direct reaction of the element does not generate sufficient amount of heat for the self-sustaining reaction.

### 5.3. Vapor-Phase Synthesis

The vapor-phase synthesis of the carbides can be performed by the reaction of the metal or chlorides with carbon-containing gases such as methane or benzene:

$$\text{Ti} + \text{CH}_4 \rightarrow \text{TiC} + 2\text{H}_2 \quad (22)$$

$$\text{ZrCl}_4 + \text{CH}_4 \rightarrow \text{ZrC} + 4\text{HCl} \quad (23)$$

When chlorides are used, the reactions proceed in the gas phase above 600 °C. Since the transition metal halogenides are very volatile, with this route it is possible to obtain nanopowders or to use it for chemical vapor deposition of carbide layers on the solid substrate.

Also, gas-gas reaction is developed to manufacture carbides, especially SiC:

$$\text{CH}_3 + \text{SiH}_3 \rightarrow \text{SiC} + 3\text{H}_2 \quad (24)$$

$$\text{CH}_4 + \text{SiH}_4 \rightarrow \text{SiC} + 4\text{H}_2 \quad (25)$$

$$\text{SiCl}_4 + \text{CH}_4 \rightarrow \text{SiC} + 4\text{HCl} \quad (26)$$

$$7\text{SiCl}_4 + \text{C}_7\text{H}_8 + 10\text{H}_2 \rightarrow 7\text{SiC} + 28\text{HCl} \quad (27)$$

$$\text{SiCl}_4 + \text{CCl}_4 + 4\text{H}_2 \rightarrow \text{SiC} + 8\text{HCl} \quad (28)$$

$$3\text{SiH}_4 + \text{C}_3\text{H}_8 \rightarrow 3\text{SiC} + 10\text{H}_2 \quad (29)$$

The vapor-phase reactions offer the advantages of synthesizing high-purity nanopowders with a good control over size, shape, and crystal structure as well as easy control of the reaction rate. However, the method does not yield appropriate amount of product for commercialization.

The vapor-phase thermal decomposition of some organometallic compounds can form SiC, for instance,

$$\text{CH}_3\text{SiCl}_3 \rightarrow \text{SiC} + 3\text{HCl} \quad (30)$$

$$28\text{HCl} \quad (31)$$

$$\text{HCl} + \text{SiC} + 2\text{CH}_4 \rightarrow 2\text{SiC} + 2\text{H}_2 + 3\text{HCl} \quad (31)$$

### 5.4. Pyrolysis of Organic Compounds

Organometallic polymer precursors offer unique opportunities for manufacturing of ceramic components using versatile plastic shaping techniques established in the engineering plastics industry such as casting, pressing, injection molding, or extrusion [79]. Pyrolysis of metal-containing polymers has been effective in the preparation of ceramic materials at temperatures significantly lower than the conventionally sintered materials. However, only SiC has attracted significant attention for this method. The main reason is only a few compounds from the group IV and VI metals that are stable have a low carbon/metal ratio and do not contain oxygen. Also, the process suffered from the inability to scale up ceramic yields. In addition, there are still several drawbacks such as high cost and toxicity. The gas-phase decomposition of organosilanes or polycarbosilanes leads to the formation of nanocrystalline SiC [80]. Typically, gas-phase reaction of the SiH$_4$ or SiCl$_4$ with hydrocarbons such as CH$_4$ results in the ultrafine β-SiC powders [81].

### 5.5. Sol–Gel Route

Sol–gel route has been developed by mixing the reactants in the liquid phase. This route mainly focuses on SiC. For example, ethylsilicate liquid and liquid phenolic resin as the sources of silica and carbon, respectively, and toulensulfonic acid as the catalyst were used [82]. The powders obtained after the carbothermal reduction of the gel were sinterable β-SiC, containing no residual SiO$_2$ and carbon at a proper starting resin content. Similar sol–gel route has also been used to synthesize cubic SiC fibers with hydrochloric acid as the catalyst [83] and tetraethoxysilane (TEOS) and novolac phenolic resin as the starting materials, and oxalic acid and hexamethylenetetramine as the catalysts [84]. Also, the other carbides were synthesized by the sol–gel route. For example, ZrC and HfC powders were prepared using Zr-n-butoxide and Hf-isopropoxide and polyhylicr alcohol as carbon sources [85]. TaC is synthesized by the liquid precursor route starting from Ta-ethylate solution and activated carbon powders [86].

This method has become very attractive for manufacturing bioamorphous ceramics [87,88] due to an eco-friendly route for advanced ceramic materials which can be produced with renewable natural resources, such as...
wood, with tailorable properties and behavior like ceramics manufactured by conventional approaches [89].

Although most research is focused on biotemplating SiC ceramics [90–92], the transition metal carbides have been successfully manufactured by this process [93].

5.6. Others

The gas-phase process can also be achieved in plasma or laser-induced reaction. The result of this process is an amorphous or only partially crystallized powder with a high specific surface area. The gas-phase processes and the sol–gel process offer very high purity powders with high sinterability. However, these powders have a very poor shaping behavior and a high cost in comparison to the traditional routes.

The above-mentioned classification of the synthetic routes is very conventional because, in most cases, the preparation method is based on more than one basic principle [94–96].

The synthesis of pure carbide powder is one of the important factors for obtaining dense ceramics. Although the quality and performance are important, the cost of mass production should be the key factor in the commercialization of carbide ceramics.

6. EXTREME ENVIRONMENT APPLICATION

The stability that chemical compounds ordinarily have is eventually limited when exposed to extreme environmental conditions such as heat, radiation, pressure, etc. Recently, there is request for many engineering materials that can survive harsh conditions such as temperatures higher than 3000 °C, or conditions in nuclear reactors.

The current increasing interest in hypersonic vehicles with new propulsion system components, and wing leading edges and nose tips needs ultra-high-temperature materials. There are more than 300 compounds with melting temperatures over 2000 °C, including refractory metals (Hf, Nb, Re, Ir, Ta, and W), oxides (HfO2, ZrO2, UO2, and ThO2), and a variety of transition metal borides, carbides, and nitrides. However, the list of materials with melting temperatures above 3000 °C is limited to less than 15 compounds, all nonoxides. Among them, the most numerous compounds are the carbides (Table 5).

However, in the real engineering application, melting temperature is only one of the many properties used in the material selection process [97], and other properties include high-temperature strength, thermal conductivity, thermal expansion coefficient, density, and cost. However, the oxidation behavior is the second-ranking property associated with the material selection for the extreme environmental situation. The main reason is that the materials applied will involve exposure to oxidizing fuels or aeroheating [98]. All nonoxide compounds will undergo oxidation and undergo a significant weakening of the mechanical properties. There are some oxides with high melting point; however, their poor thermal shock resistance, because of the low thermal conductivity as well as high thermal expansion, eliminates them in extreme environment applications. On the other hand, SiC exhibits excellent oxidation resistance up to 1700 °C because of the formation of a layer of silica glass that inhibits oxygen diffusion to the parent material [99,100]. This is the case of passive oxidation according to the reaction (Eqn 32). Long-term usage leads to the evolution of a protective oxide layer mainly consisting of silica [101]:

$$2\text{SiC} + 3\text{O}_2 \rightarrow 2\text{SiO}_2 + 2\text{CO} \quad (32)$$

However, active oxidation, which causes corrosion of the substrate materials, does not form a protective layer due to the formation and volatilization of the silicon monoxide [102]:

$$\text{SiC} + \text{O}_2 \rightarrow \text{SiO} + \text{CO} \quad (33)$$

Under reduced pressure environments, degradation of SiC and decomposition of already formed silica layer limited the application of SiC as an ultra-high temperature ceramic material. The maximum usable temperature of SiC is limited to 1700 °C, which cannot satisfy the new engineering requests.

Transition-metal carbides from the group IV and V elements have superior melting points; however, not all of

| TABLE 5 Compound With Melting Temperature Above 3000 °C [99] |
|-----------------|-----------------|------------------|
| Compound        | Melting temperature (°C) | Existing in nature |
| TaC             | 3980             | yes              |
| HfC             | 3928             | ?                |
| ZrC             | 3420             | yes              |
| NbC             | 3600             | yes              |
| TiC             | 3067             | yes              |
| TaN             | 3097             | –                |
| HiN             | 3387             | –                |
| TaB2            | >3000            | –                |
| HB2             | 3380             | –                |
| ZrB2            | 3220             | –                |
| BN              | 3000             | –                |
them can serve in harsh environments. For example, TiC and NbC form oxides TiO2 and Nb2O3, which possess low melting points, 1840 and 1485 °C, respectively. The potential candidates are TaC and HfC compounds. They possess the highest melting temperatures of any compounds (3980 and 3928 °C, respectively). In addition, they have very high hardness and elastic modulus. The main difference between HfC and TaC is the melting temperature of the oxide formed. HfO2 has a melting point of 2758 °C, while Ta2O5 has a much lower melting point, 1872 °C. HfC is a better choice and it is the reason why HfC is studied in more detail. It was shown that HfC has excellent oxidation behavior at temperatures above 1800 °C due to the densified oxide formed. However, at temperatures below 1500 °C, formed oxide grains are not able to sinter, causing them to spontaneously spall of the parent material.

7. IMPORTANCE OF NATURAL AND SYNTHETIC CARBIDES

Both terrestrial and solar carbides do not have technological significance. However, the importance is reflected in the fact that carbides may be a link with the theory of the Earth’s carbon cycle as well as the theory of galactic chemical evolution. For example, it is found that SiC is richer in 12C than relative to the standard. Isotope fractionation might have occurred through an isotope exchange reaction in a common carbon reservoir [19]. Mineral SiC may thus ultimately provide information on carbon cycling in the Earth’s mantle. Since SiC can be formed in highly reduced conditions, its occurrence is very important for understanding the existence of regions in the deep Earth, influence of the oxidation state on the composition of vapor, the geochemistry of the chalcophile and siderophile elements, and the manner of partial melting of the mantle in highly reduced environments [103,104].

On the other hand, carbides as more resistant parts in meteorite are essentially a frozen piece of a single star that ended their life before the formation of the solar system [105]. Therefore, carbides among other resistive minerals (alumina, diamond, spinels, etc.) provide a unique insight into the evolution of galaxy and can be used to constrain corresponding models. Presolar grain microstructures provide information about the physical and chemical conditions of dust formation in stellar environments [106]. The detailed characterization of these grains can yield information on the pressure, temperature, and chemical composition in the stellar atmospheres and ejecta. Therefore, the microscopic grains reveal more detail about the evolution of galaxy than by the astronomical observation with the currently possible best telescopes.

Man-made carbides that have the same composition as natural carbides are considered to be the main focus of research because of the superior melting temperatures and formation of stable high-melting temperature oxides. The combination of their properties such as high hardness, good chemical stability, and strength at high temperatures makes these materials potential candidates for a variety of high-temperature structural applications such as plasma arc electrodes, cutting tools, furnace elements, hypersonic vehicles, and high-temperature shielding [98].

REFERENCES


[63] Cheung R. Silicon carbide microelectromechanical system for harsh environments. Imperial College London; 2006.


INTRODUCTION

The applications of silicon nitrides mainly utilize their mechanical properties such as high strength, high durability, high wear resistance, as well as chemical stability at room and high temperatures. Recently, nitride ceramics have received much more attention in the field of electronics, in the context of the following social background. In this paper, highly functionalized silicon nitride and SiAlON and their applications are reviewed.

High Thermal Conductivity Silicon Nitride Ceramics

Energy and environment-related problems are serious social issues. In order to save energy as well as to reduce the emission of carbon dioxide, energy sources tend to shift from fossil fuel to electric power; hence, highly efficient use of electric power becomes extremely important. Power devices are key technologies for this purpose. Power devices conduct direct conversion and control of electric power by means of semiconductors, which can drastically save energy compared to traditional systems. Power supply and packing density of power modules are increasing more and more as their application field expands, in particular in the automobile industry.

In order to guarantee the stable operation of the power module, heat release technology in the module becomes of great importance. So far, AlN has been used as a substrate for power devices. In general, a thick metal plate as an electrode (about 0.3—0.6 mm in thickness) is directly bonded to a ceramic substrate via high-temperature brazing as illustrated in Figure 1.1. With an increase of the power density in the module, micro-cracks in the ceramic caused by the large difference in thermal expansion coefficient of metal and ceramic become a serious problem. Such a situation leads to strong demands for insulating substrates to
have higher mechanical properties, and thereby silicon nitrides are receiving a lot of attention.

Figure 1.2 summarizes the properties of commercial ceramic substrates. The thermal conductivities of commercially available silicon nitrides are less than 90 W/(mK), though their strength is more than twice as high as those of aluminum nitrides. An important issue in silicon nitride is, therefore, to increase thermal conductivity without degradation of mechanical properties. So far, some researchers have estimated the intrinsic thermal conductivity of $\beta$-Si$_3$N$_4$ crystal based on solid-state physics [1–3]. From these fundamental simulations, it could be anticipated that the intrinsic thermal conductivity of silicon nitride is over 200 W/(mK) [4].

Being different from the ideal single crystal, a sintered silicon nitride is composed of $\beta$-Si$_3$N$_4$ grains as well as two grain boundaries, and secondary phases with low thermal conductivities. In addition, inside the grains, there are several kinds of crystal imperfections such as point, line, and planer defects as shown in Figure 1.3 [5,6]. These defects scatter phonons to decrease thermal conductivity of the $\beta$-Si$_3$N$_4$ crystal. Similar to the case of high thermal conductivity AlN, lattice oxygen, or oxygen dissolving in $\beta$-Si$_3$N$_4$ lattice, drastically decreases the thermal conductivity of silicon nitrides as shown in Figure 1.4 [5,7].

In order to reduce these negative parameters, high thermal conductivity Si$_3$N$_4$ is generally fabricated with the following points in mind [4]. (1) High-purity fine Si$_3$N$_4$ is employed in order to reduce impurities. (2) Concurrent addition of rare earth oxide and alkaline earth oxides is conducted; the former plays a role of decreasing lattice oxygen and latter is needed for assisting densification. With respect to alkaline earth element, the use of MgSiN$_2$ as one of the sintering additives was found to be effective for improving thermal conductivity [8]. (3) Enhancement of grain growth is one of the important factors for improving thermal conductivity because it can decrease the number of two grain junctions as well as the lattice oxygen via solution re-precipitation process [4].

In practice, gas-pressure sintering (GPS) with long sintering time over several tens of hours is necessary in order to achieve a high thermal conductivity over 100 W/(mK). Figure 1.5 shows a relation between thermal conductivity and bending strength of GPS silicon nitrides. Considerable grain growth is needed in order to improve thermal conductivity over 100 W/(mK), which results in a drastic decrease in strength. Even in the high-purity powder it
contains about 1.2 mass% of impurity oxygen, one-third of which is lattice oxygen, which makes it difficult to fabricate high thermal conductivity Si3N4 with high strength via the GPS method.

In order to overcome this problem, a research group at AIST, Japan, has taken notice of reaction-bonding process [9–11]. When a Si powder compact is heated around 1400 °C in nitrogen atmosphere, it is nitrided to form a silicon nitride compact. As the nitridation reaction proceeds mainly via gas-solid reaction system, the external dimensions of the compact retains while each of the individual Si particles expands by about 22%. Consequently, the silicon nitride compact with higher density can be obtained without dimensional change. Well-known advantages of reaction-bonding (RB) process include the following: (1) cheaper Si powder can be employed as starting powder rather than the more expensive Si3N4 powder and (2) high dimensional accuracy due to low shrinkage after post-sintering [12]. In addition, it is possible to carry out the whole process from nitridation to post-sintering process without exposing the compacts to the air, which is favorable for controlling the oxygen content in Si3N4.

Zhou et al. have fabricated Si3N4 ceramics by nitriding Si compacts with Y2O3 and MgO as sintering additive at 1400 °C for 8 h followed by post-sintering at 1900 °C for 6 and 24 hrs, along with their mechanical properties [11]. Specimen sintered for 6 h exhibits good mechanical properties combing high strength over 700 MPa and high fracture toughness of 8.4 MPa · m1/2. The specimen sintered for 24 h has a moderate strength of about 600 MPa and extreme high fracture toughness over 11 MPa · m1/2. Well-developed rod-like grains can be seen in the fractured surface of SRBSNs, which is one reason for high KIC of the materials. In addition, no use of Al2O3 as a sintering additive might be another factor for high KIC of the specimens. It is well known that Al2O3 dissolves into β-Si3N4 lattice to form β-SiAlON, which results in strong interfacial strength. The situation in this study is opposite to the case including Al2O3.

---

**TABLE 1.1 Characteristics of Raw Si Powder. Ref. [11]**

<table>
<thead>
<tr>
<th></th>
<th>Si powder (for SRBSN)</th>
<th>Si3N4 powder (for GPS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle size</td>
<td>7 µm</td>
<td>0.1 µm</td>
</tr>
<tr>
<td>Purity (except oxygen)</td>
<td>&gt;99.9 %</td>
<td>&gt;99.9 %</td>
</tr>
<tr>
<td>Oxygen content</td>
<td>0.34 mass% (as received)</td>
<td>–</td>
</tr>
<tr>
<td>Oxygen content</td>
<td>0.70 mass% (after milling)</td>
<td>–</td>
</tr>
<tr>
<td>Oxygen content</td>
<td>0.42 mass%2 Si3N4 (after nitridation)2</td>
<td>1.2 mass%</td>
</tr>
</tbody>
</table>

1E-10 grade powder, UBE industry Co., Japan.
2Estimated oxygen content in a specimen after complete nitridation.

---

**FIGURE 1.6** Thermal conductivity of sintered reaction-bonded Si3N4 (SRBSN) and gas-pressure sintered Si3N4 (GPS) sintered at 1900 °C under 0.9 MPa N2.
In summary, compared with the GPS materials prepared from a high-purity Si3N4 powder, the SRBSN materials from Si powder with low oxygen impurity exhibited a better balance between thermal conductivity and bending strength. The reaction bonding of silicon powder compacts followed by post-sintering is a promising way of the fabrication of Si3N4 with both high thermal conductivity and high strength.

REFERENCES


Development of Silicon Nitride for High-Temperature Use

There are some candidate ceramics for high-temperature use. Silicon nitride is one of them, which shows higher fracture toughness than the other ceramics. (Table 2.1) This is why silicon nitride has been developed for high-temperature use.

A well-known silicon nitride, SN282 by Kyocera Co. Ltd., was developed for ceramic gas turbine components. [1,2] This silicon nitride has been improved for practical requirement. SN282 contains lutetium oxide, Lu2O3, as sintering additive. Microstructure is shown in Figure 2.1, and properties are summarized in Table 2.2.

Many research works were done following SN282 to reveal why SN282 shows excellent high-temperature properties. Figure 2.2 shows a phase diagram of Si3N4–Lu2O3–SiO2 system.

Some research focused on the ratio of Lu2O3 and SiO2. As this system has grass formation region in the Si3N4–Lu2SiO5–Lu2Si2O7 triangle, this region is mentioned “G” in the triangle. At an elevated temperature above 1400 °C, the glass is softened to degrade high-temperature strength. Difference between the composition “A” and “B” is small. However, microstructure, strength at room and elevated temperatures, and fracture toughness of the gas-pressure sintered specimens are quite different [4]. Properties are summarized in Table 2.2. SN282 has

<table>
<thead>
<tr>
<th>TABLE 2.1 Properties of Major Ceramics for High-Temperature Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oxidation resistance</strong></td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Silicon Nitride</td>
</tr>
<tr>
<td>Silicon Carbide</td>
</tr>
<tr>
<td>Alumina</td>
</tr>
</tbody>
</table>

FIGURE 1.7 Fractured surfaces of sintered reaction-bonded Si3N4 sintered at 1900 °C for (a) 6 h and (b) for 24 h under 0.9 MPa N2. Thermal conductivity, 4-point bending strength, and fracture toughness are 105 W/(mK), 736 MPa, and 8.4 MPa·m1/2 for the specimen (a), and 132 W/(mK), 516 MPa, and 11 MPa·m1/2 for the specimen (b).
a composition similar to “A”. In composition “A” located near the line Si₃N₄–Lu₂Si₂O₇, therefore, a small amount of glassy phase forms during sintering. This leads to the smaller grain size. Strength and fracture toughness at room temperature are low due to the small amount of grain boundary glassy phase. At elevated temperatures, however, a smaller amount of glassy phase is advantageous to maintain the strength. Composition “B” is located inside the Si₃N₄–Lu₂SiO₅–Lu₂Si₂O₇ triangle. As a larger amount of glassy phase formed, properties were contrastive to “A”.

To improve the properties of composition “A”, two techniques were tried. One is adding a small amount of Yb₂O₃ [5] and another is making anisotropic microstructure. [6] By adding 0.2 wt% Yb₂O₃, the strength of gas-pressure sintered specimen at RT was same as in “A” but strength at 1500 °C and fracture toughness at RT are improved to 480 MPa and 6 MPa·m¹/², respectively. By making anisotropic microstructure using forging technique, strength along the alignment direction at RT and 1500 °C is improved to 950 MPa and 700 MPa, respectively (Table 2.3).

Another work to change the composition far from “A” and “B” was done. Composition “C”, which is located near the line Si₃N₄–Lu₄Si₂O₇N₄, is chosen. [7,8] This composition is expected to form a small amount of glassy phase at elevated temperatures. Therefore, the hot-pressed specimen shows a very high strength of 800 MPa at 1500 °C as well as creep resistance.

On May 21, 2010, Japanese Venus Climate Orbiter “AKATSUKI (PLANET-C)” was launched. [3,9] “AKATSUKI” has a ceramic thruster, made of the silicon nitride, SN282, for the first time in the world. On June 28, 2010, the thruster jetted for 13 seconds, and on-orbit verification was successfully performed for the first time in the world. Ceramic gas turbine is far from practical use, but the silicon nitride developed for the turbine is successfully transferred to aerospace use as a thruster.

**REFERENCES**


Low-Friction Si$_3$N$_4$ Ceramics with Carbon Fiber

Ceramic materials exhibit properties that make them suitable candidates for a number of industrial applications. In the field of machinery industries, they are expected to be used where the tribological environment is severe, such as under high normal pressures and corrosive environments [1,2]. In particular, there is the possibility of using ceramics in tribological applications under nonlubricated or marginally lubricated sliding contacts.

Compared with other engineering ceramics, silicon nitride has superior mechanical properties, with high fracture strength and fracture toughness. In tribological applications, Si$_3$N$_4$ ceramics have been used as the ball component of bearing systems under lubricated conditions. However, the sliding contact of Si$_3$N$_4$—Si$_3$N$_4$ self-mated tribopairs under dry conditions produces a high friction coefficient and high wear rate because the abrasive wear is affected by the intrinsic brittle nature of ceramics [3]. Therefore, many attempts have been carried out to produce Si$_3$N$_4$-based ceramics with a solid lubrication function, and many researchers have reported self-lubricating materials. Among these materials, graphite possesses one of the highest solid lubricating functions. However, technical problems arise during the fabrication of ceramic graphite composites, since the fine graphite powder employed is difficult to mix with the raw ceramic powder due to its hydrophobic nature, and results in adhesion to the balls and pot used for the ball milling. In graphite materials, there is little work reporting on the solid lubrication properties of ceramic/carbon fiber composites. It has been reported that the solid-lubrication effect of ceramic/carbon composites is not sufficient to allow their use in tribological applications under dry sliding conditions. It has been reported that under dry sliding conditions against a stainless steel counterbody, a Si$_3$N$_4$ ceramic/carbon fiber composite did not produce any beneficial lubricating effect when compared with the results of the matrix material alone [4]. It was suggested that this was because the formation of a graphitic lubricating film on the surface of the counterbody did not occur, due to the inherent properties of the fiber employed. The properties of carbon fibers, in particular the elastic modulus, are determined by the degree of texturing of the graphite bonding layers and by the amount of graphitic bonding in the carbon fiber composition. Our group has recently shown that fibers with a high degree of orientation and high graphite content were effective for decreasing the friction coefficient under dry sliding conditions [5,6]. We introduced this Si$_3$N$_4$/carbon short fiber composite. Figure 3.1 shows SEM images of the cross section of the fiber with high elastic modulus. It can be seen that the layering is more defined in the fiber. These results indicated that the graphite content was higher, the size of graphite crystals was larger, and the graphite layer structure had a higher radial orientation in the fibers with higher tensile modulus. For fabrication of carbon fiber-dispersed Si$_3$N$_4$-based composite, the hot-press method was chosen. The composition of the starting powders was 93 mass % Si$_3$N$_4$, 2 mass % Al$_2$O$_3$, and 5 mass % Y$_2$O$_3$. The raw powder was homogeneously mixed by ball milling and then dried. After sieving the dried powders, they were mixed with 5 vol % of the respective carbon fibers using planetary ball milling. Prior to mixing, the fibers were cut to lengths of 3—5 mm. The composite powders were hot-pressed at 1950 °C for 2 h at 30 MPa pressure in a 900-kPa nitrogen atmosphere. The microstructure of the obtained composite is shown in Figure 3.2. Homogeneously dispersed carbon fibers of a few hundred microns in length were observed in the composite. The bending strength of the composite was around half that of the monolithic Si$_3$N$_4$, indicating that the carbon fibers act as defects. On the other hand, fracture toughness was increased up to about 10 MPa·m$^{1/2}$. Fiber pull-out was observed on the fractured surface of this composite as reported by many researchers [7,8] for ceramics/fiber composites, and this mechanism was responsible for the toughening effects of

![Figure 3.1](image-url) SEM images of the cross section of the carbon fibers with high elastic modulus.
the Si₃N₄/carbon short fiber composite. Table 3.1 shows the tribological test results of this composite and a conventional Si₃N₄ ceramics under dry sliding test. In the case of monolithic Si₃N₄, the initial friction coefficient indicated higher values, which were maintained in the steady state. On the other hand, although the initial friction coefficient of the Si₃N₄/carbon fiber composites also indicated high values, it gradually decreased and was in the range of 0.25.

Figure 3.3 shows worn tracks of disks of this composite and a conventional Si₃N₄ ceramics under dry sliding test. The worn track of monolithic Si₃N₄ showed a rough surface with fine crushed debris. On the other hand, for the composite specimens, the graphite fibers were preferentially worn in all cases. In addition, the worn tracks were covered with adhesive debris with a smooth appearance, which may have consisted mainly of SiO₂ formed by tribochemical reaction [9] and graphite. These results indicated that the graphite fibers inside the composite were supplied to the worn surface as a result of partial fracture followed by grain dropping during the wear test and acted as solid lubrication with adhesion on the surface depleting the graphite layer structure of the carbon fibers. The bonding energy between crystal layers of graphite is weak, such that slip between crystal layers is assumed to be one of the dominant mechanisms of low-friction sliding. The friction coefficient of graphite has been reported to be <0.1 up to 600 °C under dry conditions. To clarify the graphite lubrication effects, the friction coefficient of a Si₃N₄ dipped in graphite powder was evaluated under the same conditions and indicated a value of 0.15. The friction coefficient of the fabricated composites indicated almost the same values as that of the graphite dipped sample. It was considered from the values of friction coefficient of these composites that the sliding condition was almost a graphite—graphite contact. The friction coefficient did not decrease with further increasing of the graphite content above the 5 vol.% graphite fiber addition because the graphitic content of this composite was sufficient to maintain the graphite—graphite contact. The solid-lubrication effect of the graphite fibers restricted the increasing of Hertzian stress contact during sliding conditions. In addition, this low-friction behavior is not only a dry sliding condition but also a water lubrication condition [10].

REFERENCES


**TABLE 3.1 Tribological Test Results of this Composite and a Conventional Si₃N₄ Ceramics Under Dry Sliding Test**

<table>
<thead>
<tr>
<th></th>
<th>Friction coefficient</th>
<th>Specific wear rate (×10⁻⁶ mm³/Nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si₃N₄</td>
<td>0.7</td>
<td>90.0</td>
</tr>
<tr>
<td>Carbon fiber dispersed Si₃N₄</td>
<td>0.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**FIGURE 3.2** The microstructure of the obtained the composite.

**FIGURE 3.3** Worn tracks of disks of this composite and a conventional Si₃N₄ ceramics under dry sliding test.
Frictional Properties and Microstructure of Si$_3$N$_4$ Containing Mo and Fe Compounds Prepared by Hot Pressing [1,2]

4.1. INTRODUCTION

In the boundary lubrication region, the coefficient of friction is much larger than that observed in the fluid lubrication region. In engine design, operation in the fluid lubrication region is typically assumed. The use of lubricating oil with a lower viscosity is an effective means of reducing the friction force in this region. However, when using a lower-viscosity oil, it is difficult to prevent wear and the development of seizure under boundary lubrication conditions, which could occur when the cylinder liner and piston are made of metals such as cast iron or aluminum.

An attempt has been made to use silicon nitride, whose wear resistance and resistance to seizure development are much better than those of metals. However, in ceramics, the formation of a lubricating oil film on a sliding surface is hindered because of the poor reactivity (adsorption power) of ceramics relative to lubricating oil. Moreover, the coefficient of friction should not be held at a low level under mixed and boundary lubrication conditions to avoid generating a high shearing force during solid-to-solid contacts.

In the present work, high-strength silicon nitride, in the form of a composite that contains a molybdenum-type solid lubricant and iron compounds with high oil adsorption properties, was developed and using a new process. The microstructure of the developed material is investigated, and its mechanical properties and sliding characteristics are evaluated with the objective of clarifying its behavior in the region of boundary lubrication.

4.2. EXPERIMENTAL METHOD

4.2.1. Production of Silicon Nitride Composite Containing Molybdenum and Iron

MoO$_3$ is chosen as a solid lubricant. As it is an oxide, it has a rather low melting point (795 °C) and a rather high vapor pressure. Therefore, normal sintering in a nitrogen atmosphere by simply adding the solid lubricant to silicon nitride does not produce a dense silicon nitride sinter with dispersed MoO$_3$. On the other hand, molybdenum, which is a structural element of MoO$_3$, oxidizes easily, but the silicon nitride used as a matrix material is highly resistant to oxidation.

Therefore, MoO$_3$ is not admixed as a starting raw material. Molybdenum is first dispersed in the silicon nitride matrix; silicon nitride with dispersed MoO$_3$ is then obtained by oxidation, and silicon nitride is produced as a composite containing the molybdenum compound and an iron compound by the method described below. The starting raw materials are $\alpha$-Si$_3$N$_4$ (Ube Kosan Co., SN-E10), Y$_2$O$_3$ (Shinetsu Chemical Industries, RU-P), Al$_2$O$_3$ (Showa Denko Co., UA5305), Fe$_3$O$_4$ (Kojundo Chemical Lab. Co., Ltd., purity 99.9%), and Mo powder (Nippon Shinkinzoku Co., average grain size 1 μm). The weight ratios of the mixture composition are Si$_3$N$_4$:Y$_2$O$_3$:Al$_2$O$_3$:Fe$_3$O$_4$:Mo = 70.8:3.7:3.7:8.3:13.5. At this point, the volumetric ratio of Mo is 5%. The powders are initially weighed to obtain precise amounts; they are then mixed for 24 h in a ball mill using a polyethylene pot, balls made of silicon nitride, and ethanol as a solvent. Then, the solvent is removed from the resulting slurry by drying, and an average mixture powder is obtained by passage through a sieve with a mesh size of 150 μm. A carbon die is filled with the obtained mixture powder and hot-pressed under the following conditions: nitrogen atmosphere under a pressure of 0.98 MPa, sintering temperature of 1750 °C, retention time of 4 h, and pressure of 40 MPa. The resulting sintered body is machined to impart a certain configuration, which is explained below, to the surface to be studied.

Next, the molybdenum phase of the surface is subjected to oxidation by heat treatment of the sintered material in an air furnace at a temperature of 700 °C and a retention time of 10 h, yielding a test specimen. Also, traditional silicon nitride without added Mo and Fe$_3$O$_4$ and silicon nitride containing only Fe$_3$O$_4$ are prepared and tested for comparison.

4.3. EXPERIMENTAL RESULTS AND DISCUSSION

The structure of the prepared sintered composite material is shown in Figure 4.1. The gray region is the matrix, which consists of silicon nitride; white specks observed in the figure are the dispersed phase, including Mo. The dispersed phase is observed as evenly dispersed microscopic specks. In the gray region, pillar-shaped areas, somewhat darker in color than the surrounding area, appear. These are grains of silicon nitride enlarge by grain growth. Notable pores are not observed in the figure, which indicates that the obtained material is dense.
sintered material. The relative density calculated from the compounding ratio of the sintered composite material is 96%.

The Striebeck diagram constructed by estimating the sliding characteristics of the sintered composite material before and after the oxidation treatment is shown in Figure 4.2. The results obtained for conventional silicon nitride prepared by using Y2O3 and Al2O3 as sintering agents (referred to below as conventional material A) and those obtained for silicon nitride with added Fe3O4, which shows a lower coefficient of friction in the regions of fluid and mixed lubrication (referred to below as conventional material B), are given for comparison. In this case, the Striebeck coefficient taken on the horizontal axis is the logarithm of a value obtained by multiplying the dynamic viscosity (m²/s) of the lubricating oil, which depends on the temperature, by the maximum speed during the reciprocating motion (m/s; the speed at the middle point of a stroke) and dividing by the applied load (N).

At a Striebeck coefficient ranging from $-7.0$ to $-7.4$, the coefficient of friction of conventional materials A and B increases sharply. After the oxidation treatment, silicon nitride with added microscopic amounts of Mo and Fe3O4 (referred to below as Mo/Fe–Si3N4) shows a coefficient of friction somewhat larger than that of conventional materials at higher Striebeck coefficients ($>-8$). At lower Striebeck coefficients ($<-8$; conventional materials operating in the mixed and boundary lubrication regions), the sharp increase in the coefficient of friction is not observed, and the coefficient remains at a low level. In the figure, the point for Mo/Fe–Si3N4 farthest to the left shows a Striebeck coefficient (low speed of sliding, high-temperature, high-loading conditions) that is lower than that of either type of conventional material. Therefore, the Mo/Fe–Si3N4 material can clearly operate under conditions that are harsher than those that conventional materials can withstand. On the other hand, before the oxidation treatment, Mo/Fe–Si3N4 shows sliding characteristics similar to those observed in conventional materials; thus, the oxidation treatment clearly improves the sliding characteristics of this material.

The four-point bending strength of Mo/Fe–Si3N4 is 880 MPa, which is equal to that of conventional material A, but under boundary lubrication conditions, the coefficient of friction is less than one-third of that for conventional material A. The obtained results are explained by the evidence for effective solid-state lubrication by admixed Mo and by the absence of large defects that would degrade the performance of sintered composite silicon nitride material. It follows from the above that Mo/Fe–Si3N4 has a high strength, similar to that of conventional materials, but its sliding characteristics are much better.
Low-cost Fabrication of Silicon Nitride Ceramics

Silicon nitride is a promising material for high-temperature and high mechanical stress applications because it offers high thermal shock resistance and high strength retention at elevated temperature as well as high erosion resistance, high corrosion resistance, and low creep properties [1,2,3]. On the other hand, the material cost is higher than other ceramics, such as alumina and zirconia; so, the application area of silicon nitride ceramics was limited due to the high price.

Among the different classes of silicon nitride ceramics, reaction-bonded silicon nitride (RBSN) has some advantages, such as low sintering shrinkage, low raw material cost, and high strength retention at elevated temperature [4,5,6]. For a typical procedure for the fabrication of RBSN, at first, a silicon powder compact is produced. The silicon powder compact is heated at temperatures in the range of 1250–1450 °C in nitrogen atmosphere, when nitridation of silicon occurred. Obtained silicon nitride ceramics has high fracture strength and fracture toughness as same as the ceramics produced by conventional processing using silicon nitride powder as a starting material.

However, the application of these RBSN ceramics is limited due to the low fracture strength, high porosity, and long processing time. To overcome one of these problems, post-sintering method was proposed. This method is almost the same procedure without added sintering additives in silicon powder and re-heating at temperatures in the range of 1700–1900 °C after reaction bonding. The post-sintering method is an attractive method for cost reduction of silicon nitride ceramics. On the other hand, when a large-size silicon powder, which is of very low price, was used as a starting powder, the mechanical properties of the obtained silicon nitride ceramics are low due to the residual unreacted silicon. To solve this problem, enhancing a nitridation of silicon during reaction-bonding process was developed for obtaining a silicon nitride ceramics with high fracture strength even if a large-size silicon was used as a starting powder [7,8,9].

Figure 5.1 shows the weight gain ratio of various oxide additions on silicon powder using thermogravimetric analysis in nitrogen atmosphere. Iron oxide addition, which is a famous nitridation catalyst of silicon, indicates a high weight gain ratio. The oxide additions, which are a conventional sintering additive for densification of silicon nitride such as Al2O3, MgO, and Y2O3, a large weight gain was not observed compared with silicon powder only. On the other hand, ZrO2 addition indicated the highest weight gain than the other oxide. Silicon nitride can be formed from its elemental constituents according to the following reaction:

$$3\text{Si} + 2\text{N}_2 \rightarrow \text{Si}_3\text{N}_4 \quad (5.1)$$

According to this reaction, the increase in the mass of a silicon compact when fully nitrided is 1.665 times the original mass of silicon. From this theoretical weight gain, the actual mass increase can therefore be used as a reflection of the degree of conversion from Si to Si3N4. These results imply that ZrO2 and Fe2O3 additions to silicon powder enhance the nitridation of silicon powder. Fe2O3 has a nitridation-enchaining effect of silicon but Fe acts as a defect in sintered body. On the other hand, ZrO2 is known for being one of a sintering additive of silicon nitride ceramics.
Review and Overview of Silicon Nitride and SiAlON, Including their Applications

and a calculation by the thermodynamics from a phase diagram in each temperature as follows:

\[
\begin{align*}
\text{N}_2(g) &+ 2\text{ZrO}_2(\text{sl}) + 2\text{Si}(\text{sl}) \rightarrow 2\text{SiO}_2(\text{sl}) + 2\text{ZrN}(\text{sl}) \quad (5.2) \\
\text{N}_2(g) &+ 2\text{ZrO}_2(\text{sl}) + 4\text{Si}(g) \rightarrow 4\text{SiO}(g) + 2\text{ZrN}(\text{sl}) \quad (5.3)
\end{align*}
\]

During heating, the original ZrO\(_2\) inside the specimen as a starting material forms ZrN by one or both of the above reactions. Weiss et al. have reported on the Si–Zr–Al–O–N phase equilibria and showed that ZrN reacts with SiO\(_2\) and forms ZrO\(_2\), SiO\(_2\), and N\(_2\) at a temperature of 1673 K. This reaction implies that the formed ZrN can have a nitrogen supplying effect at temperatures in excess of 1673 K, whereby the ZrN formed by either reaction releases N\(_2\) gas at 1673 K according to the following reaction:

\[
2\text{SiO}_2 + \text{ZrN} \rightarrow \text{ZrO}_2 + \text{SiO} + 0.5\text{N}_2 \quad (5.4)
\]

In the center of a reaction-bonded compact, it is not sufficient to only supply nitrogen gas for converting the silicon to silicon nitride because the nitridation of the surface of the specimen results in an increased volume owing to the formation of silicon nitride. This results in increased densification of the surface, making it more difficult for nitrogen to penetrate to the center. With ZrO\(_2\) additions, the nitridation is promoted by forming ZrN inside the green compact at lower temperatures. The ZrN then reacts with oxygen and transforms back to ZrO\(_2\), while the SiO gas produced forms the silicon nitride by reacting with nitrogen and releases the oxygen.

It is considered that the transformation to ZrN and then back conversion to ZrO\(_2\), as described by Eqn (5.4), can occur repeatedly and the nitridation is promoted by this circulating reaction mechanism, even when coarse silicon powder is used. After the residual silicon has disappeared, ZrO\(_2\) is increased with decreasing ZrN due to the termination of reaction 2. This fact agrees with the results of X-ray diffraction. However, according to literature, this reaction does not proceed below 1673 K (1400 °C). This paradox can be explained by the exothermic reaction of silicon and ZrO\(_2\) during nitridation. It is considered that the local temperature, particularly inside the compact, may reach a higher temperature than that measured and controlled because of the heat generated through the nitridation of Si as indicated by Eqn (5.1) and, as a result, reaction 4 may be promoted.

Only ZrO\(_2\) addition can’t densify silicon nitride ceramics. For fabrication of a dense silicon nitride ceramics via reaction-bonding route, MgAl\(_2\)O\(_4\) also was added to the silicon powder with ZrO\(_2\), and a reaction-bonded process was carried out. The nitridation-enhancing effect (low temperature and short time nitridation) was confirmed by co-doping ZrO\(_2\) and MgAl\(_2\)O\(_4\), and then the reaction-bonded bodies were sintered at 1750 °C for 8 h in nitrogen atmosphere at the pressure of 0.9 MPa. The fracture strength and fracture toughness of the obtained silicon nitride ceramics are shown in Figure 5.3. The fracture strength and the fracture toughness of the conventional silicon nitride ceramics, which used a high-purity silicon nitride powder and a same additive system, were 950 MPa and 6.0 MPa m\(^{1/2}\), respectively. In the other hand, in the case of using a fine Si powder, the fracture strength and toughness are 780–880 MPa and 5.4–6.0 MPa m\(^{1/2}\), respectively. Furthermore, in the case of using a coarse silicon powder, the fracture strength and toughness are 600–750 MPa and 5.4–6.0 MPa m\(^{1/2}\), respectively. In the case of using coarse silicon with Al\(_2\)O\(_3\)–Y\(_2\)O\(_3\) system, the fracture strength and toughness are lower than those of ZrO\(_2\)–MgAl\(_2\)O\(_4\), and the values are 480 MPa and 5.0 MPa m\(^{1/2}\), respectively. This lower mechanical property of the Al\(_2\)O\(_3\)–Y\(_2\)O\(_3\) system is due to the residual silicon inside the sintered body. The inside the sintered body, ZrO\(_2\) acts as a function of nitrogen supplier. In addition, in this process it is possible to reduce the total time during the reaction-bonded process due to the
rapid nitridation effect. The power consumption during the reaction-bonded process was reduced to about 60%, and the total time is only half.

REFERENCES


Sub Chapter 6

High-Strength and High-Toughness Silicon Nitride Ceramics

Among the structural ceramics silicon nitride is characterized as a highly reliable material, thanks to rod-like β-Si3N4 grains developed in the microstructure, which play a role of reinforcement to improve fracture toughness. High fracture toughness values of 8–11 MPa·m¹/² could be achieved in silicon nitrides containing larger elongated β-Si3N4 grains by sintering at high temperatures (>2000 °C) under high nitrogen pressure (>10 MPa) [1–2], the so-called self-reinforced or in situ-reinforced silicon nitrides. Unfortunately, fracture strength of these materials was rather lower compared to materials with uniform microstructure owing to the existence of huge grains acting as flaws. With the aim to achieve a better combination of strength and toughness, fundamental research works on microstructure evolution of Si3N4 had been conducted by some research groups [3–6]. In general, α-Si3N4 powder is used as a starting raw powder since α- to β-phase transformation during liquid-phase sintering leads to the development of elongated β-Si3N4 grains on account of preferential growth rate in the [001] direction of β-Si3N4. Systematic investigations conducted by the research group at Max-Planck Institute revealed that the newly formed β-Si3N4 phase in the course of phase transformation is grown on the surface of preexisting β-particles [3,4]. The number and size of β-Si3N4 particles existing in the starting raw powder, therefore, play an important role in determining the final microstructure.

Based on this grain growth phenomenon, Hirao et al. have developed a processing strategy for precisely controlling the microstructure of silicon nitrides [7–9]. The important point of the process is to control the size and distribution of elongated β-Si3N4 grains via seeding of morphologically regulated β-Si3N4 single-crystal particles. Additionally, the seeding method has the advantages of yielding textured self-reinforced microstructure as demonstrated in Al2O3 by Seabaugh et al. [10].

The seed particles were synthesized by heating powder mixtures of α-Si3N4, Y2O3, and SiO2, followed by acid rinse treatments [11]. In this process, particle size and aspect ratio of synthesized β-Si3N4 particles were controlled by the size of raw α-Si3N4 powder and Y2O3/SiO2 ratio of added oxides, respectively. Synthesized seed consists of rod-like particles with a uniform size distribution as shown in Figure 6.1. Characteristics of synthesized seed particles are summarized in Table 6.1.

![FIGURE 6.1 Method of the synthesis for β-Si3N4 seed particles.](image-url)
The variations of microstructure with the seed size and seed amount for the seeded and cold-pressed specimens are shown in Figure 6.2 [9]. The nonseeded specimen exhibited a relatively uniform microstructure (Figure 6.2-(a)). On the contrary, all of the seeded specimens exhibited a bimodal microstructure where large elongated grains are distributed in a smaller grained matrix (Figure 6.2-(b),-(c), and-(d)). The large elongated grains exhibited a core/rim structure, which represents direct evidence of grain growth from seed particles [7]. Smaller grains should be originated from β-Si₃N₄ particles present in the silicon nitride raw powder. It is clear that the size of larger grains increases with the increase in seed size; on the other hand, the number of larger grains per unit area decreases with the increase in seed size because of the decrease in the number of added seeds in a unit volume. It should be noticed that in cases where amount of the seed is small, large grains are highly dispersed; however, the addition of a larger amount of seed particles with large diameter causes coalescence of the newly developed grains as seen in the specimen with 5 vol% L seed (Figure 6.2-(d)).

Silicon nitrides with textured microstructure were fabricated by the combination of seeding and forming processes, generating shear stress, i.e., tape casting [12,13] and extrusion [14]. Figure 6.3 shows a typical microstructure

![Figure 6.2](image1.jpg) ![Figure 6.3](image2.jpg)

**TABLE 6.1 Characteristics of Synthesized Seed Particles**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Processing parameter</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BET of (\alpha)-Si₃N₄ powder (m²/g)</td>
<td>Y₂O₃:SiO₂ molar ratio (µm)</td>
</tr>
<tr>
<td></td>
<td>Heating temp. (°C)</td>
<td>1:2</td>
</tr>
<tr>
<td>S</td>
<td>10</td>
<td>1800 °C</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>1850 °C</td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>1850 °C</td>
</tr>
</tbody>
</table>

**FIGURE 6.2** Scanning electron micrographs of silicon nitride specimens (a) without seed, with (b) 5 vol% S, (c) 2vol% L, and (d) 5 vol% L seed particles.

**FIGURE 6.3** Scanning electron micrographs of seeded and tape-cast specimen in the casting plane.
of the seeded and tape-cast specimen. The specimen exhibited an anisotropic microstructure where large elongated grains, almost lying within the casting plane, were preferentially oriented parallel to the casting direction. Moreover, a combination of seeding and extrusion processes enabled the development of a highly anisotropic microstructure where elongated grains were almost unidirectionally oriented as shown in Figure 6.4. Reflecting the high orientation degree of elongated grains, their grain shape that appeared in the plane perpendicular to the extruding direction was quite equiaxial (Figure 6.4-b). Achievement of such a highly oriented microstructure is ascribed to the higher share stress during the extrusion process. It should be noticed that small matrix grains were almost consumed by selective grain growth of larger grains generated from seed particles. Consequently, large elongated grains with high aspect ratio were dominant in the microstructure of this material.

Figure 6.5 summarizes the mechanical properties for the seeded silicon nitrides along with the nonseeded specimen (indicated by a triangle). It is clearly seen that the addition of large seeds, which led to the microstructure with a bimodal grain-diameter distribution as indicated in Figures 6.2-c and 6.2-d, is effective for improving fracture toughness ($K_{IC}$). It should be noticed that the addition of large amounts of seed particles with larger diameter caused a sudden drop in strength. This is owing to the coalescence of large elongated grains which resulted in an increase in flaw size. It could be possible to improve $K_{IC}$ retaining a high strength level of about 1 GPa by controlling the amount and size of seeds. Further increase in $K_{IC}$ with a moderate increase in bending strength could be achieved in seeded and tape-cast silicon nitrides. The high fracture toughness of these specimens could be attributed to primarily the alignment of larger elongated grains normal to the crack plane which increases the opportunities for crack deflection and crack bridging. Moreover, the seeded and extruded specimens exhibited both high strength over 1.4 GPa and high $K_{IC}$ of about 14 MPa⋅m$^{1/2}$ in the direction parallel to grain alignment as a result of a higher orientation degree of large elongated grains (as indicated by an asterisk in Figure 6.5).

In addition, the textured silicon nitride had a high Weibull modulus of 46 combined with a high strength of around 1 GPa as shown in Figure 6.6 [12]. This value is about 3 times as high as those of conventional ceramics, and comparable to that of cast iron. One possible reason for the high Weibull modulus is the narrow size distribution of large elongated grains, which are grown from seeds with...
uniform size. Ohji et al. characterized the R-curve behavior for the textured silicon nitride by the Vickers indentation flaw technique in the crack length from 30 to 500 μm [15]. The material exhibits a high toughness from the beginning of the measured crack length with little increase in the following range. On the other hand, in the case of conventionally toughened silicon nitride fabricated by gas-pressure sintering (GPS), propagation of cracks attaining a length of several hundred micrometers was required to obtain a high fracture toughness. The benefit of such a steep R-curve is that the strength distribution can be narrower, thereby resulting in a higher Weibull modulus.

REFERENCES


Development of Fine-Grained Silicon Nitride Ceramics with a Small Amount of Sintering Additive

The conventional densification of Si3N4 ceramics is generally achieved by liquid-phase sintering with metal-oxide additives. The grain boundary phase derived from the sintering additives lowers the mechanical properties at high temperature, and the corrosion and oxidation resistances. To improve these properties, we need to minimize the amount of the additives used for the production of dense Si3N4 ceramics. Spark plasma sintering (SPS) technique is one where the entire sintering process can be completed in a short time, resulting in inhibition of grain growth during sintering. The use of nanosized amorphous Si3N4 powder is expected to produce fine-grained Si3N4 ceramics that reduce the amount of sintering additives for the densification [1,2]. Nanosized amorphous Si3N4 powder (average particle size: 80 nm, total oxygen content: 4.8 mass%) was synthesized by a vapor-phase reaction from SiCl4 and NH3 gases. As a reference, submicrometer-sized α-Si3N4 powder (β-phase: <5%, average particle size: 170 nm, total

FIGURE 6.6 Weibull plots of the bending strength for the seeded and tape-cast specimen.
oxygen content: 1.2 mass%) was also used as a starting material. The Si3N4 powders were sintered with a relatively small amount of sintering additives, 1.5 mass% Y2O3 and 0.5 mass% Al2O3, at a pressure of 30 MPa under N2 by the SPS technique.

In the specimens fabricated from nanosized amorphous Si3N4 powder at 1500°C to 1700°C for 30 min at a heating rate of 150°C/min, Si2N2O was detected as a secondary phase by XRD measurement. The formation of Si2N2O may be caused by a reaction of Si3N4 with the surface oxide in amorphous Si3N4 particles. Figure 7.1 demonstrates the effect of sintering temperature on the relative density and α-phase content of samples sintered at 1500°C to 1700°C for 30 min and at 1800°C for 1 min by using nanosized amorphous Si3N4 and submicrometer-sized α-Si3N4 powders. In the submicrometer-sized α-Si3N4 powder, the relative density of the sample obtained at 1500°C was about 60%. As the temperature increased to 1700°C, the density increased to 93%. With sintering at 1800°C for 1 min, the sample had a high density of >98%. The transformation from α- to β-phase occurred with an increase in the density of the sample. At 1800°C, only the β-phase formed. In the nanosized amorphous Si3N4 powder, only β-phase formed even at 1500°C, indicating that the amorphous phase directly transformed into the crystalline β-phase. The density of the sample reached 95% even at 1500°C, while the densification of the submicrometer-sized α-Si3N4 powder required a high temperature of 1800°C. This means that nanosized amorphous Si3N4 powder was fully densified at a lower temperature than the submicrometer-sized α-Si3N4 powder with a relatively small amount of sintering additive. The high sinterability of nanosized amorphous Si3N4 powder might be due to a smaller particle size and a larger amount of total oxygen content of the nanosized amorphous Si3N4 powder. The phase transformation from α- to β-Si3N4 is mainly dependent on the densification.

Therefore, because the densification of nanosized amorphous Si3N4 powder was achieved at a lower temperature than that of the submicrometer-sized α-Si3N4 powder, the transformation to β-phase might be completed at a lower temperature. Figure 7.2 shows the SEM micrographs of the surfaces of the Si3N4 ceramics fabricated at 1700°C for 30 min at 150°C/min using nanosized amorphous Si3N4 (a) and submicrometer-sized α-Si3N4 powders (b).
The fully dense $\beta$-$\text{Si}_3\text{N}_4$ ceramics consisted of the nearly equiaxed grains with an average size of 300 nm.

Development of Advanced $\alpha$-$\text{SiAlON}$ Ceramics

$\alpha$-$\text{SiAlON}$ is a solid solution of $\alpha$-$\text{Si}_3\text{N}_4$ crystal structure in which some Si and N are replaced by Al and O, and some metal cations are incorporated as a stabilizer. The general formula for $\alpha$-$\text{SiAlON}$ is $M_x\text{Si}_{12-(m+n)}\text{Al}_{m+n}\text{O}_n\text{N}_{16-n}$, where M may be Li, Mg, Ca, Y, or most rare-earth elements. Among $\alpha$-$\text{SiAlON}$s, Ca-$\alpha$-$\text{SiAlON}$ has received considerable attention because of inexpensive raw material of calcium compound, high cation (Ca) solubility in the $\alpha$-$\text{SiAlON}$ lattice, thus low residual glass phase, and excellent thermal stability [3]. $\alpha$-$\text{SiAlON}$ has higher hardness than $\beta$-$\text{SiAlON}$ and is expected to have good performance under mechanical wear and erosion. Ca-$\alpha$-$\text{SiAlON}$ ceramics have been generally produced by sintering a powder mixture of $\text{Si}_3\text{N}_4$, AlN, and CaCO$_3$ [4]. To improve the reliability of structural ceramics such as $\text{SiAlON}$, it is very important to develop high-quality powders.

8.1. SYNTHESIS OF SPHERICAL HOLLOW $\text{SiAlON}$ POWDERS COMPOSED OF NANOSIZED PARTICLES

Ca-$\alpha$-$\text{SiAlON}$ powder was synthesized by carbothermal reduction–nitridation (CRN) of a $\text{SiO}_2$–$\text{Al}_2\text{O}_3$–CaCO$_3$ and carbon powder mixtures [5,6]. The proportions of the starting powders were calculated to give nominal composition of $\text{Ca}_x\text{Si}_{12-(m+n)}\text{Al}_{m+n}\text{O}_n\text{N}_{16-n}$ (i.e., $x = n = m/2$) in $M_x\text{Si}_{12-(m+n)}\text{Al}_{m+n}\text{O}_n\text{N}_{16-n}$ with $x$ varying from 0.3 to 1.4. The mixed powders were heated in N$_2$ gas flow using a horizontal electrical furnace. The residual carbon was removed by burning the synthesized powder at 700 °C in air. The crystalline phases appearing in the final products were Ca-$\alpha$-$\text{SiAlON}$, $\beta$-$\text{SiAlON}$, and AlN for all compositions after heating at 1450 °C for 120 min. The content of the phases depended on the starting composition. The maximum amount of $\alpha$-$\text{SiAlON}$, >85 wt%, occurred in the composition $x = 1.0$ with AlN as a minor phase. The synthesized Ca-$\alpha$-$\text{SiAlON}$ powders indicated a unique morphology of hollow spheres with 200–500 nm in diameter (Figure 8.1). This morphology of the spherical hollow powder was generally observed in compositions between $x = 0.5$ and $x = 1.2$. Further observation revealed that the Ca-$\alpha$-$\text{SiAlON}$ hollow spheres were composed of large numbers of nanosized particles of about 10–30 nm and submicron-sized particles of about 100 nm in diameter (Figure 8.2). The shells of the hollow spheres were around 50 nm thick; thus, these nanoparticles took a form with very high surface areas. Additionally, the bonding between the nanoparticles in the hollow spheres did not seem very strong because the hollow spheres could be broken down into nanoparticles by ultrasonic vibration.

The formation process of the Ca-$\alpha$-$\text{SiAlON}$ hollow spheres composed of nanosized particles during CRN process of the $\text{SiO}_2$–$\text{Al}_2\text{O}_3$–CaO system was investigated [7]. A few spherical particles <0.5 μm in diameter were generated at 1200 °C. The spherical particles could be caused by a Si–Al–Ca–O liquid phase that formed above the $\text{SiO}_2$–$\text{Al}_2\text{O}_3$–CaO eutectic temperature of 1170 °C. The number and size of the spherical particles increased when further increasing the heating temperature. From

![FIGURE 8.1](image1.png)

**FIGURE 8.1** SEM micrograph of the Ca-$\alpha$-$\text{SiAlON}$ powder prepared at 1450 °C for 120 min in composition $x = 1.0$.

![FIGURE 8.2](image2.png)

**FIGURE 8.2** TEM micrograph of the Ca-$\alpha$-$\text{SiAlON}$ powder prepared at 1450 °C for 120 min in composition $x = 1.0$. 
TEM observations in the samples prepared at 1450 °C for 0–120 min, it was clear that the solid spheres gradually changed into hollow spheres with increasing soaking time, while maintaining their spherical shape and size. XRD result showed that the morphological change from solid to hollow spheres and the formation of Ca-α SiAlON took place simultaneously. Moreover, in the spherical solid particles obtained at 1450 °C after 0 min, it was found that very fine particles were formed on the surface of the solid spheres at the initial stage of CRN. These fine particles may play a role of nuclei from which nanosized Ca-α SiAlON particles can subsequently grow. The solid spheres obtained at 1450 °C after 0 min had a nearly uniform composition. Electron diffraction analysis of the solid spheres indicated only the amorphous halo, confirming that the solid spheres at the initial stage of CRN are largely a Si-Al-Ca-O amorphous phase. An increase in density and a decrease in mass are the results of the conversion from the amorphous to crystalline α-SiAlON, which are contributed to the formation of Ca-α SiAlON hollow spheres composed of nanosized particles.

The preparation of Ca-α SiAlON powders was also performed using SiO₂ starting powders with different characteristics in particle size, shape, and crystalline state [8]. In all systems, the Ca-α SiAlON hollow spheres were always formed through an intermediate Si-Al-Ca-O liquid phase in the same formation process, regardless of the characteristics of SiO₂ powders used. Furthermore, Ca-α SiAlON powders were prepared from SiO₂-Al₂O₃-CaO glass, in order to obtain the evidence to support the assumption that the Ca-α SiAlON hollow spheres derived from the liquid phase [9]. It was revealed that the resultant particle morphology was similar to that of Ca-α SiAlON obtained from CRN of SiO₂-Al₂O₃-CaCO₃ powder mixtures.

### 8.2. FABRICATION OF DENSE SiAlON NANOCERAMICS

The synthesized Ca-α SiAlON hollow spheres composed of nanosized particles were ground to obtain nanosized Ca-α SiAlON starting powder. The nanosized powder was then sintered without sintering additive by spark plasma sintering (SPS) at 1700 °C for 20 min in N₂ atmosphere under a uniaxial pressure of 30 MPa at a heating rate of 100 °C/min [10]. The phase contents in the sample sintered by SPS were almost the same as those in the starting powder. The sintered body achieved full density of 99.2% in SPS, while significant grain growth and no densification occurred in gas-pressure sintering (GPS) of the nanosized Ca-α SiAlON powder. Moreover, crystallite sizes of the starting powder and the sintered body calculated using Scherer’s equation were 27 and 34 nm, respectively. The sample had a bimodal microstructure composed of very fine grains with <50 nm and submicron-sized grains with 100–200 nm in diameter (Figure 8.3). The grain sizes were almost equal to that of the Ca-α SiAlON starting powder, suggesting that significant grain growth did not occur during the SPS process. Furthermore, no residual glass phase was observed along the grain boundary in the Ca-α SiAlON nanoceramics by high-resolution TEM investigation. Table 8.1 demonstrates the mechanical properties of the Ca-α SiAlON nanoceramics. Vickers hardness was approximately 18 GPa, which was higher than that of Si₃N₄ ceramics fabricated by GPS. Fracture toughness was slightly lower than that of Si₃N₄ ceramics, about 3.2 MPa·m¹/₂. The average value of the flexural strength at room temperature was 680 MPa. The strength at a high temperature of 1200 °C had a high value of 700 MPa, which was almost the same as that at room temperature. Corrosion resistance of the Ca-α SiAlON nanoceramics at H₂SO₄ solution was better than that of commercial Si₃N₄ ceramics for bearing grade. Because the commercial Si₃N₄ ceramics are produced with sintering additives such as

<table>
<thead>
<tr>
<th>TABLE 8.1 Mechanical Properties of the Ca-α SiAlON Nanoceramics Prepared at 1700 °C for 20 min by SPS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vickers hardness (GPa)</td>
</tr>
<tr>
<td>Fracture Toughness (MPa·m¹/₂)</td>
</tr>
<tr>
<td>Flexural strength (MPa)</td>
</tr>
<tr>
<td>Room temperature, air Avg 680 (Min. 580, Max. 969)</td>
</tr>
<tr>
<td>1200 °C, 10 min, air Avg 680 (Min. 466, Max. 928)</td>
</tr>
</tbody>
</table>
Y$_2$O$_3$ and Al$_2$O$_3$, residual glass phase in the Si$_3$N$_4$ ceramics is easily corroded. The excellent strength at high temperature and good corrosion resistance of the Ca-\textit{z} SiAlON nanoceramics would be caused by no glassy phase at grain boundary. Consequently, it was revealed that the Ca-\textit{z} SiAlON nanoceramics fabricated by the SPS technique using the nanosized Ca-\textit{z} SiAlON powder had high hardness, excellent strength, and good corrosion resistance.

8.3. FABRICATION OF POROUS SiAlON CERAMICS

Porous Ca-\textit{z} SiAlON ceramics were fabricated by using the Ca-\textit{z} SiAlON spherical hollow particles and maintaining their unique shape [11]. The spherical hollow powder was sintered without sintering additive by SPS at 1700 °C for 1−20 min in N$_2$ under a uniaxial pressure of 5 MPa at a heating rate of 100 °C/min. The open and closed porosities of the specimen prepared at 1700 °C for 1 min by SPS were 44% and 2%, respectively. The open porosity decreased with increasing soaking time. In all specimens, the phase contents were almost the same as the starting powder. Figure 8.4 shows the SEM micrograph of the fracture surface of the porous Ca-\textit{z} SiAlON ceramics prepared at 1700 °C for 1 min by SPS. The hollow spheres with morphology similar to the Ca-\textit{z} SiAlON starting powder were observed in the specimens prepared for 1 min. With increasing soaking time, the hollow spheres gradually disappeared and total porosity decreased. The pore-size distribution measurements confirmed that the porous Ca-\textit{z} SiAlON ceramics had a large number of pores with 200−500 nm in diameter and a small number of pores with about 10 nm in diameter. This indicates that the morphology of the hollow spheres in the starting powder was maintained in the porous Ca-\textit{z} SiAlON ceramics with high porosity fabricated by the SPS technique. Figure 8.5 shows the effect of the porosity on the flexural strength of the porous Ca-\textit{z} SiAlON ceramics. The strength decreased with increasing porosity in the porous ceramics. The strength of the porous Ca-\textit{z} SiAlON ceramics was higher than that of typical porous ceramics such as Al$_2$O$_3$ reported by Coble [12]. Vickers hardness of the porous Ca-\textit{z} SiAlON ceramics decreased with increasing porosity, as well as the flexural strength. The hardness was higher than that of dense mullite and cordierite ceramics and was almost the same as that of RBSN and porous Si$_3$N$_4$ ceramics prepared using starch. Corrosion resistance of the porous Ca-\textit{z} SiAlON ceramics at H$_2$SO$_4$ solution was better than that of commercial dense Si$_3$N$_4$ ceramics, although the weight loss of the porous Ca-\textit{z} SiAlON ceramics in H$_2$SO$_4$ solution was higher than that of the developed dense Ca-\textit{z} SiAlON nanoceramics because of its high surface area. The porous Ca-\textit{z} SiAlON ceramics indicated excellent corrosion resistance, as well as the dense Ca-\textit{z} SiAlON nanoceramics, because the Ca-\textit{z} SiAlON spherical hollow powder was sintered without sintering additive.

REFERENCES

Applications

Among the examples of practical applications of silicon nitride (Si₃N₄) ceramics, several representative examples of automotive and other applications are described below.

9.1. AUTOMOTIVE APPLICATIONS

9.1.1. Glow Plugs

Conventional glow plugs require a preheating time of 20–30 s before beginning to operate, and the durability decreases if the preheating time is shortened; the resistance level of the heater is reduced, and a strong electric current flows. As a result, the heater quickly reaches a high temperature. To address this problem, a control system for the electric current was added to protect the heater. Initially, these quick-heating glow plugs became widely used. However, the heating control system was expensive, and it was difficult to shorten the preheating time to less than 6 s because the heater was made of metal.

Ceramic-type glow plugs (Figure 9.1) for diesel engines have a ceramic heater element inside the heating section. The properties of silicon nitride make it an excellent material for these glow plugs. Silicon nitride has superior resistance to heat, corrosion, and thermal shock. It also has a higher strength at high temperatures than metals do. Therefore, it retains its strength when it is exposed to a high-temperature flame in a combustion chamber. Furthermore, corrosion due to oxidation does not occur under the intense oxidation—reduction atmosphere in combustion gases, so its durability is excellent. The use of ceramic glow plugs provides a superior heating profile and a high-temperature after-glow. A ceramic glow plug can be heated to 800 °C in 2 s, which shortens the preheating time considerably. In addition, the temperature can be maintained at 900 °C for a long time after starting (i.e., the plugs provide an after-glow system), which purifies engine emissions. With these improvements, diesel cars can now be started in the same manner as gasoline cars.

A ceramic glow plug using silicon nitride ceramics for the heater body was developed in 1981 by Isuzu Motors. The use of silicon nitride parts allowed the first successful mass production for automotive use. Ceramic glow plugs proved to be overwhelmingly superior to metal glow plugs and became firmly established in the market as auxiliary starter equipment for diesel engines. The latest glow plug achieves a high temperature of 1500 °C in the after-glow mode, making it an effective tool for white smoke reduction. The ceramic glow plug also increases the engine performance, and its durability greatly improves engine reliability. Multiple ceramic companies produce glow plugs that exploit the superior heat resistance, high-temperature strength, and corrosion resistance of silicon nitride. We manufacture them with thread diameters of 8 and 10 mm. The technology obtained by the development of the glow plug has found other heating applications in addition to those in the automotive industry.

9.1.2. Turbocharger Rotors

The heat resistance and low density of silicon nitride make it a desirable material for turbocharger rotors. Its use for this part improves engine response. Manufacturers of automobiles
and ceramics around the world pursued research and development of ceramic turbocharger rotors. In October 1985, the world’s first ceramic turbocharger car was put to practical use at Nissan Motors. The rotors are manufactured by gas-pressure burning using high-quality materials.

Ceramic turbine rotors (Figure 9.2) offer lower turbo lag than metal rotors, for the following reason. When immediate acceleration requires high engine output, there is a slight delay before a rotor reaches a high rotation speed. Replacing metal rotors with small ceramics of the specific gravity decreases the inertial force and reduces the delay. In addition, these rotors are exposed to a hot gas and must be durable under those conditions. In gas-pressure sintering, Si$_3$N$_4$ is sintered at a high temperature, and grain growth is controlled by using a high nitrogen gas pressure. Therefore, the material contains few residual pores due to fine-grained nature of the silicon nitride. A high-density sintered body can be obtained with lower levels of sintering additives than typical pressureless sintering bodies by controlling the high-temperature behavior.

Lightweight exhaust valves made of silicon nitride offer superior flutter characteristics in high-speed turns, and a mileage improvement due to reduced pumping loss was expected when they were introduced. Relatively expensive heat-resistant metals are commonly used in exhaust valves, and the adoption of silicon nitride was initially regarded optimistically. However, it was one step now at a point of the cost competitiveness, and silicon nitride did not achieve practical use.

9.2. INDUSTRIAL APPLICATIONS

9.2.1. Bearings

Silicon nitride ceramic has high strength and fracture toughness. In addition, it is light and electrically insulating and offers low thermal expansion characteristics, high elasticity, and high heat conductance. These characteristics make it more attractive than bearing steel. Silicon nitride bearings (Figure 9.3) have a lower coefficient of friction and high rigidity. Thus, silicon nitride is now used in applications where heat resistance is required for super-high-speed, super-precise bearings.

9.2.2. Cutting Tool Tips

Ceramic tools maintain high strength, hardness, and a high resistance to plastic deformation even at temperatures above 1000 °C. These properties are attractive for high-speed processing in which the cutting temperature is so high that cutting by hard metal, cermet, or coated tools is impossible. The representative material applied to cutting tools is silicon nitride ceramics (Figure 9.4). In its microstructure, columnar grains intertwine with each other, and fiber reinforcement occurs. The smaller the grain size and the larger the aspect ratio (particle length/particle size), the stronger is the material. This material has high fracture
toughness, high-temperature hardness, a low coefficient of thermal expansion, and superior heat resistance. However, it reacts easily with steel at high temperatures, so it is not suitable for processing steel. Therefore, silicon nitride cutting tools are used for high-speed cutting by rough fraise processing and intermittent grinding of cast iron.

Alumina-based ceramic tools with TiAlN films applied by plasma vapor deposition (PVD) can provide superior performance. Moreover, Al₂O₃ films applied by chemical vapor deposition (CVD) can offer better stability than commonly used TiCN films, resulting in superior wear characteristics in ceramic tools. However, silicon nitride ceramics and coated silicon nitride ceramics are the most important ceramic cutting tools.

9.2.3. Metal-Casting Tools

Aluminum engine parts are manufactured by gravity/low-pressure die casting. The parameters for processes such as die casting depend on the size, shape, and desired quality of the parts. In the aluminum metal casting industry, improvements in product quality and stability and in production efficiency are important, and the extension of useful tool life obtained by using silicon nitride tools is also desirable.

Refractory metals have long been used as materials for casting furnaces. However, silicon nitride ceramics, which are more resistant to heat and corrosion, are currently used as protection pipes, ladles, stalks, and heater tubes in aluminum casting.

Silicon-nitride-based materials might be useful for fabricating tools used in the melting and holding steps of metal casting, including ladles and trough linings; thermocouple protection sheaths; more efficient and durable burners; heat exchangers for reducing fuel consumption; in-line composition sensors; tundish nozzles; tap-hole linings; and gate valves.

Figure 9.5 shows the appearance of silicon nitride parts for various practical casting applications. Many of them are more than 1 m long. The quality of the final product is stabilized by continuous temperature measurement. The ladle and stalk make it possible to use the thermal energy in metal casting process by reducing heat retention due to their low heat conductance. The heater tube enables heating from the bottom part of the smelting furnace.

The thermocouple protection tube is the most widely used example. Temperature control is critical in melting and holding furnaces. Temperature is typically measured by inserting a thermocouple protected by a ceramic sheath consisting of a closed-end tube, which has a relatively short life. A thermocouple protection tube must be stable enough in the molten metal to allow continuous submersion and continuous temperature measurement. The primary concern is catastrophic fracture due to thermal shock or mechanical impact. A piece of a broken tube could obstruct the tap hole and cause an expensive furnace shutdown. Silicon nitride tubes have shown promise in some nonferrous metals. Silicon nitride has superior thermal shock resistance and high strength at temperature, making it worthy of consideration. Closed-end Si₃N₄ tubes are available commercially. Another alternative is a ceramic matrix composite, which may be less susceptible to catastrophic fracture than monolithic ceramics. Although closed-end ceramic matrix composite tubes have been fabricated, they are expensive and are currently not in production.
Recent Progress in Zr(Hf)B₂ Based Ultrahigh Temperature Ceramics

Dipankar Ghosh* and Ghatu Subhash

*Materials Science and Engineering, University of Florida, Gainesville, Florida 32607, United States, †Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611, United States

1. INTRODUCTION

Novel materials which are lightweight retain high strength at elevated temperatures (above 2000 °C) and provide good thermal insulation properties that are crucial to the development of advanced components for civilian, defense, and aerospace applications [1,2]. Thermal protection systems that operate in the ultrahigh temperature regime (considered to begin at 1600 °C) and in a highly oxidizing environment are of immediate need for applications in hypersonic and re-entry vehicles as external shields. The design of high-performance hypersonic vehicles generally involves relatively sharp nose tips and wing leading edges (WLEs) because they help reduce the vehicle’s drag and enhance maneuverability and performance [3]. Due to such extreme demands, development of a new class of ultrahigh temperature materials (UHTMs) that can sustain the severe environments experienced by hypersonic flight, scramjet propulsion, rocket propulsion, and atmospheric re-entry vehicles is an area of renewed emphasis and interest. The daunting task of developing new UHTMs operating in hypersonic environment has often been considered a “quest for unobtainium” [4].

The structure or skin of many high-performance hypersonic vehicles must endure an operating environment with temperatures in excess of 1600 °C, oxidizing and reducing corrosive gases, and dynamic pressures in the range of 72,000 Pa. However, the engine of the hypersonic vehicle will experience the most demanding environment of the structure with temperatures at the combustor exit in excess of 2500 °C. Active cooling can be used to lower temperatures significantly. However, this creates new engineering challenges since the cooling fuel can degrade the UHTC mechanical properties and decrease overall system performance.

One of the challenges encountered in the design of sharp WLEs is that the convective heating to the surface, and hence the surface temperature, increases as the WLE radius decreases. As a result, the surface temperature on the stagnation region of such sharp leading edges can potentially exceed 2000 °C [3]. Few materials are currently available which can withstand such high temperatures and reactive environments while retaining their dimensional and structural integrity and providing adequate mechanical strength. The thermal conductivity of the materials is also very crucial for these applications, because the convective energy that enters the surface near the stagnation region is conducted away to the cooler regions of the leading edge, where it can be radiated back to the environment (Figure 1). Thus, the higher the thermal conductivity of the leading-edge material, the more efficient this process becomes [3].
Ultrahigh temperature materials have been grouped as a special class of compounds that have extremely high melting temperatures (>3000 °C) and are chemically and physically stable at temperatures above 2400 °C in reactive atmospheres (e.g., monatomic oxygen). Due to their thermo-chemical stability, UHTMs are capable of retaining their strength at high temperatures and exhibit superior oxidation resistance. Such characteristics allow these materials to be used in extreme environments including those associated with components (e.g., nozzles, propulsion components, leading edge materials for space vehicles, and thermal protection systems) for future high-performance aircrafts, hypersonic vehicles, kinetic energy interceptors, and reusable space planes. These vehicles operate in severe reactive environments with temperatures well above the melting points of traditional metals.

Ultrahigh temperature ceramics (UHTCs) constitute a special subclass in the family of UHTMs. The UHTCs mainly include borides, carbides, and nitrides of early transition metals such as ZrB₂, HfB₂, TaC, ZrC, HfC, and HfN as well as their composites, which have melting temperatures above 3000 °C. Utilization of these materials for the aforementioned applications presents a variety of unique scientific challenges associated with processing, characterization, and performance evaluation. Thus, in recent years, investigations on UHTCs have centered on the basic understanding of processing and thermo-mechanical properties. Investigations have been conducted to determine desirable processing conditions to produce controlled and defect-free microstructure, improve the understanding of microstructure evolution as a function of processing variables, provide insight into the mechanisms by which materials respond to mechanical loads and high-temperature reactive environments, and promote the development of new testing methodologies that may be unique to the UHTM community as well as appropriate for applications [5–8]. Since these materials have been recommended for high-risk and high-payoff applications, an improved scientific knowledge should enable revolutionary improvements in the design of aerospace vehicles (e.g., hypersonic flight vehicles) and the capabilities of propulsion systems (e.g., scramjet technology) that are not possible with other materials [9].

1.1. Available Ultrahigh Temperature Materials

Currently, there is a limited availability of high temperature materials in the UHTM family (see Table 1). This is due to the fact that even though there is a large pool of materials that have extremely high melting temperatures, insufficient high-temperature strength retention capability, poor thermal and chemical stability (oxidation resistance), and poor thermal conductivity preclude several of them from being included in the UHTM category [10].

Graphite-based materials are widely used as high temperature structural materials because of their dimensional stability, low density, strength retention at high temperature, ease of machinability, and low cost. However, they degrade rapidly in oxidizing environments, especially when the temperature is above 800 °C [1].

<table>
<thead>
<tr>
<th>Material</th>
<th>Melting point (°C)</th>
<th>Material</th>
<th>Melting point (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphite</td>
<td>3427 (at 2.5 bar) [12]</td>
<td>TaN</td>
<td>3093 [13]</td>
</tr>
<tr>
<td>Re</td>
<td>3180 [15]</td>
<td>ZrB₂</td>
<td>3250 [10]</td>
</tr>
<tr>
<td>TaB₂</td>
<td>3150 [11]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Coatings that can protect carbon-based materials from oxidation are an attractive option for thermal protection, but thermal expansion differences between the coatings and the substrates coupled with the possibility for high-temperature reactions make this an impractical solution for many UHT applications. SiC-based ceramics and composites also suffer from the similar problem of poor oxidation resistance at elevated temperatures [1]. Although ceramic oxides are inherently stable in oxidizing atmosphere and easy to process, they suffer from a lack of thermal shock resistance, low fracture toughness, and poor creep resistance at elevated temperatures. Refractory metals such as Re, W, and Mo have high densities and thus are not suitable for lightweight aerospace applications.

1.2. Promising Ultrahigh Temperature Ceramics

In late 1960’s, an HfB$_2$–20% SiC composite among several diborides was identified as a promising candidate that would satisfy the initial requirements of sustained operation at elevated temperatures in an oxidizing environment. However, processing was limited to conventional pressureless sintering and hot-pressing techniques that limited the manufacturability of the components. Over the past three to four decades, research in the discovery of new classes of materials and development of processing methods for UHTCs has been sporadic. As a result, there is a scarcity of fundamental research data available on the mechanical performance and manufacturability of UHTCs [4].

From early 2000, diborides and carbides of early transition metals (Zr, Hf, Ta) have emerged as the most promising candidates owing to their superior thermal and chemical stability for use in extreme environments associated with hypersonic flight (1400 °C and above in air), atmospheric re-entry vehicles (2000 °C and above in monatomic O and N), and rocket propulsion (3000 °C and above in reactive chemical vapors) [9]. These materials have melting temperatures above 3000 °C, retain their strength at temperatures above 1200 °C, and exhibit good thermal shock resistance. Borides and carbides also exhibit excellent erosion resistance under the extreme heat flux and gas velocity conditions encountered during the operation of hypersonic and re-entry vehicles [1,2]. The oxidation resistance in these materials can be enhanced significantly with the addition of SiC [10,16–18]. Thus, in the family of UHTCs, composites such as ZrB$_2$–SiC and HfB$_2$–SiC have received enormous attention [10,16–18], owing to their superior oxidation resistance above 1500 °C. Nitrides of early transition metals are also promising UHTCs due to their high thermal conductivity, thermal shock resistance, and strength [19].

1.3. Scope of Fundamental Investigations

Although there are many promising UHTCs for the aforementioned applications; they have not been fully tested. Thus, the basic investigations associated with processing and mechanical and thermal characterization that can be correlated to their performance in service are limited in open literature. Some of the issues that need to be addressed are as follows [9]:

1. What is the rationale for selection of suitable materials and compositions for the intended ultrahigh temperature applications? An important aspect of selecting the appropriate material for an application can be the identification of the key properties needed to meet the desired performance specifications.

2. How do composition and structure affect the key properties of UHTCs? How are such properties controlled through processing and microstructure refinement (grain size, additives, second phases, etc.)? The processing conditions must be identified to optimize the microstructure.

3. What are the intrinsic properties of UHTCs? Information is needed on the lattice parameters, bonding energy, thermal expansion coefficients, heat capacities, and elastic constants of compounds such as ZrB$_2$ and HfB$_2$.

4. How do UHTCs interact with other materials when operating in extreme reactive environments? Further research is needed on phase equilibrium studies to generate comprehensive thermodynamic data and phase diagrams.

5. SiC is known to enhance the properties of Zr(Hf)B$_2$ ceramics, but what other additives can be introduced in UHTC systems to improve the strength, fracture toughness, and chemical stability of UHTMs? Further investigation of this question will broaden the available spectrum of UHTMs.

6. What are the micro-mechanisms of deformation and fracture in UHTCs, both at room and elevated temperatures? How do such failure characteristics change with the introduction of secondary phases? Compounds should be examined for any possible phase transformations and/or ductile-to-brittle transitions that may affect the performance of these materials at elevated temperatures.

The present literature on UHTCs is still not mature enough to answer the above fundamental questions. Research in areas such as processing, oxidation behavior, and mechanical characterization of UHTCs is still limited. In this book chapter, a judicious effort has been undertaken to summarize many of the above aspects of fundamental and application-oriented research available in open literature on UHTCs.
2. PROCESSING OF ZR(HF)B₂-BASED UHTCS AND THEIR COMPOSITES

Densification of pure diborides, carbides, and nitrides is difficult due to their high degree of covalency in chemical bonding and low diffusion rates [11,20]. Sintering additives are commonly added to enhance densification of starting powders. Consolidation of UHTC powders has been commonly achieved by both conventional and nonconventional sintering techniques. The conventional sintering methods include: (i) pressureless sintering (PS), (ii) hot-pressing (HP), and (iii) reactive densification processes such as reactive HP. Also, nonconventional sintering methods such as spark plasma sintering (SPS) [21–26] and plasma pressure compaction (P²C) have been recently employed to produce dense UHTC compacts and their composites [21,25–31].

From the early 1950s to the 1970s, considerable effort was directed toward densification of commercially available, nonreactive diboride powders [10]. While hot-pressing (HP) was successfully utilized to produce dense diboride ceramics, its limited ability to produce large size ceramics of complex shapes forced researchers to investigate other processing schemes. Several metallic additives and liquid phase sintering aids have been found to be beneficial in the densification of these ceramic powders. Significant progress has been made toward understanding the densification mechanisms that can result in dense UHTC compacts at moderate processing conditions while minimizing second phases that may be detrimental to their high temperature properties.

In the ensuing sections, discussions will mainly emphasize the densification behavior of Zr and Hf diboride ceramics and their composites. We will discuss the difficulties associated with the sintering of these composites, and summarize the specific sintering techniques, processing variables, and sintering additives that allow moderate to full densification. Two major factors that limit full densification of these powders are: (i) low bulk and grain boundary diffusion rates and (ii) the presence of surface oxides such as ZrO₂ (for ZrB₂), HfO₂ (for HfB₂), and B₂O₃ on starting powders [32]. These surface oxides provide rapid diffusion paths along the surfaces of the powder particles and promote coarsening without densification through the evaporation–condensation mechanism at intermediate temperatures (1500–1800 °C) [33,34]. In addition, grain boundary and volume diffusion rates are low in the above temperature range and do not become appreciable until the sintering temperature reaches 1800 °C or higher [35]. Also, the hexagonal crystal structure of diborides allows anisotropic grain growth with entrapped porosity leading to coarsening rather than promoting densification. Typically, these structures require hot-pressing at 2100–2300 °C [36] to facilitate sintering. To mitigate these problems, suitable sintering additives and appropriate sintering schedules have been successfully adopted to significantly improve the densification process via pressureless sintering and reactive densification processes (e.g., reactive HP) at lower temperatures and shorter processing times.

2.1. Pressureless Sintering

Pressureless sintering (PS) technique is advantageous in producing near-net shape components, and it is cost effective, as well. However, due to the lack of any external pressure during sintering, a high sintering temperature is often required in this method to obtain reasonable densification. To compensate for this limitation, either the starting particle size is reduced or sintering additives are used. Potential sintering aids can be solid solution formers, reactive agents, or liquid phase formers [10].

Sintering of diborides has been studied for more than 40 years, and it has been shown that the densification rate can be appreciably enhanced through additions of transition and/or refractory metals (Fe, Cr [37], Ni [38]), refractory metal silicides (MoSi₂ [39,40], TiSi₂ [41] and HfSi₂ [42]), and ceramic additives (SiC, Si₃N₄ [33]). For example, fully dense ZrB₂ was produced at 2200 °C for 1 h in argon atmosphere in the presence of metals like Re (1 wt%) or a combination of Cr and Ti (0.5 wt% each) [43]. Such additions caused substitution of metal atoms into Zr sites in the ZrB₂ lattice and altered the local contraction of the lattice. This affects the surface and the volume free energies and consequently increases the driving force for sintering [43–45]. These metals can also form a liquid phase during sintering, which fills the voids between the matrix grains. This liquid phase along the grain boundaries enhances the mass transport, which accelerates the densification process [46]. In another study of pressureless sintering, ZrB₂ was sintered to 95% relative density using nonmetallic additives [47,48]. However, these studies involved the addition of significant amounts of BN (5 wt%), AlN (15 wt%), and SiC (5 wt%), which have been found to reduce melting temperature, hardness, and high temperature strength [38,43,47–49].

Another procedure that has been found to facilitate densification of ZrB₂ ceramics when using the PS technique is mechanical milling. This process not only increases the surface area of powder particles but also increases the number of point defects near the surface of the particles, thus enhancing the driving force for densification by grain boundary diffusion [50]. However, milling also increases the oxygen content due to an increase in surface area [51], and therefore it should be removed by the addition of proper reactants.

In recent literature, ZrB₂ ceramics have been sintered to almost full density using additives such as MoSi₂ [52], WC [32,53], B₄C [53], and C [36]. However, in addition to
utilizing these sintering aids, the choice of a sintering schedule is an important consideration for achieving high densification of powders. Table 2 reveals a list of recent literature on the sintering of diboride ceramics such as Zr and Hf. Figure 2 shows a typical dense ZrB₂ (~100%) microstructure sintered at 1850 °C for 1 h with 4 wt% B₄C using PS [10].

2.1.1. Removal of Surface Oxides

As mentioned before, surface oxides on diboride ceramics inhibit densification and promote coarsening [10], and it is important to remove such oxides to achieve considerable densification. Boron oxide (B₂O₃) becomes a liquid above 450 °C and starts to volatilize above 1200 °C under vacuum (low partial pressure). Thus, long isothermal hold in the temperature range of 1200–1600 °C can remove most of the surface B₂O₃ before the onset of coarsening in ZrB₂, which is estimated to be around 1750 °C [63]. On the other hand, removal of ZrO₂ and HfO₂ surface oxides is difficult and requires chemical reactions. Additives such as WC, B₄C, and C play crucial roles in removing the surface oxides and improving the densification. In general, WC is

| TABLE 2 Pressureless Sintering Conditions of Diboride Ceramics and Composites |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Material                        | Particle size (μm) | Additives        | Temperature (°C) | Time (min) | Final density (%) | Reference |
| ZrB₂                            | 0.5 (attrition milled) | WC (from grinding media) | 2150 | 540 | ~98 | [32] |
| ZrB₂                            | 2 | 2 wt% B₄C + 1 wt% C | 1900 | – | ~100 | [51] |
| ZrB₂                            | 0.5 | 1 wt% B₄C + 0.5 wt% C | 1850 | – | ~100 | |
| ZrB₂                            | 0.1–8 | 1 wt% B₄C + 0.5 wt% C | 1850 | 30 | ~100 | [52] |
| ZrB₂–20 vol% SiC                | 1.4 (ZrB₂) | 20 vol% MoSi₂ | 1850 | 120 | ~100 | [54] |
| ZrB₂                            | 0.1 (attrition milled) | B₄C | 1850 | 60 | >98% | [53] |
| ZrB₂                            | 0.5 (attrition milled) | 1 wt% C | 1900 | 120 | ~100 | [55] |
| HfB₂                            | 1.5 | 2 wt% B₄C | 2200 | 120 | 98.3 | [46], [56] |
| HfB₂                            | 0.5–5 | 20 vol% MoSi₂ | 1950 | 60 | 98 | [40] |
| ZrB₂                            | – | 5 wt% TiC | 1800 | 30 | 94.2 | [57] |
| ZrB₂–20 vol% SiC                | 4 (ZrB₂) | 5 vol% VC | 2100 | 120 | ~100 | [58] |
| ZrB₂–20 wt% SiC                 | 0.12 (ZrB₂) | 4 wt% Mo + 1.5 wt% C | 2250 | 120 | ~97.7 | [59] |
| ZrB₂–10 vol% SiC                | 2 (ZrB₂) | 4 wt% B₄C + 7.3 wt% C | 1950 | 180 | 95.6 | [60] |
| ZrB₂–20 vol% SiC                | 2 (ZrB₂) | 4 wt% B₄C + 5 wt% C | 2000 | 180 | 97.1 |
| ZrB₂–30 vol% SiC                | 2 (ZrB₂) | 4 wt% B₄C + 2.8 wt% C | 2100 | 180 | 98 | [60] |
| ZrB₂–20 vol% SiC                | 14 (ZrB₂) | 10 vol% WC | 2100 | 120 | ~100 | [61] |
| ZrB₂–20 vol% SiC                | – | 4 vol% Si₃N₄ | 2150 | 120 | ~98 | [62] |
processing conditions for ZrB$_2$ were achieved when a mixture of B$_4$C and C was used, as opposed to using B$_4$C or C alone [32,53]. Similarly, dense HfB$_2$ (98.3%) ceramics were also produced in the presence of B$_4$C, which assists in densification by removing surface oxide (HfO$_2$), as shown in reaction (2) [56]. This reaction becomes favorable above 1400 °C:

$$7\text{HfO}_2 + 5\text{B}_4\text{C} \rightarrow 7\text{HfB}_2 + 5\text{CO}(g) + 3\text{B}_2\text{O}_3(l)$$  (5)

Another advantage of using additives such as B$_4$C and C is that one of the reaction products is ZrB$_2$ or HfB$_2$ itself; thus, no secondary phases are generated except the unconsumed/unreacted additives [46]. Most of the studies on diboride ceramics using B$_4$C as a sintering aid have been conducted below 2200 °C, where solid-state sintering was believed to be the main sintering mechanism. In a recent study, it was reported that the sintering can also occur through a liquid phase when the temperature is raised above 2330 °C [46]. However, it was also shown that when liquid phase sintering occurred, there was significant grain growth resulting in the reduction of some mechanical properties such as hardness, fracture strength, and fracture toughness [46].

As discussed in the introduction, among the UHTC composites, Zr(Hf)B$_2$–SiC composites have received the most attention due to their improved mechanical properties (compared to ZrB$_2$ or HfB$_2$) and their oxidation resistance in the presence of SiC. These composites are more difficult to sinter compared to pure diboride ceramics [60] and are generally produced by HP technique. Recently, PS method has also been applied successfully with a small quantity of additives to obtain dense Zr(Hf)B$_2$–SiC composites.

Table 2 shows the processing conditions and sintering additives used to obtain dense Zr(Hf)B$_2$–SiC ceramics via the PS method. Several additives, such as Mo, B$_4$C, and C, have been used to sinter these composites, and they play different roles in improving densification behavior. Mo, for example, improves the densification of the ZrB$_2$–SiC composite by forming a solid solution with the ZrB$_2$ phase [59]. Another obstacle to overcome during sintering is that the SiO$_2$ present on the surfaces of SiC particles reacts with the B$_2$O$_3$ at elevated temperatures. This leads to the formation of borosilicate glass, which causes coarsening instead of densification. B$_4$C alone is not sufficient for the complete removal of this glassy phase. Including C as an additive has been observed to improve densification as it reacts with both the B$_2$O$_3$ and SiO$_2$ at 1000 °C under mild vacuum (20 Pa), as shown in reaction (6) [59]. Additionally, a combination of B$_4$C and C has been shown to produce the best processing conditions for obtaining dense Zr(Hf)B$_2$–SiC composites:

$$\text{SiO}_2 + 3\text{C} \rightarrow \text{SiC} + 2\text{CO}(g)$$  (6)
The optimum processing conditions to produce dense ZrB2–SiC composites with different compositions of particulate phase content (SiC) are shown in Table 2. In ZrB2–SiC composites, the SiC grain morphology changes with the SiC starting particle size. With fine SiC starting particles (0.45 μm), the resulting particulate phase morphology after sintering was in the form of whiskers, whereas with 1.45 μm SiC starting powder, the final morphology in the composite became equiaxed [59]. Recently, Zou et al. [61] showed that ZrB2 at 2000 °C/C14 reduced temperatures (1790 °C) can be produced with elongated grains for both phases when sintered in the presence of 10 vol% SiC at temperatures above 2100 °C for 2 hours. Such microstructure provides better fracture toughness by removing the surface oxides (B2O3), as shown in reaction (7) [62]:

\[
\text{Si}_3\text{N}_4 + 2\text{B}_2\text{O}_3 \rightarrow 4\text{BN} + 3\text{SiO}_2 \tag{7}
\]

### 2.2. Hot-Pressing

Table 3 shows the hot-pressing (HP) sintering conditions for Zr(Hf)B2 ceramics and their composites. This table includes only those processing conditions where a sintered density of 90% or higher was achieved. Unlike the PS technique, HP has size and geometry limitations of the sintered compacts. However, HP produces dense diboride ceramics at relatively lower processing conditions. To achieve high densification in the absence of sintering additives in the HP technique, these diboride ceramics typically require very high temperatures (>2000 °C) with moderate pressures (20–30 MPa) [49] or reduced temperatures (1790–1840 °C) with much higher pressures (800–1500 MPa) [64,65]. Reduction in particle size by attrition milling has also been found to improve the sintering conditions. For example, attrition milled ZrB2 powder (particle size < 2 μm) was sintered to almost full density at 1900 °C [66]. In general, pure HfB2 is more difficult to sinter than pure ZrB2 ceramic. When both ceramics were sintered with comparable starting powder particle size under similar processing conditions, HfB2 was only sintered to 96% at 2000 °C while ZrB2 was completely densified at 1900 °C [62,67].

As can be seen from Table 3, several sintering additives, such as Si3N4, MoSi2, C, ZrSi2, TaSi2, B4C, HfSi2, AlN, and ZrN, have been utilized to produce dense Zr(Hf)B2 ceramics. The roles of Si3N4, B4C, and C additives were discussed in the section on PS technique. When used with the HP technique, Si3N4 was found to be the most effective of these additives at lowering hot-pressing conditions for producing pure ZrB2 ceramics, as can be seen in Table 3. Si3N4 was found to produce a more significant reduction in the hot-pressing conditions when producing pure ZrB2 ceramics than other additives, also seen in Table 3. Additionally, while the PS method requires the addition of sintering aids to produce dense Zr(Hf)B2–SiC composites, the HP technique can produce these composites at lower processing conditions without any further sintering additives (see Tables 2 and 3). Disilicides of the transition metals (such as MoSi2, ZrSi2, TaSi2, and HfSi2) are also potential additives that lower the sintering temperature of diboride ceramics [10]. Improvements in densification are attributed to (i) formation of an intergranular Si–O–B liquid phase due to interaction of SiO2 and B2O3 that occurs on the surfaces of particles and (ii) ductile deformation of disilicide particles at high temperatures (>800 °C) [68,69]. This deformation contributes to densification when it forces the soft dicilicides to fill in the voids left by the diboride skeleton under pressure during sintering.

### 2.3. Nonconventional Processing of Diboride Ceramics

In recent years, nonconventional sintering techniques such as spark plasma sintering (SPS) and plasma pressure compaction (P²C®) have evolved as promising powder consolidation methods for sintering of metallic, ceramic, and intermetallic particles [21–26,28–31]. Unlike the traditional sintering techniques in which consolidation times are on the order of several tens of minutes to hours, these new techniques have successfully sintered poorly sinterable ceramics (i.e., high melting temperature ceramics and non-oxides) to full density in less than 10 minutes.

The sintering process in the SPS and P²C® techniques involves plasma activation and localized resistive heating (i.e., Joule heating) of a powder compact through the application of a low voltage direct current (DC) [21–26,28–31]. Bonding of the particles is accomplished by in-situ elimination of surface oxide (or contaminants) due to application of the pulsed DC voltage (~5 V) and consolidation of the purified powder through “Joule” heating by application of DC voltage and high current (amperage). The in-situ removal of surface oxides helps densification as the surface oxides are known to promote coarsening [10]. Surface oxide removal during sintering can also reduce the requirement of sintering additives. Another advantage of these sintering techniques is that they can retain fine and homogenous grain microstructures due to rapid consolidation [21].

Table 4 shows the processing conditions for Zr(Hf)B2 ceramics and Zr(Hf)B2–SiC composites. Only processing conditions where a sintered density of 90% or higher was achieved have been included, except for two cases of HfB2 ceramics. A comparison of Tables 2, 3, and 4 shows that the ZrB2 ceramics are sintered to almost full density at lower processing conditions (without sintering additives) with the SPS method than with PS and HP techniques. In the SPS
<table>
<thead>
<tr>
<th>Material</th>
<th>Particle size (µm)</th>
<th>Additives</th>
<th>Temperature (°C)</th>
<th>Pressure (MPa)</th>
<th>Time (min)</th>
<th>Final density (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZrB₂</td>
<td>7</td>
<td></td>
<td>2100</td>
<td>28</td>
<td>80</td>
<td>~95</td>
<td>[64]</td>
</tr>
<tr>
<td>ZrB₂</td>
<td>&lt;2 (attrition milled)</td>
<td></td>
<td>1900</td>
<td>32</td>
<td>45</td>
<td>~99.8</td>
<td>[66]</td>
</tr>
<tr>
<td>ZrB₂</td>
<td>0.1–8</td>
<td>2.5 wt% Si₃N₄</td>
<td>1700</td>
<td>30</td>
<td>15</td>
<td>98</td>
<td>[70]</td>
</tr>
<tr>
<td>ZrB₂</td>
<td>2</td>
<td>5 wt% Si₃N₄</td>
<td>1700</td>
<td>30</td>
<td>15</td>
<td>98</td>
<td>[33]</td>
</tr>
<tr>
<td>ZrB₂–10 vol% SiC</td>
<td>4–6 (ZrB₂)</td>
<td></td>
<td>1900</td>
<td>40–50</td>
<td>20</td>
<td>~100</td>
<td>[71]</td>
</tr>
<tr>
<td>ZrB₂–15 vol% SiC</td>
<td>4–6</td>
<td>4.5 vol% ZrN</td>
<td>1900</td>
<td>50</td>
<td>5</td>
<td>99</td>
<td>[72]</td>
</tr>
<tr>
<td>ZrB₂–22.4 vol% SiC</td>
<td>2 (ZrB₂)</td>
<td></td>
<td>1650</td>
<td>60</td>
<td>120</td>
<td>97.9</td>
<td>[73]</td>
</tr>
<tr>
<td>ZrB₂–22.4 vol% SiC</td>
<td>2 (ZrB₂)</td>
<td></td>
<td>1650</td>
<td>60</td>
<td>120</td>
<td>99.6</td>
<td>[73]</td>
</tr>
<tr>
<td>ZrB₂–30 vol% SiC</td>
<td>6 (ZrB₂)</td>
<td></td>
<td>1900</td>
<td>32</td>
<td>45</td>
<td>97.4</td>
<td>[74]</td>
</tr>
<tr>
<td>ZrB₂–30 vol% SiC</td>
<td>0.4 (SiC)</td>
<td></td>
<td>1900</td>
<td>32</td>
<td>45</td>
<td>98.7</td>
<td>[74]</td>
</tr>
<tr>
<td>ZrB₂–16 vol% (SiC + C)</td>
<td>5–10 (ZrB₂)</td>
<td></td>
<td>1800</td>
<td>60</td>
<td>20</td>
<td>100</td>
<td>[75]</td>
</tr>
<tr>
<td>ZrB₂–20 vol% MoSi₂</td>
<td>2 (ZrB₂)</td>
<td></td>
<td>1800</td>
<td>30</td>
<td>5</td>
<td>98.1</td>
<td>[76]</td>
</tr>
<tr>
<td>ZrB₂–20 vol% MoSi₂</td>
<td>2.1 (ZrB₂)</td>
<td></td>
<td>1800</td>
<td>30</td>
<td>30</td>
<td>99.8</td>
<td>[77]</td>
</tr>
<tr>
<td>ZrB₂–20 vol% ZrSi₂</td>
<td>2.1 (ZrB₂)</td>
<td></td>
<td>1400</td>
<td>30</td>
<td>30</td>
<td>99.1</td>
<td>[78]</td>
</tr>
<tr>
<td>ZrB₂–15 vol% TaSi₂</td>
<td></td>
<td></td>
<td>1850</td>
<td>30</td>
<td>10</td>
<td>99</td>
<td>[79]</td>
</tr>
<tr>
<td>HfB₂</td>
<td>5</td>
<td></td>
<td>1840</td>
<td>793</td>
<td>10</td>
<td>≤100</td>
<td>[80]</td>
</tr>
<tr>
<td>HfB₂</td>
<td>5</td>
<td></td>
<td>1790</td>
<td>1551</td>
<td>10</td>
<td>≤100</td>
<td>[80]</td>
</tr>
<tr>
<td>HfB₂</td>
<td>10</td>
<td></td>
<td>2160</td>
<td>27</td>
<td>180</td>
<td>90–95</td>
<td>[81]</td>
</tr>
<tr>
<td>HfB₂</td>
<td>1.37 (HfB₂)</td>
<td></td>
<td>2000</td>
<td>30</td>
<td>60</td>
<td>95.6</td>
<td>[67]</td>
</tr>
<tr>
<td>HfB₂–20 vol% SiC</td>
<td>1.37 (HfB₂)</td>
<td></td>
<td>2000</td>
<td>30</td>
<td>60(1)</td>
<td>99.2</td>
<td>[67]</td>
</tr>
<tr>
<td>HfB₂–20 vol% SiC</td>
<td>2.1 (median particle size of mixture)</td>
<td></td>
<td>2200</td>
<td>25</td>
<td>60</td>
<td>100</td>
<td>[82]</td>
</tr>
</tbody>
</table>
technique, pure ZrB₂ ceramics were sintered to 98% at 1900 °C over a consolidation period of only 3 min. With the addition of MoSi₂, the sintering temperature was reduced to 1750 °C [39]. For ZrB₂–SiC composites, literature on SPS processing is more limited than it is for HP technique. Pure HfB₂ ceramics were difficult to sinter using SPS, similarly to PS and HP techniques. However, with the addition of a small amount of sintering additive (0.1% C), 95% sintered density was achieved at 1850 °C for a consolidation period of only 15 min [86]. The addition of a higher amount of sintering additives resulted in an increase in sintered density at much lower sintering conditions (see Table 4).

For HfB₂–SiC, a significant improvement in processing conditions was achieved, as seen in Table 4.

### 3. OXIDATION BEHAVIOR OF ZR(HF)B₂ CERAMICS AND COMPOSITES

A critical requirement of the structural components used for thermal protection systems in hypersonic vehicles is their thermal stability in highly reactive environments at ultrahigh temperatures. Therefore, understanding the oxidation behavior and exploring the thermal resistance of UHTCs and their composites at elevated temperatures have been the focus of numerous investigations in recent literature. In the following sections, some of these results have been summarized briefly.

ZrB₂ starts to oxidize above 800 °C in air [10,95]. The outer surface of oxidized ZrB₂ components mainly contains a porous layer of zirconia (ZrO₂), on which a thin passive layer of B₂O₃ glass exists (see Eqn (8)). Similarly, HfB₂ also starts to oxidize in this temperature range and forms a porous layer of HfO₂ covered (on the unreacted HfB₂) with a thin layer of B₂O₃ glass, according to the reaction shown at Eqn (9) [84]. Since the porous Zr(Hf)O₂ layer does not have any oxidation resistance, further oxidation of the unreacted Zr(Hf)B₂ is controlled by the diffusion of oxygen (O₂) through the thin layer of B₂O₃ glass [10,96], which has been observed to be an effective barrier to oxygen transport [16]. Thus, the molten B₂O₃ layer provides passive oxidation protection to the underlying unreacted Zr(Hf)B₂. At temperatures above 1200 °C, the upper B₂O₃ layer starts to volatilize (B₂O₃(l) → B₂O₃(g)). Thus, the porus Zr(Hf)O₂ layer is exposed to air followed by an increase in the oxidation rate. As a result, the oxidation resistance of the pure Zr(Hf)B₂ ceramics becomes poor at elevated temperatures [16,97]:

\[
\text{ZrB}_2(s) + \frac{5}{2} \text{O}_2(g) \rightarrow \text{ZrO}_2(s) + \text{B}_2\text{O}_3(g) \quad (8)
\]

\[
\text{HfB}_2(s) + \frac{5}{2} \text{O}_2(g) \rightarrow \text{HfO}_2(s) + \text{B}_2\text{O}_3(g) \quad (9)
\]

It has been shown that when secondary phases such as SiC are added to the Zr(Hf)B₂ ceramics, oxidation resistance

<table>
<thead>
<tr>
<th>Material</th>
<th>Particle size (μm)</th>
<th>Additives</th>
<th>Temperature (°C)</th>
<th>Pressure (MPa)</th>
<th>Time (min)</th>
<th>Final density (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HfB₂−20 vol% SiC−10 vol% B₄C</td>
<td>6 (HfB₂)</td>
<td>—</td>
<td>1850</td>
<td>35</td>
<td>40</td>
<td>98.9</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td>2 (SiC)</td>
<td>—</td>
<td>1850</td>
<td>35</td>
<td>40</td>
<td>98.9</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td>1 (B₄C)</td>
<td>—</td>
<td>1850</td>
<td>35</td>
<td>40</td>
<td>98.9</td>
<td>[83]</td>
</tr>
<tr>
<td>HfB₂−20 vol% SiC−18 vol% AlN</td>
<td>5−7 (HfB₂)</td>
<td>—</td>
<td>1800</td>
<td>30</td>
<td>30</td>
<td>99.6</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td>2−4 (SiC)</td>
<td>—</td>
<td>1800</td>
<td>30</td>
<td>30</td>
<td>99.6</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td>&lt;0.1 (AlN)</td>
<td>—</td>
<td>1800</td>
<td>30</td>
<td>30</td>
<td>99.6</td>
<td>[84]</td>
</tr>
<tr>
<td>HfB₂−15 vol% TaSi₂</td>
<td>1 (HfB₂)</td>
<td>—</td>
<td>1900</td>
<td>30</td>
<td>15</td>
<td>99</td>
<td>[79]</td>
</tr>
<tr>
<td>HfB₂−5 vol% HfSi₂</td>
<td>—</td>
<td>—</td>
<td>1600</td>
<td>30</td>
<td>15</td>
<td>99</td>
<td>[42]</td>
</tr>
<tr>
<td>HfB₂−7 vol% B₄C</td>
<td>—</td>
<td>—</td>
<td>1900</td>
<td>30</td>
<td>40</td>
<td>94</td>
<td>[42]</td>
</tr>
<tr>
<td>HfB₂−30 vol% SiC−2 vol% TaSi₂</td>
<td>1.7 (HfB₂)</td>
<td>—</td>
<td>1900</td>
<td>42</td>
<td>35</td>
<td>100</td>
<td>[26]</td>
</tr>
<tr>
<td></td>
<td>&lt;10 (TaSi₂)</td>
<td>—</td>
<td>1900</td>
<td>42</td>
<td>35</td>
<td>100</td>
<td>[26]</td>
</tr>
<tr>
<td>HfB₂−2.5 wwt% Si₃N₄</td>
<td>2.08 (HfB₂)</td>
<td>—</td>
<td>1800</td>
<td>30</td>
<td>30</td>
<td>100</td>
<td>[85]</td>
</tr>
<tr>
<td>HfB₂−15 vol% MoSi₂</td>
<td>2.08 (HfB₂)</td>
<td>—</td>
<td>1900</td>
<td>30</td>
<td>10</td>
<td>98.7</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td>&lt;2 (MoSi₂)</td>
<td>—</td>
<td>1900</td>
<td>30</td>
<td>10</td>
<td>98.7</td>
<td>[22]</td>
</tr>
</tbody>
</table>
of the resulting composites is significantly enhanced in comparison to the monolithic counter parts [10]. However, oxidation of SiC is insignificant up to 1200 °C; therefore, SiC does not contribute significantly to the oxidation resistance of diborides. Consequently, the oxide scale on Zr(Hf) B2–SiC below 1200 °C consists mainly of porous Zr(Hf)O2 layer and molten B2O3 glass, which is similar to pure Zr(Hf) B2 ceramics [18,98]. As the molten B2O3 layer starts to volatilize above 1200 °C, this allows rapid oxidation of SiC (Eqn (10)) to form a thin SiO2-rich layer (it may also contain some B2O3 due to incomplete evaporation or continued oxidation of Zr(Hf)B2) [16] above the porous ZrO2 layer. Beyond 1400 °C, the thickness of the outer SiO2 layer increases while the B2O3 content decreases significantly.
Clearly, at this temperature, the structure and the composition of the oxidized surface layer are considerably different than those at lower temperatures. The outer SiO₂ layer is followed by a porous Zr(Hf)O₂ + SiO₂ layer followed by Zr(Hf)O₂ + Zr(Hf)B₂ layer and finally the unreacted Zr(Hf)B₂ + SiC. Figure 3 illustrates a schematic of the evolution of different structures during oxidation at different temperatures:

\[
\text{SiC}(s) + O_2(g) \rightarrow \text{SiO}(g) + \text{CO}(g) \quad (10)
\]

\[
\text{SiO}(g) + \frac{1}{2}O_2(g) \rightarrow \text{SiO}_2(l) \quad (11)
\]

The porous Zr(Hf)O₂ + SiO₂ layer, present just beneath the outer SiO₂ layer, results from the oxidation of the porous Zr(Hf)O₂ + SiC layer which forms initially at lower temperatures (~1000 °C) due to oxidation of Zr(Hf)B₂. As a result, the porous Zr(Hf)O₂ layer is covered and partially filled with SiO₂ [18,96,98]. The SiC-depleted Zr(Hf)O₂ and/or Zr(Hf)B₂ layers, present underneath the porous Zr(Hf)O₂ + SiO₂ layer, result from the oxidation of SiC to SiO(g), which migrates toward the outer SiO₂ layer due to the existence of oxygen partial pressure gradient across the SiC-depleted layer. At the interface between the SiC-depleted region and the above SiO₂-containing layer, SiO(g) either oxidizes to form additional SiO₂ or diffuses into the layer and reacts with dissolved oxygen to form SiO₂ closer to the surface of the outer SiO₂ layer. The relevant equations have been shown above in 3.1—3.4.

From the above discussion, it is clear that SiC plays a crucial role in improving the oxidation resistance of diborides above 1400 °C by the passive oxidation protection, which is provided through the formation a SiO₂ layer. The more continuous and thick this layer is the better will be the oxidation protection of the underlying composite. Thus, at a given partial pressure of oxygen, the outer SiO₂ layer formation will depend on the composition (i.e., available SiC) of the Zr(Hf)B₂—SiC composites as well as the exposure time. For a composite with a low SiC content, at a particular temperature, longer oxidation time will be required to develop the protective continuous SiO₂ layer and vice versa for higher SiC content. Thus, the total oxide layer thickness will be higher for low-SiC content composite and lower for high-SiC content composite.

4. MECHANICAL PROPERTIES

Although the current knowledge on processing of UHTCs has matured, investigations on mechanical performance of these ceramics and their composites both at room and elevated temperatures are still limited. In this section, mechanical properties such as fracture strength, hardness, and fracture toughness of the UHTCs will be addressed briefly. This will include the influence of microstructure (grain size of matrix and particulate phase) and particulate phase content on the mechanical properties, effect of high temperature on the mechanical properties, and finally a discussion on the fundamental deformation and fracture mechanisms in these ceramics.

4.1. Elastic Modulus and Hardness

Pure ZrB₂ has Young’s modulus (E) values in the range of 489–493 GPa and HfB₂ has E values in the range of 480–510 GPa [10]. It has been shown that the elastic modulus is affected by the presence of particulate phases [10]. For example, the presence of 10 vol% SiC reduced the modulus of ZrB₂—SiC composite. This could be attributed to the lower stiffness of SiC compared to ZrB₂ as well as to the presence of porosity in the composite [66]. However, with the increase in the vol% of SiC, the modulus of the composite started to increase and approached the value for pure ZrB₂ ceramics.

Hardness values of Zr(Hf)B₂ ceramics and their composites are mainly available from Vickers indentation measurements, and available values have been provided in Table 5. In general, hardness values of Zr(Hf)B₂ ceramics...
<table>
<thead>
<tr>
<th>Material</th>
<th>Grain size (µm)</th>
<th>Density (%)</th>
<th>Hardness (Vickers, GPa)</th>
<th>Fracture toughness (MPa·m$^{1/2}$)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZrB$_2$</td>
<td>~ 7.5</td>
<td>~ 95</td>
<td>~ 95</td>
<td>5.46 ± 0.41 (SENB)</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 4.2</td>
<td>~ 98</td>
<td>~ 98</td>
<td>6.5 ± 0.1</td>
<td>[99]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 3</td>
<td>~ 96</td>
<td>~ 96</td>
<td>7.1 ± 0.2</td>
<td>[99]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 2</td>
<td>~ 97</td>
<td>~ 97</td>
<td>6.81 ± 0.99</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 2</td>
<td>~ 98</td>
<td>~ 98</td>
<td>8.50 ± 0.84</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 2</td>
<td>~ 100</td>
<td>~ 100</td>
<td>8.15 ± 0.43</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 3</td>
<td>98</td>
<td>~ 98</td>
<td>6.6 ± 0.14</td>
<td>[105]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 2</td>
<td>95</td>
<td>~ 95</td>
<td>6.02 ± 0.22</td>
<td>[105]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~ 2</td>
<td>98.3</td>
<td>~ 98</td>
<td>6.7</td>
<td>[106]</td>
</tr>
<tr>
<td>HfB$_2$</td>
<td>~ 3</td>
<td>97.6</td>
<td>18.6 ± 1.1 (9.8 N)</td>
<td>3.3 ± 0.2</td>
<td>[46]</td>
</tr>
</tbody>
</table>
have been seen to fall in a wide range 8.7–16 GPa depending on the density, grain size, indentation load, porosity, etc. [10]. The maximum reported Vickers hardness values were found to be 21–23 GPa for monolithic ZrB2 and ~28 GPa for monolithic HfB2 ceramics [35,66]. The addition of secondary phases, such as SiC, Si3N4, B4C, MoSi2, HfN, and HfC, has been seen to cause slight to moderate increases in hardness mainly due to improvements in the sintered density and reduction in grain size (see Table 5).

### 4.2. Fracture Strength

Fracture strength values of UHTCs such as Zr(Hf)B2 ceramics and their composites are mainly available as flexural strength data from 3-point and 4-point bend tests. A direct comparison is difficult to make due to variations in processing conditions, microstructure, test method, and specimen size. However, this section presents a general overview on the fracture strength, microstructural effects (grain size and particulate phase), and fracture characteristics of Zr(Hf)B2 ceramics and composites.

Table 6 shows a list of tensile fracture strengths for Zr(Hf)B2 ceramics and composites both at room and high temperatures up to 1500 °C. The strength of monolithic ZrB2 at room temperature has been seen to vary between 400 to 600 MPa depending on the processing conditions, density, grain size, etc. [10]. The addition of secondary phases such as SiC, SiC whiskers, Si3N4, AlN, YAG, and MoSi2 has been seen to improve the fracture strength considerably. For example, the addition of SiC was observed to increase the fracture strength from 713 ± 48 MPa for 10 vol% SiC to 1089 ± 152 MPa when 30 vol% SiC was added (see Table 6 [66]). Similarly, a combination of SiC and Si3N4 was also shown to improve the flexural strength of ZrB2 composites as can be seen from Table 6 [33]. Recently, SiC whiskers (SiCw) also have been shown to improve the strength of ZrB2 ceramics. However, an increase in SiCw content over 10 vol% did not improve the fracture strength, as opposed to the SiC particulate phase [99].

Fracture strength data for HfB2 ceramics are rather limited compared to ZrB2 ceramics. Similar to ZrB2, the addition of secondary phases such as MoSi2, SiC, HfN, and HfC improved the fracture strength of HfB2 composites over monolithic HfB2 (see Table 6). In general, HfB2 composites seem to have lower flexural strength values than ZrB2 composites.

#### 4.2.1. Effect of Grain Size on Fracture Strength

One of the factors that strongly affects the fracture strength is grain size. A general trend of increasing fracture strength with decreasing grain size can be seen from Figure 4 [10]. Fracture strength also increases with the addition of secondary phases as mentioned previously. For example, an
<table>
<thead>
<tr>
<th>Material</th>
<th>Grain size (µm)</th>
<th>Density (%)</th>
<th>Flexural strength (MPa) at RT</th>
<th>Flexural strength (MPa) at HT</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZrB₂</td>
<td>–</td>
<td>~94</td>
<td>416 ± 32</td>
<td>–</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB₂</td>
<td>–</td>
<td>87</td>
<td>350 ± 30</td>
<td>340 ± 40 (800 °C)</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>320 ± 60 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>310 ± 10 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–5 vol% Si₃N₄</td>
<td>3 ± 0.1</td>
<td>98</td>
<td>600 ± 90</td>
<td>490 ± 40 (800 °C)</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400 ± 20 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>240 ± 30 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–20 vol% SiC + 4 vol% Si₃N₄</td>
<td>2.4 ± 0.1</td>
<td>98</td>
<td>730 ± 100</td>
<td>550 ± 150 (800 °C)</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>430 ± 110 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250 ± 10 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–18.5 vol% SiC–3.7 vol% Si₃N₄–1 vol% Al₂O₃–0.5 vol% Y₂O₃</td>
<td>2.5 ± 0.1</td>
<td>98</td>
<td>710 ± 110</td>
<td>680 ± 20 (800 °C)</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>630 ± 20 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>280 ± 20 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂</td>
<td>–</td>
<td>86.5</td>
<td>351 ± 31</td>
<td>342 ± 42 (800 °C)</td>
<td>[109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>317 ± 57 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>312 ± 11 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–4 wt% Ni</td>
<td>–</td>
<td>98</td>
<td>371 ± 24</td>
<td>616 ± 8 (600 °C)</td>
<td>[38] [109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>624 ± 37 (800 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>237 ± 3 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15 ± 1 (1200 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–41 wt% TiB₂–4 wt% Ni</td>
<td>–</td>
<td>~100</td>
<td>599 ± 167</td>
<td>582 ± 28 (800 °C)</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250 ± 14 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂–13 wt% B₄C–4 wt% Ni</td>
<td>5–15</td>
<td>99.6</td>
<td>643 ± 86</td>
<td>573 ± 99 (800 °C)</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>148 ± 10 (1000 °C)</td>
<td></td>
</tr>
<tr>
<td>ZrB₂</td>
<td>6</td>
<td>99.8</td>
<td>565 ± 53</td>
<td>–</td>
<td>[66]</td>
</tr>
<tr>
<td>ZrB₂–10 vol% α-SiC</td>
<td>–</td>
<td>93.2</td>
<td>713 ± 48</td>
<td>–</td>
<td>[66]</td>
</tr>
<tr>
<td>ZrB₂–10 vol% α-SiC</td>
<td>~3</td>
<td>~100</td>
<td>835 ± 35</td>
<td>300 ± 35 (1500 °C)</td>
<td>[71]</td>
</tr>
<tr>
<td>ZrB₂–20 vol% α-SiC</td>
<td>–</td>
<td>99.7</td>
<td>1003 ± 94</td>
<td>–</td>
<td>[66]</td>
</tr>
<tr>
<td>ZrB₂–30 vol% α-SiC</td>
<td>~6 (ZrB₂)</td>
<td>99.4</td>
<td>1089 ± 152</td>
<td>–</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>~3 (SiC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZrB₂–30 vol% α-SiC</td>
<td>–</td>
<td>97.2</td>
<td>1063 ± 91</td>
<td>–</td>
<td>[101]</td>
</tr>
<tr>
<td>ZrB₂–25.15 vol% α-SiC</td>
<td>~10 (ZrB₂)</td>
<td>97.67</td>
<td>506 ± 43</td>
<td>–</td>
<td>[102]</td>
</tr>
<tr>
<td></td>
<td>&lt;3 (SiC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZrB₂–5 vol% α-Si₃N₄</td>
<td>3</td>
<td>98</td>
<td>600 ± 90</td>
<td>400 ± 20 (1000 °C)</td>
<td>[103]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>240 ± 30 (1200 °C)</td>
<td></td>
</tr>
</tbody>
</table>
Increase in the strength of ZrB$_2$ with the addition of SiC can be attributed to the presence of SiC phase as well as to the reduction in ZrB$_2$ grain size (see Table 6 [66]). However, for ZrB$_2$–SiC composites, it has been shown in recent literature that the size of SiC particulate phase plays a more crucial role in controlling the fracture strength than the ZrB$_2$ grain size [71].

In monolithic Zr(Hf)B$_2$ ceramics, due to thermal anisotropy, thermal stresses are developed within the grains during cooling from the sintering temperature. When SiC is added, due to its lower coefficient of thermal expansion, tensile stress develops at the interfaces and compressive residual stress develops within the matrix. Also, thermal stress within the matrix decreases as the distance from the matrix–particulate interface increases. As a result, interface areas become potential sites for crack initiation compared to matrix. However, this can be reduced with a decrease in the SiC particle size [7]. This thermal analysis is in good agreement with the work of Rezaie et al. [101], who showed a strong correlation between the room temperature flexural strength of ZrB$_2$–30 vol% SiC composites and the average SiC grain size. It was revealed that the critical flaw size for the composites, calculated from the measured fracture strength and fracture toughness

### TABLE 6 Room and High-Temperature Fracture Strength of Zr(Hf)B$_2$ Ceramics and Composites—Cont’d

<table>
<thead>
<tr>
<th>Material</th>
<th>Grain size (µm)</th>
<th>Density (%)</th>
<th>Flexural strength (MPa) at RT</th>
<th>Flexural strength (MPa) at HT</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZrB$_2$–20 vol% β-SiC–5 vol% α-Si$_3$N$_4$</td>
<td>2.4</td>
<td>98</td>
<td>730 ± 100</td>
<td>430 ± 110 (1000 °C)</td>
<td>[103]</td>
</tr>
<tr>
<td>ZrB$_2$–18.5 vol% β-SiC–3.7 vol% α-Si$_3$N$_4$</td>
<td>2.5</td>
<td>98</td>
<td>710 ± 110</td>
<td>630 ± 110 (1000 °C)</td>
<td>[103]</td>
</tr>
<tr>
<td>ZrB$_2$–20 vol% SiC–5 vol% AlN</td>
<td>2 (ZrB$_2$) 1 (SiC)</td>
<td>100</td>
<td>835 ± 26</td>
<td>–</td>
<td>[104]</td>
</tr>
<tr>
<td>ZrB$_2$</td>
<td>~7.5</td>
<td>~95</td>
<td>629 ± 31</td>
<td>–</td>
<td>[99]</td>
</tr>
<tr>
<td>ZrB$<em>2$–10 vol% SiC$</em>{wp}$</td>
<td>~4.2</td>
<td>~98</td>
<td>786 ± 24</td>
<td>–</td>
<td>[99]</td>
</tr>
<tr>
<td>ZrB$<em>2$–30 vol% SiC$</em>{wp}$</td>
<td>~4</td>
<td>~96</td>
<td>767 ± 16</td>
<td>–</td>
<td>[99]</td>
</tr>
<tr>
<td>ZrB$<em>2$–15 vol% SiC$</em>{wp}$</td>
<td>2–3</td>
<td>~97</td>
<td>545 ± 44</td>
<td>–</td>
<td>[100]</td>
</tr>
<tr>
<td>ZrB$<em>2$–20 vol% SiC$</em>{wp}$ (HP)</td>
<td>~3</td>
<td>98</td>
<td>753 ± 16</td>
<td>–</td>
<td>[105]</td>
</tr>
<tr>
<td>ZrB$<em>2$–20 vol% SiC$</em>{wp}$ (SPS)</td>
<td>~2</td>
<td>95</td>
<td>708 ± 11</td>
<td>–</td>
<td>[105]</td>
</tr>
<tr>
<td>ZrB$<em>2$–20SiC$</em>{wp}$–3YAG</td>
<td>–</td>
<td>98.3</td>
<td>783</td>
<td>–</td>
<td>[106]</td>
</tr>
<tr>
<td>ZrB$_2$–20 vol% MoSi$_2$</td>
<td>2–3</td>
<td>99.1</td>
<td>531 ± 46</td>
<td>655 ± 17 (1200 °C)</td>
<td>[52]</td>
</tr>
<tr>
<td>HfB$_2$–2 wt% B$_4$C</td>
<td>3.4</td>
<td>97.6</td>
<td>483 ± 41</td>
<td>–</td>
<td>[46]</td>
</tr>
<tr>
<td>HfB$_2$ (SPS)</td>
<td>10–20</td>
<td>68</td>
<td>340</td>
<td>345 (1500 °C)</td>
<td>[22]</td>
</tr>
<tr>
<td>HfB$_2$–15 vol% MoSi$_2$ (SPS)</td>
<td>~1</td>
<td>94.7</td>
<td>511 ± 119</td>
<td>573 ± 26 (1500 °C)</td>
<td>[22]</td>
</tr>
<tr>
<td>HfB$_2$–15 vol% MoSi$_2$ (HP)</td>
<td>~1.5</td>
<td>98.7</td>
<td>742 ± 151</td>
<td>664 ± 28 (1200 °C)</td>
<td>[22]</td>
</tr>
<tr>
<td>HfB$_2$–20 vol% SiC</td>
<td>10–20 (HfB$_2$) 2–10 (SiC)</td>
<td>–</td>
<td>453 ± 46</td>
<td>–</td>
<td>[82]</td>
</tr>
<tr>
<td>HfB$_2$–30 vol% SiC</td>
<td>2 (HfB$_2$) ~100</td>
<td>590 ± 50</td>
<td>600 ± 15 (1500 °C)</td>
<td>[25]</td>
<td></td>
</tr>
<tr>
<td>HfB$_2$–19.5 vol% β-SiC–3 vol% HfN</td>
<td>0.5–4 (HfB$_2$)</td>
<td>99.5</td>
<td>650 ± 55</td>
<td>465 ± 45 (1500 °C)</td>
<td>[107]</td>
</tr>
<tr>
<td>HfB$_2$–6 vol% HfC–22 vol% SiC</td>
<td>3 (HfB$_2$) 1 (SiC)</td>
<td>99.2</td>
<td>770 ± 35</td>
<td>310 ± 15 (1500 °C)</td>
<td>[108]</td>
</tr>
</tbody>
</table>
data, was much closer to SiC grain size than ZrB₂ grain size (see Figure 5). This clearly suggested that in ZrB₂–SiC composites, SiC size plays a more important role in controlling the fracture strength than the ZrB₂ grain size [101].

4.2.2. Crack Healing Due to Oxidation

In a recent study, it has been shown that ZrB₂–SiC composites have the ability to self-heal the surface flaws at elevated temperatures due to oxidation. Such self-healing behavior as a result of surface oxidation can potentially improve the strength and thus the high-temperature mechanical behavior [110]. Crack healing is very sensitive to environment, temperature, and time.

Flexural strength measurements were performed on ZrB₂–SiC composites to investigate the effect of environment on crack healing behavior. These flexural tests were performed both on samples containing cracks induced by Vickers indentation and on samples containing cracks that were heat-treated in different atmospheres (air and vacuum). Figure 6(a) shows the comparison of the flexural strengths of the ZrB₂–SiC composites at various conditions [110]. It can be clearly seen that initially flexural strength dropped significantly when surface cracks were induced. However, upon heat treatment in air, the strength was not only recovered but was slightly higher than the polished specimen (no pre-crack) (see Figure 6(a)). Also, air was significantly more effective in healing the surface flaws than a vacuum, where strength recovery was marginal. This is because specimens in air were oxidized, which resulted in complete surface crack healing (indicated by the complete strength recovery). Figure 6(b) shows the effect of temperature on crack healing behavior, suggesting 1000 °C as the optimum temperature for crack healing [110].

4.2.3. High-Temperature Fracture Strength

Zr(Hf)B₂ ceramics and their composites undergo oxidation at elevated temperatures starting from ~800 °C. The oxidation process changes the surface morphology, which continuously evolves as a function of temperature [10,96,98]. Therefore, changes in the fracture strength of these ceramics and composites are expected, which can be seen from Table 6. In general, fracture strength has been seen to degrade with increasing temperature. However, strength often does not degrade up to a certain temperature and sometimes even improves upon oxidation (probably due to crack healing as discussed in previous section). In the following sections, some of the studies on the effect of oxidation temperature, environment, exposure time, and secondary phases on the fracture strength of Zr(Hf)B₂ ceramics will be discussed.

![Figure 4](image)

**Figure 4** Room temperature flexure strength of UHTCS as a function of grain size [10]. For color version of this figure, the reader is referred to the online version of this book.

![Figure 5](image)

**Figure 5** Comparison of calculated critical flaw sizes \( a_{cr} \) with the average and maximum ZrB₂ and SiC grain sizes for ZrB₂–SiC composites prepared from (a) grade A and (b) grade B ZrB₂ precursors [101].
Ramírez-Rico et al. [111] studied the effect of temperature on compressive fracture strength of several ZrB$_2$—SiC composites. Figure 7(a) shows the average compressive strengths as a function of temperature for the ZrB$_2$—SiC composites. For all the composites, the strength degraded as the temperature increased. The ZrB$_2$—SiC composite with polished surfaces showed much higher strength than the composites with unpolished surfaces (up to 1400 °C), suggesting the influence of surface finish on the strength. However, beyond this temperature, strength remained the same for polished and unpolished specimens indicating oxidation effects dominate at these temperatures [111]. Also, ZrB$_2$—SiC composites (with and without any surface polishing) showed higher strengths over the entire temperature range than the ZrB$_2$—SiC and ZrB$_2$—SiC$_{fiber}$ composites. The poor high-temperature behavior of ZrB$_2$—SiC$_{fiber}$ is attributed to excessive cracking, oxidation through the cracks, and elimination of the fiber's coating. As a result, the high-temperature strength of ZrB$_2$—SiC$_{fiber}$ is low compared to the other composites.

Figure 7(b) shows the room temperature compressive strengths of the ZrB$_2$—SiC and ZrB$_2$—SiC—C composites as a function of oxidation time at 1400 °C. For both the composites, the compressive strength, in general, degraded continuously as the exposure time increased. However, for unpolished ZrB$_2$—SiC composite, longer oxidation time was required for strength degradation compared to ZrB$_2$—SiC—C composite [111]. The strength degradation with exposure time was attributed, in general, to the increase in oxide layer thickness in the composites.

Guo et al. [112] studied the effect of submicron size SiC (800—900 nm) phase on the oxidation and mechanical behavior of ZrB$_2$—SiC$_{subm}$ (SiC$_{subm}$ indicates submicron SiC) composites in comparison to a ZrB$_2$—SiC$_m$ composite containing micron size SiC (SiC$_m$, 1.6 μm) particles. To evaluate the effect of SiC particle size on the mechanical

---

**FIGURE 6** Effect of (a) pre-cracking and crack healing environment and (b) crack healing temperature on the flexural strength of a ZrB$_2$—SiCw composite [110]. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 7** Compressive strengths of (a) ZS, ZSC, and ZSS as a function of temperature and (b) ZS and ZSC after oxidation in atmospheric air at 1400 °C [111].
properties, room-temperature flexural strengths of oxidized ZrB2−SiC composites were determined. It was shown that the ZrB2−20 vol% SiCsubm composite had higher room- and high-temperature flexural strengths than the ZrB2−20 vol% SiCm composite. While the room-temperature flexural strengths of all the ZrB2−SiCsubm composites were higher after oxidation, for the ZrB2−SiCm composite, the flexural strength degraded upon oxidation [112]. The ZrB2−20 vol% SiCsubm composites were observed to develop thinner oxide layers than the ZrB2−SiCm composite (see Figure 8). While thinner oxide layers in the ZrB2−20 vol% SiCsubm healed the surface flaws present before oxidation and thus improved the strength, the thicker oxide layer in the ZrB2−SiCm composite weakened the strength [112].

Sciti et al. [113] studied the effect of high-temperature oxidation on flexural strength in ZrB2−20 vol% MoSi2 composites. Figure 9 shows the flexural strengths as a function of oxidation temperature in the range of 700−1400 °C. It can be seen that up to 1200 °C, oxidation did not alter the strength of the composite, although microstructural analysis of the oxidized specimens revealed structural degradation. However, a severe decrease in flexural strength (up to 50%) was observed when the oxidation temperature went beyond 1200 °C, which is a direct consequence of oxidation and corresponding structural degradation [113]. This behavior can be attributed to the rapid subsurface oxidation above 1200 °C, which was accelerated by the formation of cracks and the presence of fast oxidizing species such as ZrC and MoB in the near surface region of the as-sintered specimens [113].

Thus, beyond a certain temperature range (1000−1200 °C), fracture strength of Zr(Hf)B2 ceramics and composites decreases rapidly due to oxidation and surface degradation. However, in the intermediate temperature range, fracture strength often does not degrade or can even improve depending on several factors, as discussed above. In processed ceramics, room temperature strength is largely controlled by the internal defects and surface flaws. On the other hand, high-temperature strength is strongly influenced by the changes in surface morphology and induced defects due to active oxidation. However, oxidation often heals the surface flaws up to a certain temperature range without compromising the structural integrity and thus results in strength enhancement at elevated temperatures. It should be noted that while the presence of SiC is beneficial for improving the oxidation resistance in the temperature range of 1200−1500 °C, the oxide layer formed in this temperature range has less structural integrity than the one that forms around 1000 °C. This lower structural integrity that occurs in Zr(Hf)B2−SiC composites due to oxidation leads to a rapid strength degradation beyond 1200 °C. When metallic phases are present, moderate softening of the second phase in the intermediate temperature range is often beneficial to improve the strength of the ceramic. For example, in the presence of Ni, strength of ZrB2−4 wt% Ni composite increased significantly from room temperature to 800 °C. However, excessive softening due to a further increase in temperature was found to be detrimental, and strength at 1000 °C was observed to be well below the room temperature fracture strength [38].

---

**FIGURE 8** Oxidation behavior in ZrB2−SiCm [112].

**FIGURE 9** Change in flexural strength due to oxidation in the temperature range of 700−1400 °C [113].
In a recent study, Hu and Wang [114] conducted room- and high-temperature (1800 °C) flexure experiments on three ZrB$_2$–SiC composites with different ZrB$_2$ and SiC grain sizes. These studies revealed that at room temperature, ZrB$_2$–SiC composite with the finest grain size exhibited higher strength (as expected). At elevated temperatures an opposite trend was observed [114]. The three composites, ZrB$_2$–15 vol% SiC (ZS1) (ZrB$_2$–4.5 μm and SiC 1 μm), ZrB$_2$–30 vol% SiC (ZS2) (ZrB$_2$–3 μm and SiC 1.5 μm) and ZrB$_2$–15 vol% SiC (ZS3) (ZrB$_2$–8 μm and SiC 3 μm), showed flexural strengths of 865 ± 125 MPa, 705 ± 120 MPa, and 500 ± 40 MPa, respectively, at room temperature. At room temperature the flexural strength was influenced by the SiC particle size, as discussed previously [101]. However, at 1800 °C, the flexural strengths were observed to be 112 ± 12 MPa (13% room temperature strength retention), 48 ± 10 MPa (7% room temperature strength retention), and 217 ± 16 MPa (43% room temperature strength retention), for ZS1, ZS2, and ZS3, respectively. At higher temperatures, the flexural strength was highest for the ZrB$_2$–SiC composite with the largest ZrB$_2$ grain size and the lowest for the one with the finest ZrB$_2$ grain size. Clearly, SiC particle size was not critical at higher temperature, unlike the flexural strengths of ZrB$_2$–SiC composites at room temperature.

From the flexural studies at 1800 °C, it was observed that the bend bars of ZS1 and ZS2 composites exhibited marked plastic deformation, while the ZS3 fractured almost elastically, as can be seen from Figure 10. The ZS1 and ZS2 composites exhibited mixed fracture mode (transgranular and intergranular). On the other hand, the ZS3 composite showed transgranular fracture mode at room temperature, which changed completely to intergranular at 1800 °C [114].

The above studies by Hu and Wang [114] revealed that the factors controlling fracture behavior at room temperature for ZrB$_2$–SiC composites are not the same at elevated temperatures. For example, while finer ZrB$_2$ grain size and higher SiC content have been shown to improve the room-temperature fracture strength, they were found to be less effective at higher temperatures. This is because at higher temperatures, the fracture or deformation in these ceramics is controlled by grain boundary sliding, grain rotation, and cavitation. For ZrB$_2$ ceramics/composites with finer ZrB$_2$ grain size, excessive grain growth at elevated temperatures has been observed, leading to fracture and significant strength degradation [114]. Thus, slightly coarser grain size is beneficial for high-temperature strength retention, although this is a tradeoff for the room-temperature strength. Since the grain boundary strength at high temperature is weaker than the strength of the grains, increasing grain size will reduce the grain boundary areas and prevent grain sliding and rotation, improving the high-temperature strength [114]. In addition, increasing the particulate phase content in ZrB$_2$ composites can be useful in preventing the exaggerated growth of matrix phase at elevated temperatures. However, the ZrB$_2$–SiC interface areas have been shown to contain amorphous films [73], which can also degrade the composite strength at high temperature due to softening. Therefore, the microstructural design of UHTCs cannot be entirely predicted from its room-temperature behavior, and a tradeoff in room-temperature mechanical properties is possibly required to improve the high-temperature deformation and fracture behavior [114].

**4.2.4. Theoretical Modeling of Temperature-Dependent Fracture Strength**

Recently, Li et al. [115] developed a theoretical framework to determine the temperature-dependent fracture strength of UHTCs. In general, there are two methods to change the thermodynamic state or microstructure of a system: application of mechanical load and thermal

![FIGURE 10](image-url) (a) Deformation and fracture behavior of the samples after flexural tests at 1800 °C and (b) the corresponding load–displacement plots [114]. For color version of this figure, the reader is referred to the online version of this book.
heat transfer. Since chemical bonds can be broken either by mechanical load or by thermal load, a form of equivalence can be drawn between strain energy and heat energy. Li et al. [115] assumed total maximum storage energy per unit volume, \( W_{\text{Total}} \), as the critical failure energy. Due to the equivalence between strain energy and heat energy, as mentioned above, the total energy can be described as the summation of the strain energy from applied load \( (W_{\text{th}}(T)) \) and the corresponding equivalent heat energy \( (W_T(T)) \), which can be expressed as

\[
W_{\text{Total}} = K W_T(T) + W_{\text{th}}(T) \]

[115]. Here, \( K \) is the ratio of strain energy to heat energy. The final expression for fracture strength was given as

\[
\sigma_{\text{th}}(T) = \sigma_{\text{th}}^0 \left[ \frac{E(T)}{E_0} \left( 1 - \frac{1}{Tm} \int_0^{Tm} \frac{1}{Cp(T)} dT \right) \right]^{1/2}
\]

(12)

where \( \sigma_{\text{th}}(T) \) and \( E(T) \) are fracture strength and Young’s modulus at temperature \( T \), respectively. \( Cp \) is specific heat capacity at constant pressure and \( Tm \) is the melting temperature [115]. Thus, if the temperature dependency of Young’s modulus of a material is known, then fracture strength can be estimated theoretically. Figure 11 shows a comparison between theoretical predictions and experimental results of fracture strength in Zr(Hf)B₂ ceramics, revealing excellent agreement between the model predictions and experimental data [115].

4.3. Fracture Toughness

The fracture mode in Zr(Hf)B₂ ceramics and composites is predominantly transgranular; however, for finer grain size matrix phase, a mixed mode of fracture (both transgranular and intergranular) has also been observed [100,114]. Figure 12 shows fractured surfaces in ZrB₂–SiC composites, revealing a transgranular mode of fracture in the ZrB₂ phase [116]. Fracture toughness values of Zr(Hf)B₂ ceramics and composites are mainly available for room temperature. The reported fracture toughness values were mostly obtained using direct crack method (DCM) and Chevron notch beam (CNB) techniques. The room-temperature fracture toughness of monolithic ZrB₂ ceramics is usually low, falling in the range from \( 2.8 \pm 0.3 \text{ MPa} \cdot \text{m}^{1/2} \) to \( 3.5 \pm 0.3 \text{ MPa} \cdot \text{m}^{1/2} \) depending on the test method, specimen size, sintered density, grain size, etc. However, a recent study has reported a fracture toughness value as high as \( 5.46 \pm 0.41 \text{ MPa} \cdot \text{m}^{1/2} \) for monolithic ZrB₂ ceramic (see Table 5 [100]).

The addition of second phases has been seen to improve the fracture toughness (see Table 5), a trend that was also observed for hardness and fracture strength. For example, several additives such as SiC, B₄C, Si₃N₄, Y₂O₃, AlN, SiC whiskers, MoSi₂, and HfN have been shown to improve the fracture toughness of Zr(Hf)B₂ ceramics. However, the maximum improvement in fracture toughness was observed with the addition of SiC nano whiskers (SiCₙw) and AlN [100]. In the presence of SiCₙw, the fracture toughness for ZrB₂–15 vol% SiCₙw was found to be \( 6.81 \pm 0.99 \text{ MPa} \cdot \text{m}^{1/2} \), and a fracture toughness as high as \( 8.5 \pm 0.84 \text{ MPa} \cdot \text{m}^{1/2} \) was reported for ZrB₂–10 vol% SiCₙw–3 vol% AlN composite [100]. In general, the increase in fracture toughness in the presence of secondary phases is attributed to several mechanisms, such as crack interactions with the second phases, crack deflection, and crack bridging. For example, extensive crack interactions with the whisker, such as crack deflection, whisker pullout, and bridging, can be seen clearly in Figure 13. It is believed...
that these interactions absorb crack propagation energy during the fracture process and lead to these observed improvements in fracture toughness [105]. Interestingly, SiCnw was observed to be more effective than SiC particles at enhancing the fracture toughness (see Table 5).

5. DEFORMATION BEHAVIOR OF UHTCS

So far, the previous sections have focused on several mechanical properties of UHTCs both at room and high temperatures. Issues related to microstructural control (grain size, second phase) of properties have also been discussed. However, the fundamental deformation and fracture mechanisms that evolve in these ceramics, when subjected to mechanical loading, have not yet been addressed. Thus, this section focuses briefly on the inelastic deformation mechanisms that are rarely observed in other high-temperature structural ceramics. Indentation and scratch processes have been conducted on numerous engineering materials to unearth the deformation and failure mechanisms that operate during wear, abrasion, impact and machining, and control material removal [117–124]. Thus, these studies are useful in identifying the key damage and failure processes which become active during material removal. Investigations into the wear characteristics of ultrahigh temperature ceramics in particular have been extremely limited. Such studies are necessary because hypersonic vehicle structures made of these materials often undergo abrasion and erosion due to sharp particles impinging during take-off, landing, re-entry of vehicles into the earth’s atmosphere, etc. Under these abrasion-dominated wear scenarios (abrasive wear), sharp particles impacting on the exposed surfaces can result in inelastic deformation and material removal.

In recent literature, few studies are available that have focused on the indentation [6] and the scratch [8] behavior of Zr(Hf)B₂ ceramics to unearth the deformation and failure mechanisms. Ceramics are inherently brittle and their failure proceeds without any plastic deformation at a macroscopic scale. Indentation and scratch experiments, however, are known to induce plastic deformation in brittle ceramics due to their point contact load application. Although plastic deformation mechanisms in structural ceramics have been identified through TEM studies, any evidence of macroscopic plasticity is rarely observed in ceramics, especially those with high melting temperatures [125]. In contrast, Zr(Hf)B₂ ceramics have been shown to display unique metal-like plastic deformation features from indentation and scratch studies, which are commonly observed in ductile materials. Such deformation patterns are not observed in typical structural ceramics. Figure 14 shows a static indentation imprint in a polycrystalline ZrB₂ ceramic. In this micrograph, multiple slip-line patterns can be observed in the regions surrounding the indent. Such slip-line patterns are known to develop in metals as a result of...
of dislocation activity. While several ceramics (e.g. Al₂O₃) are known to generate dislocations at room temperature due to mechanical deformation, macroscopic exhibition of dislocation plasticity at room temperature in Zr(Hf)B₂ with such extremely high melting temperature (as shown in Figure 14) is truly unique compared to other structural ceramics. Such macroscopic slip-steps are clear evidence of dislocation-induced plastic deformation and manifestation of dislocation mobility at room temperature in these ceramics. Owing to the extremely high melting temperature of these UHTCs, such room temperature dislocation mobility is unexpected [125].

Clearly, microplasticity was identified as a dominant inelastic process in the Zr(Hf)B₂ ceramics from indentation [6] and scratch [8] experiments. To confirm dislocations in these ceramics, Ghosh et al. [5] conducted TEM studies on a ZrB₂–5 wt% SiC composite where deformation was induced by a microscratch process in the load range of 50–250 mN using a Berkovich nanoindenter. Figure 15(a) shows a SEM micrograph of a scratch groove in the ZrB₂–5 wt% SiC composite containing slip-line patterns in the vicinity of the groove. A focused ion beam (FIB) technique was employed to extract TEM specimens from these regions of scratch grooves for TEM analysis (see Figure 15(b)).

The TEM investigations revealed a high density of dislocations within the deformed ZrB₂ grains in the regions directly beneath the scratch groove. Figure 16(a) shows a bright-field TEM image, at [0001] zone axis, revealing a large area of dense dislocation activity and clustering within the deformed ZrB₂ grains. Dislocation bands in three different orientations were clearly visible, suggesting slip on multiple planes. Similar multiple and intersecting slip bands were also observed at a macroscopic scale on the deformed surfaces (see Figure 16(b)). These slip bands, observed from the TEM and SEM images, appear to be at an orientation of approximately 60° to one another. Such intersecting slip bands and their threefold symmetry are an indication of prismatic or pyramidal slip in ZrB₂. From rigorous dislocation analysis, the slip system was confirmed to be {1010} <1120> [5]. Further, TEM investigations also revealed another slip system in ZrB₂ ceramic, {1120} (0001), less commonly reported in HCP materials [5].
The current TEM investigations clearly revealed prismatic slip as the only identified room temperature plastic deformation mechanism in polycrystalline ZrB$_2$. The above TEM studies are also consistent with the work by Haggerty and Lee in ZrB$_2$ single crystals [126]. There is also considerable evidence [127,128] supporting that prismatic slip is favored over basal slip in ZrB$_2$. Xuan et al. [127] performed Vickers microhardness measurements in ZrB$_2$ single crystals, on basal (0001), and on prismatic planes ($\{10\overline{1}0\}$) and ($\{11\overline{2}0\}$), from room temperature to 1000 °C. It was observed that over the entire temperature range, the hardness values were the same on both types of prismatic planes but were lower than those on the basal plane. A similar result was also obtained in the work of Nakano et al. [128]. When a basal plane is indented, the only possible slip planes are pyramidal planes. On the other hand, indentation on ($\{10\overline{1}0\}$) and ($\{11\overline{2}0\}$) planes can cause slip on other prismatic planes from the ($\{10\overline{1}0\}$) and ($\{11\overline{2}0\}$) families as well as on pyramidal planes. Therefore, it is clear that in ZrB$_2$ prismatic slip is more favorable. The work of Xuan et al. [127] and Nakano et al. [128] also suggested that dislocation motion is probably equally favorable on both ($\{10\overline{1}0\}$) and ($\{11\overline{2}0\}$) prismatic planes. The current TEM investigation clearly revealed that prismatic slip was the only identified mechanism responsible for room temperature plastic deformation in polycrystalline ZrB$_2$. More details of TEM analysis in ZrB$_2$ ceramics can be found elsewhere [5].

Ghosh et al. [5] provided a rationale on the origin of room temperature dislocation activity and mobility in ZrB$_2$ ceramics based on crystal structure and chemical bonding. The nature of chemical bonding strongly influences the properties of metals and ceramics. The nonmetallic character of bonding (mainly covalent) is well reflected in ceramic properties such as high elastic modulus, chemical stability at room and high temperatures, and extremely low electrical conductivity [5]. Interestingly, UHTCs have a unique combination of properties that represent properties of both metals and ceramics [10]. Their high melting temperature (>3000 °C), high elastic modulus (>400 GPa), and superior thermo-chemical stability are typical of ceramics. In addition, UHTCs have surprisingly high electrical conductivity ($>10^6$ S/m) and thus low electrical resistivity ($<10^{-5}$ ohm-cm), comparable to metals. Such electrical properties clearly are in contrast to typical structural ceramics such as aluminum oxide (Al$_2$O$_3$), silicon carbide (SiC), and silicon nitride (Si$_3$N$_4$) as illustrated in Figure 17(a) [5].

The high electrical conductivity suggests considerable metallicity in UHTCs. While the exact mechanism of high electrical conductivity in Zr(Hf)B$_2$ ceramics is not known, Zhang et al. [129] related the metal-like conductive nature of Zr(Hf)B$_2$ ceramics to the Zr(Hf)$-$Zr(Hf) metallic bonding. These ceramics have AlB$_2$-type ($P6_{3}/mm$ space group) hexagonal crystal structure [10]. The unit cell contains a six-member graphite-like ring or a net of boron (B) atoms in two dimensions with top and bottom close-packed layers of metal (M) atoms (see Figure 17(b)). Vajeeston et al. [130] and Zhang et al. [129] suggested from theoretical analysis of chemical bonding that three types of chemical bonds exist in this crystal structure: (i) B$-$B, (ii) M$-$B, and (iii) M$-$M (M=Zr, Hf) [129,130]. While the B$-$B bonds are purely covalent, the M$-$B bonds are a mixture of covalent and ionic characters. However, due to small differences in electronegativity values (1.33, 1.30, and 2.04 for Zr, Hf, and B, respectively), ionicity is less than 8%. In contrast, the M$-$M bonds are predominantly metallic due to nearly free-electron nonbonding $d$-orbital states, which induces significant metallicity in these ceramics. Covalent and metallic bonds are known to strongly influence the properties of materials. Ghosh et al. [5] argued that metallic bonding in Zr(Hf)B$_2$ ceramics could be responsible for the metal-like mechanical deformation behavior as was observed from indentation and scratch studies at room temperature [6,8]. This study correlated such bonding characteristics to the observed room temperature plastic deformation and the associated dislocation mobility in Zr(Hf)B$_2$ ceramics. It should be noted that one of the reasons why such room temperature macroscopic plastic deformation features are not readily observed in other structural ceramics (e.g., Al$_2$O$_3$, ZrO$_2$, SiC, and Si$_3$N$_4$) is that they possess strong covalent and ionic bonding instead of direct metal$-$metal bonds [5]. Such bonding makes dislocation mobility extremely difficult, particularly at room temperature where the thermal energy is limited.

The role of metallic bonding on dislocation mobility can be further realized when one considers the dislocation core structure in these ceramics [5]. Dislocation mobility is strongly influenced by the width of a dislocation core, which depends on the type of chemical bonding. The presence of metallic bonding makes the dislocation core wider, whereas covalent bonding narrows it [131]. For a wide dislocation core consisting of nondirectional metallic bonds, the energy barrier for dislocation mobility will be lower than that of a narrow core structure made up of highly directional covalent bonds [131]. Cubic SiC has the required number of independent slip systems to exhibit room-temperature plasticity. However, due to its pure covalent bonding, the width of the dislocation core is narrow, which causes immobile dislocations at room temperature [131]. An accurate description of this dislocation core structure and its motion in the Zr(Hf)B$_2$ crystal structure is complex and requires an atomistic approach. However, the dislocation core structure due to prismatic slip in ZrB$_2$ not only involves B$-$B and Zr$-$B bonds but also involves the Zr$-$Zr bonds. Such metallic bonds can make the dislocation core relatively wider compared to a core made up of only B$-$B or Zr$-$B bonds. As a result, the Zr$-$Zr bonds are assumed to play a crucial role in enhancing the dislocation mobility at room temperature by
lowering the chemical barrier. Due to this nonlocalized nature of Zr–Zr bonds and the associated increased dislocation mobility during mechanical deformation, dislocations are not restricted to a small material volume (typical of ceramics). Rather, dislocations extend over a relatively larger region allowing the dislocation activity to be readily visible even at room temperature at distances further away from the region of load application, as seen in Figure 14.

Although macroscopic evidence of dislocation mobility was observed in Zr(Hf)B₂ ceramics, it is well known that plastic flow in polycrystalline materials is highly heterogeneous and occurs as intermittent bursts, sometimes referred to as “slip avalanches” [132,133]. This eventually leads to highly localized deformation in the form of surface slip-steps or slip-lines. It is known that such strain localization can reduce ductility and often

**FIGURE 17**  (a) Electrical conductivity of UHTCs, metals and structural ceramics and (b) schematic of ZrB₂ crystal structure. For color version of this figure, the reader is referred to the online version of this book.
triggers failure in materials. Ghosh et al. [116] explored whether such plastic strain localization is influenced by any microstructural length scale such as grain size in ZrB2 ceramics. Grain size and slip-line spacing relationships are available for some metals and alloys [134–137]; however, such literature does not exist for ceramics due to the difficulty associated with the slip-line formation at room temperature. Figure 18 shows SEM and optical micrographs (in Nomarski illumination) of fractured and metallographically polished surfaces revealing the grain morphology and relative grain sizes in four ceramics with different average ZrB2 grain sizes: two polycrystalline ZrB2 ceramics (grain sizes of 20 μm and 41 μm), one ZrB2–5 wt% SiC composite (grain size 5 μm), and one ZrB2–10 wt% SiC composite (grain size 2 μm). Variation in grain sizes among all the ceramics is clearly evident from these micrographs.

Figure 19 illustrates the optical micrographs of typical slip-line patterns formed in the vicinity of indented regions in all the ceramics, where multiple sets of intersecting slip lines were observed. However, due to fine slip-line spacing in the ZrB2–10 wt% SiC composite, slip lines were not visible under the optical micrographs in this ceramic. Only the indent is shown for comparison with other micrographs. Figure 20 shows higher-magnification SEM micrographs revealing slip-line patterns in individual ZrB2 grains in all the four materials. Note that in both the ZrB2–SiC composites, the observed slip-line patterns have evolved only within the ZrB2 phase (matrix) but not in SiC phase. Therefore, even though ZrB2–SiC composites are used, only the ZrB2 phase contributes to slip-line formation. Thus, the comparison is made only among ZrB2 phase grain sizes in the composites with the ZrB2 grain sizes in the monolithic materials.
Figure 21 presents the summary of the measured average slip-line spacing vs. average grain size which clearly reveals a decrease in slip-line spacing with a decrease in grain size [116]. These results also suggested an increase in density of dislocations, i.e., an increase in the localized plastic strain as well as the strain homogenization within the grains, with a decrease in grain size. Thus, the work of Ghosh et al. [116] clearly revealed a definite influence of grain size on slip-line spacing and, thus, local plastic strain in polycrystalline ZrB₂ ceramics. Ghosh et al. [116] rationalized grain size dependence of slip-line spacing from different viewpoints: (i) plastic strain accommodation in polycrystalline materials in contrast to a single crystal, (ii) using the concept of geometrically necessary dislocations (GNDs) to correlate both strain gradient and dislocation density with slip-line spacing, and (iii) influence of interface boundaries on slip-line spacing in the ZrB₂–SiC composites. Here, the rationalization for grain size dependence of slip-line spacing is discussed based on the concept of GNDs, although more details can be found elsewhere [116].

Dislocations are induced during deformation to accommodate the plastic strain. Conceptually, they have been classified into statistically stored dislocations (SSDs) and geometrically necessary dislocations (GNDs) [138–140]. SSDs are considered for homogeneous strain, whereas GNDs are treated as the additional defects that are required to accommodate the strain gradient due to nonuniform plastic strain. In polycrystalline materials, the requirement of strain compatibility among the grains results in nonuniform strain within each grain [116]. In the indentation process, the plastic strain is highly nonuniform within the deformed zone. It is known that density of GNDs scales with the magnitude of nonuniform plastic strain (i.e., strain

Figure 19 Optical micrographs (Nomarski illumination) of (a) ZrB₂-10 wt% SiC, (b) ZrB₂-5 wt% SiC, (c) HP-ZrB₂, and (d) SPS-ZrB₂. Note the evolution of multiple slip-lines in the vicinity of indented regions. Due to extremely fine grain size and spacing of slip-lines, they cannot be seen in ZrB₂-10 wt% SiC under optical microscope: only indent is shown for consistency. For color version of this figure, the reader is referred to the online version of this book.
gradient) [141]. Nix and Gao [139] have attributed GNDs to the formation of slip lines on the indented surfaces where large strain gradients exist. Figure 22 shows a schematic, partially adopted from Nix and Gao [139], of a conical indentation along with the GND structure which is idealized as circular dislocation loops. The formation of slip lines (and slip-steps) is not only limited to the indented surface but also extends to the outside surfaces in the vicinity of the indentation, similar to the observations in the current work (see Figures 14, 19 and 20). Here, the higher dislocation density near the indenter is indicated by a smaller symbol (•) and the lower dislocation density in the regions away from the indenter is shown by a bigger symbol (○).

The idealized slip-line spacing on the indented surface has been indicated as \( S \) and the spacing on the surface outside the indent is shown as \( S_1 \) (see Figure 22). If \( \theta \) is the angle between the surface of the indenter and the specimen plane, then the spacing \( \cdot S \) can be derived as [139]

\[
S = \frac{Ba}{h} = \frac{nba}{h}
\]

(13)

where \( h \) is the indentation depth, \( a \) is the contact radius, and \( B \) is the apparent Burgers vector expressed as \( B = nb \), where \( n \) is the number of “real” planar dislocations that have slipped on the glide plane as a result of a Frank—Read process [142] and \( b \) is Burgers vector. Thus, it is clear that the slip-line spacing is related to indentation depth and diagonal length. Similarly, outside the indentation (see Figure 22), the ratio between apparent Burgers vector \( B_1 \) (\( B_1 = mb \), where \( m \) is the number of “real” planar dislocations that have slipped on the glide plane similar to \( n \) and \( m \neq n \)) and \( S_1 \) can be expressed as

\[
\tan \theta_1 = \frac{B_1}{S_1}
\]

(14)

Therefore, \( S_1 \) can be further expressed as

\[
S_1 = \left( \frac{m}{n} \right) \left( \frac{h}{a \tan \theta_1} \right) S
\]

(15)

It should be noted that \( S \) and \( S_1 \) are idealized slip-line spacings in a continuum. However, in a polycrystalline...
material, these quantities can be assumed as average slip-line spacing. The slip-line spacing ($S$) was extremely fine on the indented surfaces and, therefore, was nonmeasurable. Instead, $S_1$ was more readily detectable and measurable as can be seen from Figures 14 and 17. Note from Eqn (15) that the slip-line spacing ($S_1$) outside the indented region is also related to $h$ and $a$.

The total density of dislocations ($\rho_T$), in presence of a strain gradient, is expressed as

$$\rho_T = \rho_G + \rho_s$$

(16)

where $\rho_s$ is the density of statistically stored dislocations and $\rho_G$ is the density of geometrically stored dislocations. It has been shown that $\rho_s$ depends on the average strain of indentation which is related to the shape of the indenter ($\tan \theta$) [143]. In contrast, $\rho_G$ is related to the indentation depth ($h$) as follows:

$$\rho_G = \left( \frac{3}{2bh} \right) \tan^2 \theta$$

(17)

Therefore, in a given material and with a particular indenter type, the density of GNDs is related to the inverse of indentation depth which has also been utilized to explain the depth dependence of hardness (i.e., indentation size effect) [139]. Also, the density of GNDs scales with the average strain (in each grain) divided by the grain size. In other words, at a given strain level, density of GNDs scales with the reciprocal of grain size. Therefore, at a given indentation load, although the average strain may remain the same, the density of GNDs will increase with decrease in grain size.

**FIGURE 21**  Slip-line spacing in all the ceramics revealing the increase in average spacing with increasing average grain size [116]. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 22**  Schematic of a conical indentation in a crystalline solid. In the plastic zone, the dislocation density is higher near the indenter compared to regions away from it [116,139].
6. SUMMARY

Zr(Hf)B₂ ceramics and Zr(Hf)B₂–SiC composites have been identified as the materials with the most potential for ultrahigh temperature applications in hypersonic and supersonic vehicles. Significant progress has been made toward the processing of these ceramics and composites with a high sintered density (>95%). In the pressureless sintering method, sintering additives have been used to produce dense compacts, which often have been found to reduce the sintering temperature. In hot-press techniques, processing conditions (temperature and time) are usually lower than in PS due the presence of external pressure. While for pure Zr(Hf)B₂ ceramics, mechanical properties have also been shown to play a major role in achieving high-density ceramics and composites at lower processing conditions. Finally, with the advantage of nonconventional sintering techniques (e.g., SPS), processing conditions have been further reduced than in PS and HP techniques.

A considerable effort has been put forward to understand the thermal resistance of Zr(Hf)B₂ ceramics and composites. These ceramics undergo oxidation above 800 °C; however, the resulting surface oxide layer (B₂O₃) provides resistance to further oxidation. However, above 1200 °C, the oxidation resistance of pure Zr(Hf)B₂ ceramics degrades due to the evaporation of the B₂O₃ layer. The high-temperature oxidation resistance has been shown to improve considerably when SiC was added. SiC also oxidizes at higher temperatures, but the resulting oxide layer (SiO₂) is stable up to 1800 °C. Therefore, Zr(Hf)B₂–SiC composites provide better thermal stability at elevated temperature regimes than the pure Zr(Hf)B₂ ceramics. Mechanical properties of these ceramics and composites are mainly available as elastic modulus, hardness, fracture, strength, and fracture toughness. Grain size plays a crucial role in improving the mechanical properties. Similar to oxidation behavior, mechanical properties have also been shown to improve for the Zr(Hf)B₂–SiC composites than the pure ceramics. Therefore, SiC has twofold beneficial effects where it not only improves the oxidation resistance but also enhances the mechanical properties. When SiC is present, mechanical properties are influenced strongly by the particulate phase size than the matrix grain size. Additionally, increase in volume fraction of the particulate phase also improves the mechanical properties.

Zr(Hf)B₂ ceramics and composites revealed readily detectable unique metal-like plastic deformation features in terms of slip lines, which were observed to evolve within the Zr(Hf)B₂ matrix phase. Such observations clearly suggested room temperature dislocation activity and mobility, which are highly unusual in ceramics having extremely high melting temperatures. Existence of dislocations was confirmed through TEM analyses and prismatic slip was identified as the plastic deformation mechanism in poly-crystalline ZrB₂. It was argued that in addition to having extremely high electrical conductivity, formation of typical metal-like plastic deformation features at room temperature was also a reflection of the metallicity present in chemical bonding in these ceramics. In this context, a possible role of Zr–Zr metallic bonds (present in the ZrB₂ crystal structure) on the dislocation mobility has been discussed. It was also shown that ZrB₂ grain size has a pronounced effect on the slip-line spacing. The average slip-line spacing within the ZrB₂ phase was observed to increase with increasing ZrB₂ grain size for pure ZrB₂ ceramics as well as for ZrB₂–SiC composites. This observation was rationalized in the context of geometrically necessary dislocations.

ACKNOWLEDGMENT

The authors acknowledge Amy Pittman for editing and helping with manuscript preparation.

REFERENCES

Handbook of Advanced Ceramics


Ceramic Bearings and Seals

Mathias Herrmann
Fraunhofer-Institut für Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Chapter Outline
1. Introduction 301
2. Ceramic Bearings 301
3. Ceramic Sliding Bearings and Mechanical Seals 309
4. Corrosion of Ceramic Sealing and Bearing Materials 314

References 324

1. INTRODUCTION
Structural ceramic materials are widely used in different wear applications—in mills, mixers, linings of tubes for abrasive slurries etc., bearings, sealing devices and inserts for cutting of metals, knives paper and textiles, and orthopedic and dental implants. The market for ceramic wear parts is about one third of the market for engineering ceramics in Europe [1]. The main groups of wear parts are seals and bearings, cutting tools and bioceramics (Table 1). These components have a sphere of influence in the market that is more than 100 times higher than the figures given in Table 1 because in many cases only ceramic parts allow the reliable functioning of a device or process. Due to their higher costs ceramic wear parts are usually used in applications where other materials fail. The materials undergo different wear mechanisms depending on the environment and loading conditions. Special applications require special materials with special microstructures, properties and surface structures. The following chapter presents an overview of the microstructures, properties and advantages of ceramic materials used in bearings and seals. A brief summary of corrosion mechanisms occurring in ceramic materials is provided at the end of the chapter.

2. CERAMIC BEARINGS
The principle of using rolling elements to reduce friction was known by the ancient Egyptians. Steel bearings have been in use since 1883 and the first ceramic bearings were prepared in the middle of the 1980s [1–4].

Bearings may consist of ceramic rollers or balls and ceramic raceways or, in the case of hybrid bearings, ceramic rolling components and metal raceways (Figure 1). Hybrid bearings offer most of the advantages of ceramic bearings at a lower cost and hence most ceramic bearing components are hybrid bearings (99% versus 1% full-ceramic bearings [3]). They are mainly used in semiconductor fabrication equipment, special pumps, instruments and applications with elevated temperatures or highly corrosive environments.

In 2009 the world market for rolling contact bearings was approximately €20 B, and full-ceramic and hybrid bearings only made up €140 M (<1%). This share is expected to increase to €1 B by 2020 [3]. Depending on series volume and bearing type, a hybrid bearing is 2.5–25 times more expensive than metal bearings. This cost issue is the main reason for the limited use of such bearings.

The stresses generated during application are highly localized. Thus ceramic ball bearings typically do not fail catastrophically, but rather exhibit cracking or chipping in the race or the cylindrical rollers (Figure 2). The stress in the race strongly depends on the accuracy of the race geometry, the elastic constants of the materials and the friction. Therefore high-precision finishing is required to produce ceramic bearings with reproducible properties.

The high local stress generated in the bearing dictates the requirements of the materials used:
- high strength,
- high fracture toughness,
- low local defect sizes and
- high fatigue resistance.

Silicon nitride ceramics meet these requirements and are the main ceramic materials used in bearings. Different stabilized
ZrO₂ materials are also used, but have a lower maximum allowable load than Si₃N₄. The ZrO₂ bearings have the advantage of possessing a similar thermal expansion coefficient to those of the surrounding steel components, allowing for thermal stress compensation at high-temperatures.

A comparison of the properties of a typical silicon nitride bearing material and bearing steel is given in Table 2. The effects of the properties on bearing behavior are also discussed. Ceramic and hybrid bearings are found to be highly advantageous for use in high-speed, high-precision applications, in dry or near-dry (minimum quantity of lubrication) conditions and in corrosive, abrasive or high-temperature environments.

Figure 3 illustrates one of the advantages of the hybrid bearing over steel bearings. As mentioned in Table 2, lower friction forces and much less variation in the forces over time are found for the hybrid bearing than for steel bearings under the same conditions. Such bearings are relatively expensive and are used in applications in which other materials do not come into consideration (see Table 3). The bearings are used in numerous industrial sectors and have started to penetrate the consumer market (Table 3).

2.1. Silicon Nitride Bearing Materials

In this section the demands on Si₃N₄ ceramic materials used for bearings and the consequences for the microstructure are discussed. Silicon nitride ceramics for ball bearings are sintered using different sintering additives which form liquids during sintering at 1700–2000 °C, allowing densification by the liquid phase sintering mechanism [5]. After cooling, a three-dimensional network of additives (Figure 4) is formed. Most of the additives are concentrated in the grains, but they also form amorphous films of a few nanometers in thickness between the grains (see also chapter Silicon nitride ceramics). The grain size of the Si₃N₄, the grainboundary phase volume and composition and the material defect size are important parameters determining the behavior of the bearings.

As can be deduced from Figure 2a, the local tensile stress during application can reach levels of 1 GPa and higher. Hence the defect sizes in the bearings must be minimized. Typically the surface defect size must not exceed 20 μm in high-performance bearings. Hot isostatic pressing, i.e. under an external pressure of up to 1000–2000 atm, is used to produce most ceramic balls for bearings to minimize the defects. Gas pressure sintering (with a nitrogen gas pressure of up to 100 atm) is typically used for producing the races to achieve materials with low defect sizes and high strengths (800–1400 MPa).

Failure of the balls and raceway is caused by a combination of cyclic compression and tensile stresses as well as wear caused by sliding friction. The sliding friction is caused by slip of parts of the ball or the cylindrical rollers...

---

**TABLE 1** Market for Engineering Ceramics in Europe and the US in 2010 (millions of euros) and Expected Growth Rate [1]

<table>
<thead>
<tr>
<th></th>
<th>US market</th>
<th>European market</th>
<th>AAGR in % (European market)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural ceramics</td>
<td>1737</td>
<td>2643</td>
<td>3.1</td>
</tr>
<tr>
<td>Mechanical and wear parts</td>
<td>826</td>
<td>764</td>
<td>3.8</td>
</tr>
<tr>
<td>Bearings</td>
<td>13</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Pump seals</td>
<td>110</td>
<td>100</td>
<td>1.5</td>
</tr>
<tr>
<td>Cutting tools</td>
<td>110</td>
<td>78</td>
<td>3.6</td>
</tr>
<tr>
<td>Bioceramics</td>
<td>73</td>
<td>115</td>
<td>10</td>
</tr>
<tr>
<td>Other wear parts</td>
<td>269</td>
<td>406</td>
<td>2.5</td>
</tr>
</tbody>
</table>

---

**FIGURE 1** Schematic view of a ball bearing [3]. For color version of this figure, the reader is referred to the online version of this book.
FIGURE 2  Calculated stress distribution in the raceway and typical failure caused by overloading: calculated stress in the ball-raceway contact zone [41] (a), localized stress causing overloaded bearings and in turn resulting in surface cracking perpendicular to the direction of rolling (b) or in chipping of the rolling elements [10] (c). For color version of this figure, the reader is referred to the online version of this book.
### TABLE 2 Comparison of Properties of Metal and Si₃N₄ Ceramic Bearings [16,38,39]

<table>
<thead>
<tr>
<th>Property</th>
<th>Metal bearing</th>
<th>Si₃N₄ ceramic/hybrid bearings*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density, g/cm³</td>
<td>7.8</td>
<td>3.1–3.4 Reduced weight/reduced centrifugal forces</td>
</tr>
<tr>
<td>Hardness (HK 0.5)</td>
<td>750</td>
<td>1500</td>
</tr>
<tr>
<td>Elastic modulus, GPa</td>
<td>210</td>
<td>290–320 Reduced plasticity results in higher stiffness, smaller deformation at rolling contact point, 11% and 24% higher stiffness for hybrid and full-ceramic bearings respectively than for metal bearings [2], increased speed capability in comparison with conventional all-steel bearings</td>
</tr>
<tr>
<td>Chemical activity of surface</td>
<td>Metallic bonds result in microwelding and subsequent microfracture without lubrication, Higher friction, more lubrication necessary, higher wear</td>
<td>Dry running or minimum quantity of lubrication possible, lubrication with media possible Lower friction results in smaller increase in temperature during operation</td>
</tr>
<tr>
<td>Electrical properties</td>
<td>Conductive</td>
<td>Nonconductive/nonmagnetic</td>
</tr>
<tr>
<td>Corrosion resistance</td>
<td></td>
<td>Generally higher corrosion resistance in comparison with bearing steels</td>
</tr>
<tr>
<td>Thermal stability, °C</td>
<td>180</td>
<td>&gt;800 high-load durability at high temperature</td>
</tr>
<tr>
<td>Linear expansion coefficient, 10⁻⁶/K</td>
<td>12.5</td>
<td>2.8–3.5 smaller tolerances caused by temperature rise result in reduced vibration</td>
</tr>
</tbody>
</table>

*Typical values for Si₃N₄.

### TABLE 3 Application Fields for Ceramic and Hybrid Bearings as a Function of the Main Requirement During Application

<table>
<thead>
<tr>
<th>Main parameter</th>
<th>Effect</th>
<th>Applications</th>
</tr>
</thead>
</table>
| High-speed              | Lower centrifugal force, higher stiffness, lower friction | - Medical equipment  
- Spindles in machine tools, industrial equipment (spin tester), cars (Formula 1) |
| High temperature        | Heat resistance (ceramics), no need for lubrication | - Metal production facilities  
- Industrial equipment |
| Corrosive environments  | Corrosion resistance, no need for lubrication | - Chemical equipment  
- Agitator motors in polyethylene production, lubrication by media at 200 °C  
- Food packaging industry (aseptic usage, dry running)  
- Textile machinery  
- Semiconductor production |
| Vacuum environments     | No or strongly reduced need for lubrication  | - Semiconductor production  
- Turbomolecular pumps |
| Use in magnetic and electrical fields | Nonmagnetic  
Nonconductive | - Medical equipment (MRI)  
- Special electric motors |
| Low friction            | No or minimal need for lubrication          | - Sport/consumer applications  
- Food industry |
| High-precision/high-load application | High stiffness | - Windmills (>1 MW)  
- Measuring instruments |
on the raceway. Slip is affected by the design and usually the slip ratio is around 1–10% or less.

Cyclic loading (ball on plate) of Si$_3$N$_4$ materials can be used to simulate the conditions causing ball bearings to undergo fatigue. Cycling loading experiments show that in materials with coarse-grained structures a network of microcracks forms and grows with increasing number of cycles. The extent of microcrack formation is much lower in materials with fine-grained structures, but overloading results in formation of conical cracks several micrometers in length [6]. Materials with relatively fine-grained microstructures are hence preferable for ball bearings [3,7].

Figure 5 shows an experimentally observed dependence of the wear mode on the maximum Hertzian pressure and the slip ratio in rolling-sliding dry contact conditions in air [8]. The operating conditions for ball bearings must be chosen in such a way that only mild wear takes place. The mild wear region is characterized by a smoothing of the surface. Typically the roughness reaches a steady state in the range of 1/10 of the grain size, i.e. 10–100 nm [8–13]. Figure 6 illustrates the change in the surface roughness of the raceway during operation. The smoothing is caused by mechanical wear as well as tribochemical reactions. Tribooxidation takes place in silicon nitride and the other nonoxide ceramics in the contact zone. The reaction of silicon nitride with oxygen and water vapor can be described by the following equations:

$$\text{Si}_3\text{N}_4 + 3\text{O}_2 \rightarrow 3\text{SiO}_2 + 2\text{N}_2$$

$$\text{Si}_3\text{N}_4 + 6\text{H}_2\text{O} \rightarrow 3\text{SiO}_2 + 4\text{NH}_3$$

The SiO$_2$ can form a lubricating film, debris or rolls depending on the conditions and thus can strongly reduce the friction coefficient [8].
In water the formed SiO₂ dissolves and hence SiO₂ and NH₃ can be removed from the friction surface. As the load and running time is increased the grain boundaries undergo damage and are partially removed (Figure 6c) [7,10,14]. This weakens the material and can be the starting point for crack formation or breaking out of grains (Figure 2). Similar behavior is found for LPSSiC (liquefac phase sintered SiC) [15]. Severe wear is typically reflected by breaking out of complete grains and an increased roughness and results in rapid bearing failure, as shown in Figure [9].

The described overall wear scenario shows that the bearing behavior strongly depends on the loading conditions, the lubricating media and the microstructure (grain size as well as grain boundary phase amount and composition).

Figure 7 demonstrates the difference between the coefficients of friction (COFs) under dry running conditions and in water [16]. Even water appears to be an effective lubricant for ceramic bearings. The COFs in water-lubricated and grease-lubricated systems are similar. A low COF not only reduces the energy
consumption, but also strongly affects the loading capacity of the bearing. Model calculations show that the tensile stress on a ball bearing depends on the COF as well as the ball size, the raceway geometry and the load [16–18]. The model calculations reveal that the bearing loading capacity can be increased by a factor of 10 if the COF for sliding friction between ball and raceway is reduced from 0.6 to 0.23 or even by a factor of 500 if the COF is lowered to 0.13 (oil lubrication [16,18]). In Figure 7b, c the loading capacities of different silicon nitride materials under dry running conditions and in water (with stepwise increases in the load after $10^3$ revolutions) up to crack formation are shown [16]. The maximum allowable load is ten times higher for water lubrication than for dry running conditions. This can be attributed to the strong reduction in the COFs in water-lubricated bearings compared with those of bearings operated in dry running conditions (Figure 7a).

Figure 7 also gives an indication of the influence of the microstructure on bearing behavior. Reduction of the coefficient of friction of the Si$_3$N$_4$ materials through the addition of TiN (0.95 vol.%) or BN results in better performance in dry running conditions, but has little effect under wet conditions. TiN additives in bearings are quite common and not only reduce the COFs, but also change the colors of the materials [5,19]. The addition of BN to bearing materials is less common because it inhibits the sintering of silicon nitride and lowers the material hardness and strength.

Experimental results have shown that liquid phase sintered silicon carbide exhibits similar loading capacities to those of Si$_3$N$_4$ materials [16]. However, ball bearings are not usually made from these materials.

The last few years have seen the development of \(\alpha\)-SiAlON materials with improved fracture toughness and strength [5,20–23]. The grain boundary phase content is much lower (normally less than 2–4 vol.%) in these materials than in the \(\beta\)-Si$_3$N$_4$ materials used in ball bearings. These materials were recently shown to have higher wear resistance levels than those of \(\beta\)-Si$_3$N$_4$ materials under various conditions [24,25]. Although they have not yet been used in bearings, these materials still show potential for an increase in lifetime and corrosion resistance.

In comparison with standard materials [10], fine-grained or nanostructured \(\beta\)-Si$_3$N$_4$ ceramics have lower COFs and result in higher bearing loading capacity and lifetime. Under dry friction (Roell Amsler, 10% slip) the standard Si$_3$N$_4$ material can be loaded up to 1.2 GPa, whereas the nanosized material can be loaded under the same conditions up to a Hertzian pressure of 2 GPa (Figure 8) [26]. Similar effects are found for \(\alpha/\beta\)-silicon nitride materials produced by spark plasma sintering [27,28]. The sintering time is insufficient to allow conversion of the metastable \(\alpha\)-modification to \(\beta\)-Si$_3$N$_4$. The resultant materials exhibit grains in the submicrometer to nanometer range, higher hardness and reduced coefficient of friction and wear. The new effective densification techniques [27,28]; see chapter Spark Plasma Sintering] allow new material concepts, e.g. cBN-Si$_3$N$_4$ materials, to be realized. These possibilities have only begun to be explored.

**FIGURE 8** Coefficients of friction of different Si$_3$N$_4$ ceramics under dry rolling wear conditions: fine-grained materials (grain size $\leq$100 nm) show lower coefficients of friction and higher loading capacities. For color version of this figure, the reader is referred to the online version of this book.
The dry friction—and hence the lifetime and loading of the bearings—can be tailored through the use of Si$_3$N$_4$/SiC composites. SiC inclusions reduce the wear and the COFs [25,29]. Materials with the same composition but different rare earth oxide additives show a successive reduction in the coefficient of friction from La to Lu [29]. This is mainly due to the accompanying increase in grain boundary phase strength [29,30].

Besides the grain boundary phase composition, the grain boundary phase content has an important influence. The ceramic ball bearing lifetime under dry running conditions (at a Hertzian pressure of 1.75 GPa) has been investigated as a function of the amount and extent of crystallization of the grain boundary [7]. Reduction of the amount of the grain boundary phase by 30–50% and prevention of local crystallization of the amorphous grain boundary phase has been shown to result in a twofold increase in lifetime over that of standard materials [7].

Under certain conditions electrical conductivity is desired to avoid electrostatic effects. Different possibilities, i.e. addition of silicides or TiN, have been explored [5,10,19]. These materials must be added at concentrations higher than the percolation threshold, i.e. higher than 15–25 vol.% [5,19]. This changes the mechanical properties as well as the wear and tribological behavior [5]. Over the last few years attempts have been made to prepare Si$_3$N$_4$–carbon nanotube (CNT) composite materials containing less than 5 vol.% of CNTs [14,31]. For these materials special processing techniques to achieve a homogeneous distribution and special sintering techniques to achieve full density without decomposition of the CNTs are necessary.

Besides pores or impurities caused by the quality of the powders or by the processing, local inhomogeneities called ‘snowflakes’ in the materials can be critical for high-performance bearings. These inhomogeneities are visible in dark field optical images (Figure 9). Detailed SEM investigations show that these areas contain microporosity [32–34]. The snowflakes can occur for different reasons, but most often they are caused by local crystallization of the grain boundaries [32–34].

As described at the beginning of this chapter, stress is localized in the near-surface area during application. Surface finishing and the stresses resulting from these processes hence have a strong influence on bearing behavior. Introduction of compressive stresses into the

![FIGURE 9](image) Optical and SEM micrographs of a Si$_3$N$_4$ ceramic containing ‘snowflakes’: dark field images of different ceramics (a, b), SEM image of snowflake region (c).
surface of the ceramic, e.g. by shot peening [35], can greatly increase the loading capacity. Shot peening has been shown to result in a fourfold increase in bearing loading capacity. This offers the possibility of improving the behavior of ceramic ball bearings without changing the material.

The fact that ceramic bearings have very high levels of reliability is indicated by the use of hybrid bearings in the space shuttle main engine pumping liquid hydrogen [2]. Nevertheless, proof tests could improve the behavior of ball bearings operating at their limits. A method has recently been developed for balls [36]. Danzer et al. have developed a method for determination of the strength of ceramic balls enabling direct comparison of the quality of different batches or materials in the form of balls [37]. This in turn allows for better quality control of balls and hence bearings.

Standardization of silicon nitride materials for ball bearings is underway and will give engineers a better basis for exhausting the possibilities of hybrid and full-ceramic bearings (Table 4). However, this standardization process does not encompass the amount and nature of the grain boundary phase, both of which have a strong effect on wear behavior and corrosion. Most currently used materials are class I materials (Table 4), but for certain applications class 2 or class 3 materials can be used at lower cost without causing any changes in performance. As mentioned at the beginning of this chapter, even though the cost of ceramic bearings has dropped dramatically over the last few years, it is still higher than the cost of metal bearings (the price of a ½-inch ball is approximately 2 to 4 times that of M50 steel balls [1]). Efforts are currently being made to reduce the cost of the materials for bearing applications by lowering the raw material costs and improving the processing and sintering behavior.

### 3. CERAMIC SLIDING BEARINGS AND MECHANICAL SEALS

Numerous machines and systems have rotating parts which must be protected or isolated from the environment. This function is fulfilled not only by roller bearings, but also by sliding bearings and mechanical seals. Sliding bearings support transverse (radial) and thrust (axial) loads on the rotating part by sliding a rotating face [42–49] on the stationary face. In these kinds of bearings pure sliding friction takes place, whereas in roller bearings rolling friction is the main source of friction. The design of seals and bearings ensures proper lubrication by a liquid (usually oil, grease or the pumping media) or gas under the given conditions to minimize friction. Lubricated bearings typically have COFs of less than 0.01 [50].

Mechanical seals consist of two rings or discs which rotate relative to each other. One of the mating faces is stationary and the other is attached to the rotating shaft. The rings are pressed together by a defined force often realized by a spring so that contact is maintained (Figure 10).

The two contacting surfaces are usually flat and polished. A fluid, either the liquid used in the system or a lubricant from an external source, is introduced into a gap between the two surfaces. During normal operation a continuous thin film of 1–3 μm in thickness is formed

---

**TABLE 4** Properties of Silicon Nitride Materials Suitable for Roller Bearing Balls (ISO FDIS 26602:2008) [40]

<table>
<thead>
<tr>
<th>Property</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>g/m³</td>
<td>3.0–3.6</td>
</tr>
<tr>
<td>Elastic modulus</td>
<td>GPa</td>
<td>270–330</td>
</tr>
<tr>
<td>Poisson’s ratio</td>
<td></td>
<td>0.23–0.29</td>
</tr>
<tr>
<td>Thermal conductivity (room temperature)</td>
<td>W/m K</td>
<td>20–40</td>
</tr>
<tr>
<td>Coefficient of thermal expansion (room temperature to 500 °C)</td>
<td>×10⁻⁶/K</td>
<td>20–3.7</td>
</tr>
<tr>
<td>Transverse rupture strength³ 4-point σ₄ (Weibull modulus)</td>
<td>MPa  σ₄</td>
<td>Material Class I: 760 (12)</td>
</tr>
<tr>
<td>Hardness HV5, HV10</td>
<td>GPa</td>
<td>Material Class I: 14.2</td>
</tr>
<tr>
<td>Fracture Toughness Kᵥ (ASTM 1421 or JIS R 1607)</td>
<td>MPa√m</td>
<td>Material Class I: 6</td>
</tr>
<tr>
<td>Microstructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pore size</td>
<td>μm</td>
<td>Material Class I: 10</td>
</tr>
<tr>
<td>Inclusions of size 25 &lt; x ≤ 50 Number/μm²</td>
<td>Material Class I: ≤4</td>
<td></td>
</tr>
<tr>
<td>Inclusions of size 100 &lt; x ≤ 200 Number/μm²</td>
<td>Material Class I: ≤0</td>
<td></td>
</tr>
</tbody>
</table>

³Distance between supports: 40 mm
between the two mating faces. A wide range of different constructions can be produced for different application conditions. Despite their differences, the materials used in seals and sliding bearings have many features in common.

The product PV of the nominal contact pressure P and the mean sliding velocity at the interface V is frequently used as a means of comparison of materials and seal behavior. The maximum PV value for a given lifetime is only valid for a specific seal design and media at a given temperature. Therefore it only provides a rough indication of the overall performance of a seal or material [1,45,50].

Different materials are used in industry for sliding bearings and mechanical seals. The choice of material strongly depends on the application conditions, desired lifetime and economic constraints. A rough overview of the most commonly used materials is given in Table 5. Ceramics clearly represent just one of a myriad of different materials used. Like ceramic ball bearings, ceramic seals and sliding bearings are used in high-performance applications in which high wear and corrosion resistance is required. The need for hermetically sealed pumps and systems results in an increased demand for seals lubricated by the pumped medium. This in turn necessitates the use of materials which can cope with the severe conditions of corrosive media, abrasive particles or running either with mixed film lubrication (e.g. gas formation near the boiling point of the media) or no lubrication. Under these conditions ceramic seals and sliding bearings are the first choice.

In Figure 11 the normalized PV_{max} values under dry running conditions are given for different materials. The data show the relatively low PV limits for combinations of two hard components under dry running conditions. Hence materials of differing coefficient of elasticity and hardness

| Properties of Different Seal and Sliding Bearing Materials [1,44,50] |
|----------------|----------------|----------------|----------------|----------------|
|                | Density, kg/m³ | Young’s modulus, GN/m² | Bending strength, MPa | Thermal conductivity, W/mK | Hardness (Vickers/ Shore) | Thermal expansion coefficient, 10⁻⁶ K⁻¹ |
| Carbon graphite, resin-impregnated | 1800 | 23 | 65 | 11.0 | 90–100 (Sh A) | 3.0 |
| Carbon graphite, antimony-filled | 2500 | 33 | 90 | 13.5 | 85–95 (Sh A) | 3.5 |
| PTFE, 25% glass | 2250 | – | – | 0.4 | 70–75 (Sh A) | 44–92 |
| Stellite 1 Ni resist | 8690 | 248 | – | 15.0 | 600 (V) | 11.3 |
|                | 7300 | 96 | – | 10.0 | 150 (V) | 19.0 |
| 18/2 martensitic steel AISI 431 | 7720 | 210 | – | 18.8 | 261–319 (V) | 11.1 |
| Lead bronze BS1 1400 1985 1B4 | 9000 | 68 | – | 71.0 | 65–90 (V) | 18.4 |
| Alumina, 99.5% | 3870 | 365 | 300–400 | 30.0 | 2500 (V) | 6.9 |
| Tungsten carbide, Co binder | 14,700 | 630 | 1700–2500 | 88.0 | 1500–1600 (V) | 5.1 |
| Tungsten carbide, Ni binder | 14,700 | 600 | 1700–2500 | 73.0 | 1300–1500 (V) | 4.8 |
| Silicon carbide, reaction-bonded | 3100 | 413 | 400–500 | 200.0 | 2500–3500 (V) | 4.3 |
| Silicon carbide, sintered | 3100 | 390–420 | 400–450 | 100.0 | 2500 (V) | 4.8 |
| Silicon carbide, graphite | 2800 | 130–200 | 100–130 | 153.0 | – | 3.4 |
are frequently used in sliding bearings and mechanical seals; e.g. combinations of resin-impregnated carbon or plastics and various metals are used for low loads. The data also indicate the relatively good behavior of SiC/C sliding pairs.

Ceramic sliding bearings and mechanical seals have been in use since the end of the 1970s/beginning of the 1980s [1,50,51]. The main materials from which ceramic seals are made are sintered silicon carbide (over 55%), Si$_3$C$_4$ and C$_x$SiC$_z$ (25–30%) [1]. The second most important ceramic material for sliding bearings and mechanical seals is alumina (approximately 15%), although SiC can be used at higher speeds and has a superior wear behavior. Al$_2$O$_3$ or ZrO$_2$ materials are mostly used in applications with low chemical and thermal loads, mainly because of the relatively low cost of Al$_2$O$_3$ ceramics in comparison with SiC. Hard metals or cermets (e.g. WC/Ni) are used in applications in which the seal components must have high fracture toughness values.

The cost of a ceramic component strongly depends on the production volume. For example SiC seals for automotive water pumps are produced on a large scale (75% of cars in Europe have silicon carbide pump seals) and have a unit cost of €0.30–0.50 [1].

Table 6 gives a more detailed overview of different SiC-based materials used in seals and bearings. The first SiC seals were made from Si-infiltrated SiC (SiSiC). This material is prepared by the infiltration of a SiC green body with silicon at 1550–1650 °C under a vacuum. The resulting dense material contains SiC as well as a small amount of unreacted silicon which strongly reduces the stability of the material in basic solutions [50,53,54]. Hence nowadays these materials are only used in special applications. Most applications utilize silicon carbide seals made by solid state sintering. This kind of material is sintered at a temperature of greater than 2000 °C using carbon-, boron- or aluminum-based additives [50,53,54]. The SSiC materials all contain z-SiC polytypes as the dominant phase. Depending on the starting composition and processing method they may contain residual graphite or boron carbide. The materials exhibit low COFs and high chemical resistivity, thermal conductivity and thermal shock resistivity. SSiC is mainly used in seals and bearings due to these properties.

Liquid phase sintered silicon carbide exhibits higher strength and fracture toughness than SSiC and can be sintered at lower temperatures (1800–1950 °C) due to the use

![FIGURE 11 Normalized PV limits for different material combinations under dry running conditions (data from [1]; normalized based on SiC/SiC combination). For color version of this figure, the reader is referred to the online version of this book.](image-url)

**Table 6 Physical Properties of SiC Ceramics used in Tribological Applications [50,52]**

<table>
<thead>
<tr>
<th>Material type</th>
<th>Microstructural phases/features</th>
<th>Density, g cm$^{-3}$</th>
<th>Porosity, %</th>
<th>Young’s modulus, GPa</th>
<th>Thermal expansion coefficient at 20–500 °C, $10^{-6}$/K</th>
<th>Thermal conductivity at 20 °C, W/mK</th>
<th>Flexural strength at 20 °C, MPa</th>
<th>Knoop hardness HK0.1, GPa</th>
<th>KIC, MPam$^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKasic C</td>
<td>SiC, fine-grained</td>
<td>&gt;3.1–3.35</td>
<td>&lt;3</td>
<td>410–420</td>
<td>4.1</td>
<td>125</td>
<td>400–500</td>
<td>24.5</td>
<td>4</td>
</tr>
<tr>
<td>EKasic F</td>
<td>SiC, coarse-grained</td>
<td>&gt;3.15</td>
<td>0.5–0.6</td>
<td>410–420</td>
<td>4.0</td>
<td>110</td>
<td>400–430</td>
<td>25</td>
<td>3.5</td>
</tr>
<tr>
<td>EKasic G</td>
<td>SiC + graphite</td>
<td>&gt;3.0</td>
<td>&lt;3</td>
<td>410–420</td>
<td>4.1</td>
<td>110</td>
<td>250–300</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>EKasic P</td>
<td>SiC, pores</td>
<td>&gt;2.70–2.90</td>
<td>10–14</td>
<td>410–420</td>
<td>4.0</td>
<td>90</td>
<td>210–230</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>SSiC</td>
<td>SiC, Si</td>
<td>3.05–3.15</td>
<td>&lt;1</td>
<td>400–420</td>
<td>4.3</td>
<td>60</td>
<td>350</td>
<td>–</td>
<td>3.5–4</td>
</tr>
<tr>
<td>SiC 30</td>
<td>SiC, Si, graphite</td>
<td>2.65</td>
<td>0.5</td>
<td>190–210</td>
<td>4$^1$</td>
<td>125</td>
<td>140</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>LPSSiC</td>
<td>SiC, aluminates</td>
<td>&gt;3.2</td>
<td>&lt;1</td>
<td>420–430</td>
<td>3.5</td>
<td>75</td>
<td>550</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

$^1$ Temperature range: 20–1000 °C.
of rare earth oxide or MgO with aluminum oxide or nitride additives (2–10 vol.%) which form a liquid phase at high-temperature and solidify to form an oxide grain boundary phase during cooling [50,55–58]. This grain boundary phase reduces the very high corrosion stability of the SiC [50,59] and hence makes these materials suitable for use in special applications (e.g. gas seals) only [50,60]. The additives allow the microstructure to be tailored flexibly and any desired wear, friction [15,65–67], mechanical [55–58] and electrical properties [61–64] to be achieved.

LPSSiC with higher strength and fracture toughness levels than those of SSiC materials is used in gas seals. Under normal operating conditions the gas seal rings are separated by a very thin gas film (less than 5 µm in thickness) and run at high-speed (200 m/s). The high centrifugal forces create high-precision machining, high edge stability and low surface roughness are required. These requirements are best met by fine-grained LPSSiC materials.

In water-based environments the tribological behavior of SiC is determined by tribochemical reactions resulting in the formation of hydrated silica (SiO$_2$ * xH$_2$O). These reactions yield products which can be dissolved in the media (especially in basic solutions) and result in a certain amount of surface polishing [50,59,65–70]. Disruption of ideal running conditions is highly problematic because in such cases the lubrication film is disrupted and the sliding faces must run either in dry or in mixed film lubrication conditions. Friction causes high local temperatures and thermal stresses or thermal shock. This in turn can result in particle breakout and increased wear of the sliding phase. SiC ceramics have a very good combination of properties (high thermal conductivity, low thermal expansion and moderate strength) and thus can survive short periods of dry running or mixed friction [51,60,69,71].

Seal and bearing behavior can be significantly improved by the right microstructural design and surface morphology. Great efforts have been made over the last 30 years to understand the relationships between wear and microstructure. The main requirements for proper functioning of seals and bearings [50] are:

- stabilization of the hydrodynamic lubrication film and prevention of dry running, or
- reduction of friction and wear if the lubrication film breaks down.

The above requirements can be met through optimization of the design, especially the morphologies of the sliding faces, and the materials.

The surface morphology can be altered through introduction of controlled porosity [72,74,75], production of surface structures/textures by treatment with a laser [23,76] and tailoring of the microstructure [15,50,51,68–70,97]. The structure and roughness of the surface formed during operation strongly depends on the grain size of the material. During the run-in period tribochemical reactions occur to yield a ‘steady state’ surface dependent upon the load/velocity ratio. This is a decisive factor for the film stability [26,72,76].

The lifetime for seals made from SSiC materials with different microstructures in deionized water is given in Figure 12. Under these conditions the SiC material with a fine-grained structure exhibits the lowest lifetime. The behavior of the silicon carbide material with a bimodal coarse-grained microstructure (grain size up to 1500 µm; see Figure 13) is significantly better. These coarse-grained materials are particularly superior to fine-grained structures in applications in hot water in which tribochemical reactions (formation of SiO$_2$) can weaken the grain boundaries [66,76]. The large grains function as reinforcements because they extend into areas where no chemical attack can weaken the grain boundaries. The coarse-grained structure also has a higher surface roughness, producing reservoirs for residual liquid in the seal gap if the hydrodynamic film is interrupted. The same effect can be seen in the case of controlled residual porosity (Ekasis P; see Figures 12 and 13). Wear can be improved especially under dry running or mixed film lubrication conditions through incorporation of graphite into the material (Ekasis G; see Figures 12 and 13). These inclusions act as solid lubricants and are protected against excessive wear by the surrounding hard SiC matrix.

However, under dry running conditions and under conditions with mixed lubrication SiC/SiC shows insufficient behavior during application. Under dry running the coefficient of friction of SiC/SiC is too high. After the hydrodynamic film is disrupted the wear increases sharply and destruction of the seal is possible. Under these conditions the combination of SiC with a graphite-containing material yields higher system reliability even though the wear rate is higher under standard conditions [44,50,77,78].
Besides graphite, SiC/C composites can be used for such purposes. Siliconized graphite is produced by infiltration of graphite with liquid silicon. These materials contain 60–65% silicon carbide, 30–35% graphite and up to 5% residual free silicon and exhibit good properties under dry running conditions [52]. The remaining silicon reduces the stability of the materials in basic solutions such as SiSiC materials.

Carbon fiber-reinforced graphite or SiC materials are also used for large seals in power plants [77,78] to achieve high reliability.

The dry running behavior can also be improved through formation of a graphite layer on the SiC surface. This can be achieved by thermal treatment in furnaces, resulting in evaporation of Si from the SiC surface. Such layers can also be produced locally by laser treatment [69,78,79] or by chemical reaction of the SiC with chlorine-containing compounds at 600–900 °C or in water under hydrothermal conditions [80,81]. An alternative to chemical modification is application of different coatings. Table 7 gives a brief comparison of possible coatings.

**TABLE 7 Overview of Different Coatings for SiC Bearings and Seals [60]**

<table>
<thead>
<tr>
<th>Type</th>
<th>Method of production</th>
<th>Feature</th>
<th>Possible application area</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphite</td>
<td>Decomposition of SiC (thermal decomposition or chemical reaction)</td>
<td>Layer thicknesses of up to 100 μm, independent of geometry, cost-effective</td>
<td>Standard applications for seals and friction bearings in hermetic pumps</td>
<td>[60,79–82,98]</td>
</tr>
<tr>
<td>Diamond-like carbon (DLC)</td>
<td>Physical vapor deposition (PVD), plasma-enhanced chemical vapor deposition (PECVD)</td>
<td>Good dry running conditions, quite high hardness</td>
<td>Seals and friction bearings in hermetic pumps</td>
<td>[83–85,89–91]</td>
</tr>
<tr>
<td>Crystalline diamond coatings</td>
<td>Chemical vapor deposition (CVD)</td>
<td>Good tribological properties, superhard material, thermal stability in air up to 800 °C</td>
<td>Seals and bearings under high loads and/or in dry running conditions</td>
<td>[71,86–88,92]</td>
</tr>
</tbody>
</table>
Diamond-like carbon (DLC) coatings represent a family of carbon-based coatings containing carbon and hydrogen as well as different metals. A detailed overview of DLC coatings is given in the literature [83,90]. The coatings contain carbon in the sp² hybridized (graphite) and sp³ (diamond) hybridized states and hence combine the properties of both modifications of carbon. The coatings exhibit low COFs during dry running [60,84,85,87]. Due to the low coating thickness (approximately 3 μm) the dry running time is limited. Such coatings are currently being used successfully in gas seals. Under normal running conditions the seal rings have no contact, but during start-up and shutdown contact occurs. The DLC coatings must prevent destruction of the seal by reducing the coefficient of friction in this relatively short period [60].

Crystalline diamond coatings several micrometers (usually 10—20 μm) in thickness can be deposited on SiC seals using chemical vapor deposition (Figure 14). Diamond is formed at a gas (H₂ + CH₄) pressure of 1—100 mbar. The gas is activated by a hot tungsten filament [88,92]. CVD diamond-coated seals can operate under dry running conditions or in slurries containing highly abrasive particles. During the run-in period the diamond tips are cut, resulting in smoothing of the surfaces. Further wear and the COFs are very low, enabling a long lifetime to be reached [71,88]. Such systems have been introduced in the market by Burgmann Industries GmbH & Co. KG. In Figure 15 the wear coefficients of diamond coatings and other ceramic materials are compared. In Figure 16 the COFs for standard SiC seals under dry running, insufficient lubrication and mixed friction conditions are given. The data indicate very low COFs in comparison with SiC and other sealing materials. This and the high wear resistance result in a more than tenfold increase in lifetime [88]. Under extreme application conditions seals for pumps for steam-assisted gravity drainage used in deep drilling operations exhibit a twenty fold increase in lifetime [71,88].

Diamond coatings also have the advantage of reducing or completely avoiding the electrochemical corrosion of SiC seals that can take place in high-purity water [93—95] (see also next chapter).

The long lifetime and greatly lowered costs of maintenance for the system more than compensate for the higher production costs for these coated seals in applications with severe operating conditions.

It was recently observed that the coefficient of friction of SiC could be influenced through application of an electric potential [96].

4. CORROSION OF CERAMIC SEALING AND BEARING MATERIALS

Corrosion is defined as the local or large-area reaction of a material with the surrounding media resulting in the degradation of the materials or component and potentially even in the destruction of components. The corrosion behavior is therefore not a property of the material, but rather a property of the system including the material, the media and conditions such as temperature, pressure, volume or flow rate of the medium and tribological or mechanical load. Corrosion is a very complex process which strongly depends on both the microstructures and compositions of the ceramics and the properties of the system. Common structural ceramics are resistant to typical organic lubricants or solvents. More pronounced corrosion can be observed in water-based solutions. Depending on the conditions the slow dissolution in acids or the formation of hydroxides can take place as shown by the following reactions:

\[ \text{MO}_x + 2y \text{H}^+ \rightarrow M_{(aq)}^{2y^+} + y \text{H}_2\text{O} \]
\[ \text{MO}_x + y \text{H}_2\text{O} \rightarrow M(\text{OH})_{2y} \]

These processes are very similar to that of the dissolution of the protective coatings of metals.

Nonoxide ceramics react with water, acids or bases to form oxides or hydroxides. For example, Si₃N₄ reacts with...
water to form hydrated silica and ammonia. The hydrated silica results in a passivating layer.

In SiC ceramics, as in many other carbides (TiC, WC,…), the corrosion reaction is a redox reaction, rather than just hydrolysis. Additional electrochemical corrosion reactions can occur because of this and the fact that SiC ceramics are semiconductors.

Besides surface reactions which, for instance, can affect wear resistance, local reactions at pores or inclusions can result in the formation of pits and a subsequent reduction in the strength of the ceramic materials. Chemical reactions are also known to accelerate subcritical crack growth and therefore can strongly reduce the lifetime of a component used under tensile stress in corrosive media.

Figure 17 shows the change in residual strength after corrosion in different media for 100 h for various ceramic materials. High-purity alumina ceramics and solid phase sintered SiC materials show no change in strength due to their high stability and the absence of a three-dimensional network of grain boundary phases (as is found in liquid phase sintered materials such as Si₃N₄ and LPSSiC ceramics). All other materials show a media-dependent reduction in strength due to the formation of new defects or growth of existing defects through the corrosion reaction.

The stability of alumina ceramics is strongly dependent upon the purity and additives. Alumina ceramics with glassy silicate grain boundaries exhibit a much lower corrosion resistance in comparison with pure materials due to the dissolution of the grain boundary.
phases. MgO-containing Al₂O₃ ceramics form highly stable spinel and hence show similar behavior to that of the pure materials [99–101].

For ZrO₂-containing materials it is well known that especially under hydrothermal conditions the metastable tetragonal phase transforms to the monoclinic phase. This is associated with a volume increase and results in local tensile stresses. These stresses in turn cause microcracking and ultimately result in macroscopic cracks. MgO- and CeO₂-stabilized ZrO₂ materials are slightly more stable and are hence preferentially used in bearings and seals in cases in which ZrO₂ must be used [2,3]. This degradation strongly depends on the microstructure of ZrO₂ ceramics [137–140].

As already mentioned Si₃N₄ and liquid phase sintered SiC (LPSSiC) ceramics are densified via liquid phase sintering. During cooling from the sintering temperature the liquid forms a three-dimensional network of crystalline or amorphous grain boundaries. In Si₃N₄ materials the sintering additives and the SiO₂ on the surfaces of the Si₃N₄ starting powders form the grain boundary phase. Thus the grain boundaries are mainly amorphous [5,10,31,111]. In LPSSiC materials the SiO₂ on the SiC powder surfaces is nearly completely reduced during sintering. Hence the grain boundary is made up of crystalline aluminates [15,50,53,55–59,61–63]. The different chemical stabilities of these grain boundary phases are the main reason for the different behavior of the different Si₃N₄ materials and also for the differences between the LPSSiC and SSiC materials shown in Figure 17.

Corrosion also strongly affects the friction and wear behavior. In Figure 18 the wear rate and the friction coefficient in water at room temperature up to 300 °C are shown for alumina, SiC and Si₃N₄ ceramics. The data illustrate the strong influence of the reactivity with water on the wear volume. For the SiC and Al₂O₃ ceramics the coefficient of friction only changes slightly with temperature, but the specific wear volume increases strongly due to the reaction of water with the ceramic surface. The investigated Si₃N₄ material shows the most pronounced increase in friction coefficient and wear volume. This can be attributed to the lower stability of the grain boundary (with MgO and Al₂O₃ additives) of this material in comparison with the other two investigated materials.

Si₃N₄- and SiC-based ceramics are the main materials used in bearings and seals and hence further discussion will concentrate on the corrosion behavior of these materials.

4.1. Corrosion Resistance of Silicon Nitride Ceramics

With the advent of Si₃N₄ use in ball bearings in the 1990s came increased efforts to investigate the corrosion stability of these ceramics, mostly in Japan [104–110,133–136]. These investigations revealed the strong relationship between the grain boundary phase and the chemical stability. In Si₃N₄ ceramics the grain boundary phase is distributed among the triple junctions and the thin films between the grains. The triple junctions form a three-dimensional network and have cross-sectional diameters of less than 1–2 μm. The thin grain boundary films between adjacent grains have thicknesses typically in the range of 1–2 nm, strongly depending on the grain boundary composition [5,10,31,111]. In materials with 5–10 vol.% additives more than 90% of the grain boundary phase is concentrated in the triple junctions. These microstructural features must be taken into account in the interpretation of the corrosion resistance.

Conventional sintered materials with Y₂O₃/Al₂O₃ have much lower strengths after corrosion in acids at elevated temperatures due to the dissolution of the grain boundary phase in acids. This dissolution of the grain boundary phase can occur to depths of several millimeters without the three-dimensional network of the Si₃N₄ grains being destroyed (Figure 19). Even materials with completely dissolved grain boundaries exhibit strengths of 400–500 MPa. Investigations showed that the extent of
corrosion in acids can be greatly reduced through the use of MgO or MgO and Al2O3 as sintering additives [59,103,112]. Only at the beginning of the twenty-first century were comprehensive investigations of the relationship between grain boundary phase composition and corrosion behavior carried out [103,112—120]. They revealed that the corrosion behavior of silicon nitride ceramics with amorphous grain boundary phases exposed to H2SO4, HCl, other mineral acids except HF and bases at temperatures to at least 100—130 ºC depends on the stability of the grain boundary phase only due to the high stability of the silicon nitride grains themselves. The stability of the grain boundary phase can be described using the basic relations known in glass science [118,121—124]. Typically the glass network modifiers (in Si3N4 ceramics Y3+, rare earth, Mg2+, and, as intermediates, Al3+) dissolve in the acids and a hydrated network remains. In a second step the Si–O–Si bonds in the network break and the SiO4 units dissolve in the acids. This process is shown schematically in Figure 20.

The destruction of the network is normally the process with the lowest rate. Therefore more strongly linked networks are more stable. A strong correlation between the observed linear reaction rates and the structures of the grain boundaries has been found. Based on these data and the proposed corrosion mechanism it can be concluded that the main factor influencing the stability of grain boundary phases in acids is the amount of bridging anions and network-forming cations, i.e. the SiO2 content of the grain boundary phase. This finding is confirmed by the fact that

![Figure 19](image1.png) SEM micrograph of a polished surface after corrosion in acid (1 N H2SO4 (a) and in 1 N NaOH (b)) (Si3N4 material with Y2O3/Al2O3).

![Figure 20](image2.png) Schematic diagram illustrating the corrosion process in acids or bases. For color version of this figure, the reader is referred to the online version of this book.
HIPed Si₃N₄ with no sintering additives shows very high corrosion resistance in acids [133]. The nature of the network-modifying ions (e.g. Y³⁺, rare earth, Mg²⁺ and/or Ca²⁺) has a much lower influence.

The kinetics of corrosion is rather complex and is often described by different laws. A detailed analysis of the different behavior is given in the literature [112–117,125]. For short corrosion times usually a reaction-controlled mechanism is observed, whereas for longer times the process can be greatly slowed by passivation mechanisms (Figure 21). The change in mechanism depends on the grain boundary composition, the temperature and the concentration of the acid (Figure 22).

The reason for this passivation is the formation of SiO₂-rich deposits in the pores. The detailed mechanisms and conditions of formation of these layers are discussed in the literature [102,103,112–117,125].

Such hydrated networks are also known to form in glasses. The formation of these hydrated networks is much more pronounced in the pores of the Si₃N₄ ceramics than in glasses. This is due to kinetic factors [102,103] as well as hydrolysis of the surfaces of the Si₃N₄ grains (a few nanometers in 100 h). The strong remaining skeleton of Si₃N₄ grains (see Figure 19) stabilizes this layer. The formation of a SiO₂-rich passivating layer in Si₃N₄ ceramics also leads to a much higher corrosion resistance in acids of high concentration than in acids of medium concentration (Figure 21).

The anions in HCl, HNO₃ or H₂SO₄ do not appear to have any special effects. Fluorine anions strongly influence the corrosion of Si₃N₄ ceramics due to their ability to dissolve SiO₂ [107,115,119]. Even at low concentrations the F ions prevent the formation of passivating layers and dissolve the thin films between the Si₃N₄ grains, resulting in a non-adhering corrosion layer from which the Si₃N₄ grains can be removed very easily. This also results in a significant reduction in strength [115]. The corrosion rate is found to be lower in H₃PO₄ solutions than in HCl and other mineral acids due to the formation of rare earth phosphates with low solubility [5].

The influence of the microstructure on corrosion can be summarized as follows: ceramics with high corrosion resistance in acids are those with low amounts of grain boundary phase and large numbers of bridging anions. This is shown schematically in Figure 23.

The mechanisms underlying corrosion resistance in bases and under hydrothermal conditions are the same. For standard Si₃N₄ materials the weight loss and the corrosion depth are much lower in bases at least up to concentrations of 1 mol/l and temperatures up to 130 °C than in acids. This can also be explained by the stability of the grain boundary phase and the known relations from glass science. At the concentrations mentioned above Y₂O₃ and Al₂O₃ form hydroxides passivating the grain boundaries. Under very strong alkaline conditions Al₂O₃, MgO and SiO₂ constituents of the grain boundary can easily be dissolved and therefore under these conditions the corrosion stability is reduced. In summary the materials that are the most stable in acids due to a high grain

**FIGURE 21** Corrosion rates of Si₃N₄ ceramics with different additives in 1 N H₂SO₄ at 90 °C [125] (The labelling shows the kind of additives (Mg: MgO, Al: Al₂O₃, Y: Y₂O₃, Si: SiO₂) and the number corresponds to the volume content).
FIGURE 22  Time dependence of the thickness of the corrosion layer and weight loss during corrosion (a) as a function of temperature in 0.5 mol H₂SO₄ (material with 6 wt.% Y₂O₃ and 4 wt.% Al₂O₃ at 90 °C) and (b) for materials with grain boundary phases with different SiO₂ and constant Y₂O₃/Al₂O₃ contents at 90 °C in 0.5 mol H₂SO₄ following Schilm et al. [102] (mol % SiO₂ given in the diagram) and (c) dependence of weight loss on corrosion duration for the material with 6 wt.% Y₂O₃ and 4 wt.% Al₂O₃ in H₂SO₄ at different concentrations at 90 °C following Schilm et al. [26] (concentrations given the diagram).
boundary phase SiO₂ content are less stable in basic solutions.

Under hydrothermal conditions at temperatures of >200–250 °C the grains and the thin films between the grains are seen to dissolve, resulting in a corrosion layer [59,109,113,126]. Under these conditions the SiO₂-rich grain boundaries are less stable than the Y₂O₃/Al₂O₃ grain boundary phases. The ceramics that are most stable in bases and under hydrothermal conditions, but less stable in acids, hence are those in which the grain boundary phase has a high rare earth ion content. Up to now no silicon nitride ceramics with the same universal corrosion stability as that
of SSiC could be developed. However, ceramics with low amounts of Y2O3/MgO/Al2O3 represent a good compromise with high stability in acids, bases and hydrothermal conditions [59,108–110].

The influence of the grain boundary phase on the corrosion under different conditions is illustrated in Figure 23. Crystallization of the grain boundary can also yield an improvement in corrosion behavior. Very little information about the influence of grain boundary crystallization on corrosion resistance can be found in the literature [103,127]. This could hence be a way to increase the corrosion stability of Si3N4 ceramics. The considerable reduction of the amount of grain boundary phase in α- and β-SiAlON materials also improves the corrosion resistance. These materials are hence additional candidates for further improvement of the corrosion behavior of ball bearing materials.

There is also little data available on the increase in wear due to corrosion, but it seems to be obvious that chemically induced grain boundary damage increases wear because in many of the observed wear mechanisms the wear starts with destruction of the grain boundaries.

The presented corrosion data show the main correlations between composition, structure and corrosion behavior, allowing materials to be preselected for the given conditions. Nevertheless, the materials still have to be evaluated to determine their suitability for the given applications.

### 4.2. Corrosion Stability of Silicon Carbide

The different kinds of silicon carbide ceramics existing show different corrosion behavior. Whereas the solid phase sintered material exhibits very high corrosion stability in different media and at different temperatures, silicon-infiltrated silicon carbide and LPSSiC show lower stabilities (Table 8) due to their compositions and microstructures. The SSiC used in seals and bearings consists of large SiC grains and pure grain boundary phases. Occasionally existing secondary phases such as graphite or boron carbide are normally only present as local inclusions with grain sizes of less than 5–10 μm and therefore corrosion of these phases does not have a weakening effect on the SSiC. The high grain boundary purity and the high corrosion stability in acids, bases and hydrothermal conditions inherent to SiC are the reason for the superior corrosion resistance of these materials in comparison with Si3N4 materials and other structural ceramics. SSiC is the only dense Si-containing ceramic with good stability in HF-containing liquids. SiC can be passivated not only by the formation of a SiO2 layer (which can easily be dissolved in HF-containing solutions), but also by the dissolution of SiO2 and subsequent formation of carbon-containing passivating layers. This explains the high stability in HF-containing solutions and under hydrothermal conditions.

In silicon-infiltrated SiC, Si-infiltrated graphite and LPSSiC ceramics a continuous secondary phase of lower stability determines the overall corrosion resistance of the material.

The first two kinds of materials contain residual silicon which dissolves in basic solutions at higher temperatures. The Si can even decompose water to generate hydrogen.

\[ \text{Si} + (n + 2)\text{H}_2\text{O} \rightarrow \text{SiO}_2 \times n\text{H}_2\text{O} + 2\text{H}_2 \]

The silica formed dissolves in highly basic solutions and therefore no passivation takes place in bases. In contrast, in neutral and acidic solutions passivation takes place due to the slow solubility of SiO2 in the solutions. Because the solubility increases greatly with increasing temperature [131] the corrosion stability decreases strongly with increasing temperature even in neutral solutions.

### TABLE 8 Corrosion Stabilities of Different Kinds of SiC Materials in Different Media [50,59,65–67,93–95,97,128,129]

<table>
<thead>
<tr>
<th>Material class</th>
<th>Acid</th>
<th>Bases</th>
<th>Hydrothermal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiSiC</td>
<td>(+)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td>Si-infiltrated graphite</td>
<td>(+)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td>SiC</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
</tr>
<tr>
<td>LPSSiC</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
</tr>
</tbody>
</table>

The table shows the stabilities of different kinds of SiC materials in different media. The symbols (+) and (-) indicate whether the material is stable or not stable, respectively. The formation of SiO2, carbon-containing layers, and the dissolution of SiO2 and Si are also indicated.
FIGURE 24 Hydrothermal stability of SSiC, LPSSiC (2% Additives) and Si3N4 (Y2O3/Al2O3-additives) at different temperatures [59] (mean corrosion rate calculated on the bases of linear time dependence of corrosion (based on weight loss)).

![Graph showing corrosion rate at different temperatures](image)

FIGURE 25 Formation of pits in LPSSiC ceramics after hydrothermal corrosion in water at 200°C (polished cross section).

![Image showing pitting](image)
The stability of LPSSiC is determined by the oxide grain boundary phases formed. They are nearly fully crystallized in the sintered materials (usually Y₃Al₅O₁₂; Al₂O₃; Y₂Al₂O₅; and/or Y₂Al₂O₉) and are quite stable in acids and bases of medium concentration. However, they can dissolve in concentrated acids and bases at elevated temperatures [128,129]. Under hydrothermal conditions the grain boundary phases are much more stable than the grain boundaries in Si₃N₄ materials, but nevertheless a strong reduction in strength is observed. A detailed analysis has shown that the strength-determining defects are pits formed from residual porosity in the material (Figure 24). Therefore retention of the properties under corrosive conditions is connected with minimization of defects (segregation of additives and pores).

In the work [128,129] the volume of the corrosive solution was quite low in comparison with the sample volume. Therefore the data observed there are valid for concentrated solutions near the solubility limit of the components. In open systems such as pumps and seals higher corrosion rates have to be expected.

Kitaoka [104–106] and Presser et al. [65–67] investigated the acceleration of wear under hydrothermal conditions in detail for SiSiC. They observed that under these conditions thin (tens to hundreds of nm thick) SiO₂ × nH₂O layers are formed. Silica was also found to form in microcracks below the surface. This layer partially dissolved in the media and was observed under mild and severe conditions. However, under severe conditions SiC debris (up to a few hundred nanometers in size) was found in the layer. These particles were loosely bonded by the SiO₂ × nH₂O [65–67].

Kitaoka showed that SiO₂ × nH₂O forms between room temperature and 300 °C in water and nearly completely dissolves in the liquid during corrosion at 120–300 °C. At 120 °C and 300 °C gases were also seen to form. The main gaseous reaction products observed were H₂, CO₂ (25% of the amount of H₂ (mol/mol)), CH₄ (less than 10% of H₂) and, additionally at 300 °C, C₂H₆ (less than 1% of the amount of H₂). In the presence of dissolved oxygen in the water CO₂ and CO were very dominant reaction products.

As already mentioned in the introduction SiC ceramics are semiconductors. The specific resistivity at room temperature can be in the range of more than 1 MΩm to less than 1 Ωm depending on the doping of the grains and the grain boundary phase [54,61–64,93–95]. The reaction of SiC with acids, bases and water is a redox reaction and therefore a chemical potential in contact with water can be formed as in the case of metals (see Figures 25–28; standard potentials are given in [94]). Therefore electrochemical corrosion can cause material degradation. The corrosion rates of SiC ceramics are generally low in comparison with those of metals. Nevertheless, damage patterns for SiC seals have been reported under conditions (pure water at room and elevated temperatures) in which pure chemical corrosion as described above would not be expected [95]. This is strongly suggestive of electrochemical corrosion as the primary cause. The detailed
mechanism underlying this behavior has not yet been elucidated, but a possible reason could be, besides electrochemical corrosion, potential differences caused by a streaming potential. The initial formation of local pits could be enhanced by cavitation.

Studies have been carried out to investigate the electrochemical corrosion behavior of SSiC, SiSiC and LPSSiC [54,94,95,130,132]. SiC ceramics were found to exhibit corrosion potentials (open circuit potentials) of greater than $-150$ mV (in comparison with NHE, normal hydrogen electrode) in 1 N acids and higher than $-200$ mV in 1 N NaOH [94,95,130]. For LPSSiC somewhat lower values were observed ($-450$ mV in 0.5 mol H$_2$SO$_4$ (1 N H$_2$SO$_4$) and $-660$ mV in NaOH [94,95,130,132]). During immersion in the solution these potentials shifted to positive values (up to $+500$ mV in acid and up to 0 mV in 1 M NaOH) due to the formation of surface layers. Through treatment in HF solutions the potential could be returned to the original value [95]. At least in basic and neutral solutions the formation of a SiO$_2 \times n$H$_2$O surface layer was directly shown by SEM/EDX investigations. This SiO$_2 \times n$H$_2$O layer was found to passivate the SiC surface partially and resulted in reduction of the corrosion current [94]. The most pronounced corrosion of SiC ceramics was found to occur if the applied voltage was in the region of water decomposition to oxygen. Under these conditions the formation of CO gas was also observed [132].

Evidence of corrosion is provided by the micrographs in Figure 28. Immersion of SiC in acid without application of a voltage yielded no visible reactions, whereas application of 2 V in H$_2$SO$_4$ resulted in the formation of a 100-nm-thick SiO$_2$ layer. In 1 N NaOH severe etching of the grains took place.

Electrochemical corrosion is greatly reduced through reduction of the conductivity of the SiC ceramics [94]. Due to the dependence of the conductivity of the grains on the doping level (Al, B) in SiC materials core rim structures are observed [130].

**REFERENCES**


Ceramic Bearings and Seals


Pfeifer W. IWM, Freiburg.


[78] VDI-Richtlinie: VDI 2840 Carbon films – basic knowledge, coating types and properties.


Chapter | 3.4 Ceramic Bearings and Seals


Chapter 4.1

Hydrogen-Production Technologies Using Amorphous Silica Membranes

Kazuki Akamatsu and Shin-ichi Nakao

Department of Environmental and Energy Chemistry, Faculty of Engineering, Kogakuin University, 2665-1 Nakano-machi, Hachioji, Tokyo 192-0015, Japan.

Chapter Outline

1. Introduction 331
2. Pore-Size Control of Silica Membranes Prepared by Chemical Vapor Deposition 332
3. Membrane Performance Under Hydrothermal Conditions 334
4. Prediction of the Performance of Membrane Reactors 335
5. Dehydrogenating Organic Hydrides 335
6. Decomposing Hydrogen Sulfide 337
7. Methane Steam Reforming 339
8. Summary 340

References 340

1. INTRODUCTION

The establishment of a hydrogen-based energy system for efficiently producing and transporting high-purity hydrogen is eagerly awaited. For example, hydrogen production using natural energy or renewable energy has attracted much attention, and, at the same time, organic hydrides have become candidates for hydrogen carriers to transport the produced hydrogen to the usage points [1–3]. In this system, it is necessary to develop a technology for producing high-purity hydrogen from the organic hydrides. In addition, hydrogen production by the decomposition of hydrogen sulfide is also a promising technology in petroleum refineries, which has been proposed but is still not in operation even after nearly 30 years [4,5]. Furthermore, there remains room for technological improvements to decrease the reaction temperature in the conventional industrial hydrogen production system by steam reforming of hydrocarbons [6,7]. In general, these reactions are endothermic equilibrium reactions; so, the conversions are limited by the thermodynamics. Therefore, these reactions have to be conducted at higher temperatures to obtain higher conversions. In addition, the produced hydrogen has to be separated and purified from other gases for any applications; thus, some separation and purification processes as post-treatment have to be performed. As one of the key technologies to solve these problems and to achieve the establishment of a hydrogen-based energy system, high-purity hydrogen production using membrane reactors has gained much attention currently.

High-purity hydrogen-production technologies in membrane reactors or catalytic membranes with hydrogen-selective silica membranes have been studied for around 20 years. Figure 1 shows a conceptual illustration of a membrane reactor. A membrane reactor is a system that integrates “reaction” with catalysts and “separation” with membranes. As reaction and separation occur simultaneously in a membrane reactor, an endothermic equilibrium reaction can be shifted forward beyond the thermodynamic equilibrium by extracting one of the generated components, hydrogen in this case, from the reaction system using a membrane. This enables us to achieve higher conversion at lower reaction temperature than the conventional reactor temperature. In addition, it also makes the hydrogen-production system compact because it is not necessary to build additional separation and purification processes for post-treatment. It is not an exaggeration to say that the catalysts for producing hydrogen and the membranes for extracting the produced hydrogen determine the performance of the membrane reactor.
Three types of hydrogen-selective membranes are well known: polymeric membranes based on polyimide [8–10], metallic membranes such as palladium or its alloys [11–13], and ceramic membranes such as amorphous silica [14,15]. However, the polymeric membranes cannot be used at high temperature where hydrogen-production reactions are conducted. For example, hydrogen production using such membrane reactors by dehydrogenating organic chemical hydrides is supposed to be conducted at 500–600 K, and that by decomposing hydrogen sulfide or steam reforming is at around 800–1000 K. Palladium or its alloyed membranes have been extensively studied and their hydrogen-selective performances are excellent because of their unique hydrogen separation mechanism. However, palladium is expensive, and it is apparent that its total reserves are insufficient for such use. On the other hand, ceramic membranes can be used even under such high-temperature conditions and have a cost advantage. Among ceramic membranes, hydrogen-selective amorphous silica membranes have been studied for around 20 years. The silica membranes generally consist of porous supports such as alumina on which amorphous silica as separation layers is deposited. The preparation methods for such silica layers are classified into two types: one is the sol–gel method [16,17] and the other is the chemical vapor deposition (CVD) method [18,19]. In the sol–gel method, colloidal silica sol prepared by hydrolysis and condensation is coated onto the porous support, and silica membranes are obtained after firing. In contrast, in the CVD method, reaction between the silica precursor in saturated vapor and oxygen or ozone takes place and the silica is deposited into the pores of the substrate and onto the porous substrate [20].

The performance of hydrogen-selective silica membranes is evaluated in terms of the hydrogen permeance and the separation factors of hydrogen relative to other gases that are to be separated. In both cases, larger values are preferable. For the past 20 years, much effort has been devoted to the development of hydrogen-selective amorphous silica membranes and there are some reports on excellent performances that matched palladium or its alloyed membranes [21,22]. Unlike metallic membranes, such as palladium, the separation mechanism of the silica membranes is assumed to be a “molecular-sieve,” which means that only the gas species with smaller size than the pore size of the membranes can permeate through the membranes and that the other gas species cannot. In the case of hydrogen-selective silica membranes, amorphous silica networks are assumed to be “pores” that small gas species such as hydrogen can permeate through. At the same time, there exist a few voids, where larger gas species such as nitrogen can permeate through. On reviewing the articles on amorphous silica membranes by many researchers, there are many reports on successful separation of hydrogen from nitrogen, which means that the pore size of the membrane is distributed mainly around 0.3 nm.

When applied to membrane reactors for hydrogen production, a larger hydrogen permeance contributes to a larger equilibrium shift and a larger hydrogen separation factor contributes to higher purity of the obtained hydrogen in the permeate side, in general. Here, it should be noted that the desirable pore size differs depending on the reaction system. In other words, we should take into account the sizes of the gas species that have to be separated from hydrogen. To achieve a larger separation factor, the pore size should be larger than hydrogen and at the same time smaller than the gases that should be separated from hydrogen. Furthermore, to achieve a larger hydrogen permeance, the pore size should be larger. Therefore, a membrane whose pore size is just a little smaller than the sizes of the gas species that should be separated from hydrogen is preferable.

In this review, the strategy for developing silica membranes appropriate for each hydrogen-production system is discussed first, mainly from the viewpoint of “pore-size control.” Then, examples of development of membrane reactors with pore-size-controlled silica membranes are presented: dehydrogenating organic hydrides, decomposing hydrogen sulfide, and methane steam reforming.

2. PORE-SIZE CONTROL OF SILICA MEMBRANES PREPARED BY CHEMICAL VAPOR DEPOSITION

As mentioned above, strict pore-size control is required for each reaction system. For example, when applying the dehydrogenation reaction of organic hydrides, the produced hydrogen (kinetic diameter 0.289 nm) should be separated from cyclohexane (kinetic diameter 0.60 nm) and benzene (kinetic diameter 0.59 nm), or methylcyclohexane (kinetic diameter 0.60 nm) and toluene (kinetic diameter 0.59 nm). Therefore, the pore size should be designed at around 0.5–0.55 nm. On the other hand, when applying to methane steam reforming, the produced hydrogen should be separated from CO₂ (kinetic diameter 0.33 nm), CO (kinetic diameter 0.376 nm), or unreacted methane (kinetic diameter 0.376 nm).
diameter 0.38 nm), which means that the pore size should be designed at around 0.3 nm. However, there are few successful reports on pore-size control of silica membranes prepared by the CVD method because of the difficulty of characterization in the subnanometer scale.

Among these, Sea et al. first reported successful pore-size control using phenyl-substituted ethoxysilanes as precursors [23–25]. They prepared silica membranes with tetraethoxysilane, phenyltrietoxysilane, or diphenyldimethoxysilane. They showed that larger pores were constructed by using the phenyl-substituted ethoxysilanes as precursors. In this way, the methodology of pore-size control by changing the functional groups of the precursor is considered reasonable, but the analysis of the reaction or the mechanism of micropore formation has hardly been discussed.

Ohta et al. employed phenyl-substituted methoxysilanes to prepare amorphous silica membranes by the CVD method. Tetramethoxysilane (TMOS), phenyltrimethoxysilane (PTMS), and dimethoxydiphenylsilane (DMDPS) were used as precursors [26]. The precursors contain zero, one, and two phenyl groups in one molecule, respectively. In the study, it was expected that the existence of the phenyl groups would result in a looser silica network, and larger pores would be built. After optimization of membrane preparation conditions for each precursor, the permeation characteristics through their membranes were compared. Figure 2 shows the permeation results at 573 K of H₂, O₂, N₂, CF₄, and SF₆ through the TMOS-, PTMS-, and DMDPS-derived membranes prepared at 873 K. In every gas species, the permeance was highest through the DMDPS-derived membrane, followed by the PTMS-derived membrane, and then the TMOS-derived membrane. This order coincided with the number of phenyl groups in the precursor. In addition, it is noteworthy that hydrogen permeance through the DMDPS-derived membrane was as high as 1 × 10⁻⁶ mol m⁻² s⁻¹ Pa⁻¹ and this value matches that through palladium or its alloyed membranes. Figure 3(a)–(c) show the temperature dependence of permeation for H₂, O₂, N₂, CF₄, and SF₆ through the TMOS-, PTMS-, or DMDPS-derived membranes. As for hydrogen permeation, the activation energy through the TMOS-derived membrane was 19 kJ mol⁻¹, while it decreased to 11 kJ mol⁻¹ in the case of the PTMS-derived membrane and finally it became 0.8 kJ mol⁻¹ in the case of the DMDPS-derived membrane. These results also indicate that the pore size of the silica membrane increased as the number of phenyl groups in the precursor molecule increased.

To deposit an amorphous silica layer, the silica precursors are assumed to form silanones, and then they have to undergo oligomerization reactions. Quantum chemical calculations

---

**FIGURE 2** Single-component permeances through TMOS-, PTMS-, and DMDPS-derived membranes at 573 K. (from [26]).

**FIGURE 3** Arrhenius plots of single-component permeances through (a) TMOS-derived membrane, (b) PTMS-derived membrane, and (c) DMDPS-derived membrane at 373–573 K. ○, H₂; □, O₂; ◊, N₂; ▽, CF₄; △, SF₆. (from [26]).
were conducted to investigate whether steric hindrance because of the bulky group would occur when using precursors such as PTMS or DMDPS containing the phenyl group. If this is the case, it is clear that pore-size control can be achieved by changing the chemical structure of the precursors. In this study, bond energies of the silica sources were estimated by quantum chemical calculations with the CBS-QB3 method available in Gaussian 03. The results showed a strong tendency for gas-phase conversion of TMOS and PTMS to dimethoxysilanone, (MeO)₂Si=O, and DMDPS to methoxyphenylsilanone, Ph(MeO)Si=O, transient intermediates that are assumed to be major precursors of amorphous silica. This result was, of course, just based on the bonding energy, not based on the kinetic viewpoint. Therefore, it can be speculated that the PTMS-derived membrane was produced from the intermediates of dimethoxysilanone including a small amount of methoxyphenylsilanone, and that the DMDPS-derived membrane was mainly from methoxyphenylsilanone. This can explain the result that the PTMS-derived membrane has somewhat larger micropores than the TMOS-derived membrane and the DMDPS-derived membrane has much larger micropores than the TMOS membrane. Indeed, the ratio of methoxyphenylsilanone to dimethoxysilanone is not clearly understood; however, the phenyl group of the precursors probably resulted in enlargement of pore sizes in the silica membranes.

3. MEMBRANE PERFORMANCE UNDER HYDROTHERMAL CONDITIONS

If we wish to apply appropriate silica membranes to a membrane reactor for hydrogen production, especially for the steam-reforming reaction, a severe problem remains unsolved. Silica membranes are not hydrothermally stable. Exposure to moisture leads to deterioration of the membrane performance, in general. It is often pointed out that this is because of the “densification” of the silica. Essentially, the structure of the silica membrane would change by rearrangement of siloxane bonds or dehydration of silanol groups into siloxane groups under hydrothermal conditions. However, this is just an assumption at this stage and more detailed studies are necessary.

To avoid this deterioration under hydrothermal conditions, several types of metal-doped silica membranes have been investigated. Among them, nickel-doped silica membranes prepared by the sol–gel method were shown to maintain both high hydrogen permeance and high hydrogen selectivity after exposure to steam at 873 K [27].

Besides this, the silica membranes prepared by CVD using TMOS and oxygen at higher temperature were shown to keep both high hydrogen permeance and high hydrogen selectivity under humid conditions. For example, TMOS-derived membranes prepared at 773 K are damaged simply by exposure to air at room temperature. In contrast, membranes prepared at 873 K are stable even under 76 kPa of steam vapor at 773 K for 82 h, as shown in Figure 4 [28]. This is one of the target conditions for applying to membrane reactors for methane steam reforming. Actually, the TMOS-derived membranes were tested at higher temperatures than the preparation temperature, such as 973 K, or 1073 K [29]. The permeation tests revealed that hydrogen permeance decreased slightly at such higher temperatures above the fabrication temperature under dry conditions; however, once the membrane was subjected to higher temperatures, hydrogen permeance did not change even when it was exposed to high temperature again. On the other hand, in humid conditions, a slight deterioration of hydrogen permeance was observed at such higher temperatures. Recently, hexamethyldisiloxane (HMDS) was employed as a silica precursor and its performance under hydrothermal conditions was investigated [30]. The membrane exhibited the same degree of H₂/N₂ selectivity under hydrothermal conditions as the TMOS-derived membrane, and additionally it showed much higher permeance for hydrogen than the TMOS-derived membrane.

In view of these results, the HMDS-derived membranes and the TMOS-derived membranes for use in membrane reactors to deal with reaction systems containing H₂O have to be prepared at a higher temperature than the application temperature. However, the reason why these membranes prepared at higher temperature are stable under hydrothermal conditions is not clear at this stage, even though the microporous structure of the membranes is assumed to be exerting some influence.

![Figure 4](from [28]).
4. PREDICTION OF THE PERFORMANCE OF MEMBRANE REACTORS

Prediction of the performance of a membrane reactor is one of the most important points. To do this, the simplest 1-D model is constructed just by considering permeations through the membrane and reactions. In other words, mass balance is just applied to reaction and permeation. However, hydrogen-production reactions are endothermic, in general. Therefore, the method of heat supply for the enthalpy of the reactions is very important for the development of membrane reactors. In addition, these enthalpy changes are relatively large in most cases; so, it is often pointed out that the large enthalpy changes result in inhomogeneous temperature distribution profiles in the reactors. Furthermore, these models cannot be used to discuss 3-D complicated configurations of the membrane reactors. There are some reports on the prediction of the performance of membrane reactors using computational fluid dynamics to consider heat profiles in the membrane reactors, which is one of the crucial factors for determining the reactivity [31]. Such detailed analysis of the temperature profiles enables us to design membrane reactors where heat supply can be achieved effectively.

5. DEHYDROGENATING ORGANIC HYDRIDES

Organic hydrides are one of the promising candidates for hydrogen carriers because of their high storage capacities and easy handling. Representative examples of organic hydrides are as follows:

\[
\begin{align*}
\text{C}_6\text{H}_{12} & \rightleftharpoons 3\text{H}_2 + \text{C}_6\text{H}_6 \quad \text{(Cyclohexane – Benzene system)}, \\
\text{C}_6\text{H}_{11} – \text{CH}_3 & \rightleftharpoons 3\text{H}_2 + \text{C}_6\text{H}_5 – \text{CH}_3 \quad \text{(Methylcyclohexane – Toluene system)}, \\
\text{C}_{10}\text{H}_{18} & \rightleftharpoons 5\text{H}_2 + \text{C}_{10}\text{H}_8 \quad \text{(Decalin – Naphthalene system)}.
\end{align*}
\]

According to thermodynamics, these reactions should be conducted at 473–573 K. For palladium or its alloyed membrane, hydrogen brittleness emerges as a problem in this temperature range. Therefore, a ceramic membrane is preferred. From the viewpoint of equilibrium conversions, the dehydrogenation of decalin occurs easily, followed by methylcyclohexane, then cyclohexane. In addition, hydrogen-storage efficiency is the highest in the case of decalin, and meets the target of the US Department of Energy hydrogen plan. However, in the decalin–naphthalene reaction system, naphthalene is in the solid state at ordinary temperatures and pressures, which is not desirable. As mentioned in Section 2, the DMDPS-derived membrane is favorable for membrane reactors to dehydrogenate cyclohexane or methylcyclohexane because of its larger pore size. Figure 5 shows the relationships between the reaction temperature and the conversion of cyclohexane at 0.1 MPa when a nonporous tube, a TMOS-derived membrane, or a DMDPS-derived membrane was used [32]. The reactor with a nonporous tube would act as a packed-bed reactor because there is no extraction of hydrogen. In using the nonporous tube, the conversions agreed well with the equilibrium values at all temperature conditions. On the other hand, in using the TMOS-derived membrane, the conversions were above the equilibrium values. When the DMDPS-derived membrane was used, the conversions were much higher than those when the TMOS-derived membrane was used. It should be noted that this order coincides with the order of hydrogen permeance. This means that larger hydrogen permeance resulted in larger equilibrium shifts, and this can also be explained by using the simple reactor model, as described in Section 4. Figure 6 shows the relationships between the reaction temperature and the conversion of methylcyclohexane at 0.1 MPa when a nonporous tube or a DMDPS-derived membrane was used [33]. In the case of the methylcyclohexane–toluene system, a similar

![Graph showing the relationship between reaction temperature and conversion](image-url)
tendency was observed. Figure 7 shows the relationships between the reaction pressure and the conversion of methylcyclohexane at 533 K when a nonporous tube or a DMDPS-derived membrane was used. This dehydrogenation reaction generates 3 mol of H₂ and 1 mol of toluene from 1 mol of methylcyclohexane. Therefore, higher pressure is unfavorable to achieve higher conversion from the viewpoint of thermodynamics. On the other hand, extraction of hydrogen in a membrane reactor, which results in an equilibrium shift, is effective when there is a larger partial pressure difference in hydrogen. Thus, higher pressure is favorable to achieve higher conversion from the viewpoint of the driving force of permeation in membrane reactors. In the case of using a nonporous tube, conversion of methylcyclohexane decreased as reaction pressure increased. This is because there is no hydrogen extraction in this case, and the dehydrogenation reaction in this reactor was subject only to thermodynamics. When using the DMDPS-derived membrane, the conversion of methylcyclohexane exceeded the equilibrium value, and the conversion was virtually constant at pressures ranging from 0.1 to 0.25 MPa. Its value was as high as 0.99. This means that the unfavorable effect of thermodynamics was completely compensated by the larger hydrogen extraction because of the higher pressure.

Actually, these experiments were conducted using carrier gas and sweep gas to demonstrate the effectiveness of the membrane reactor with hydrogen-selective silica membranes. However, these operations with carrier gas and sweep gas are not realistic for practical use because the hydrogen should be produced with high purity. If the hydrogen is produced using carrier gas and sweep gas, the produced hydrogen will contain a lot of carrier gas and sweep gas and some post-treatment to purify the hydrogen would be required. Therefore, the neat organic hydride should be supplied in a pressurized condition, and the membrane reactors should be operated without carrier gas or sweep gas. This operation can produce high-purity hydrogen at atmospheric pressure, and it can be directly supplied to any application, for example, to fuel cells. Figure 8 shows an example of such an operation. The time course for cyclohexane conversion measured by gas chromatography, reaction pressure, hydrogen purity, production rate of H₂, and permeance of H₂ before and after each day’s run is shown [34]. For the first day’s run, the conversion and hydrogen purity were around 0.8 and 99.9%, respectively, and these values were stable for 6 h. In addition, this performance was also reproduced for the second, third, seventh, and eleventh day’s runs. Mass-spectrometric measurement of the gas on the permeate side, diluted with nitrogen, showed that a small amount of benzene and cyclohexane was obtained. The concentrations of benzene and cyclohexane in the permeate gas were estimated to be 0.118% and 0.043%, respectively. Consequently, hydrogen with high purity at 1 atm was stably obtained in the permeate side during this test. Actually, it was reported that the permeate gas produced by the membrane reactor was directly
supplied to a commercial fuel cell, and stable electric power generation with output greater than 4 V was successfully observed. In Figure 9, results in the methylcyclohexane–toluene system are also shown [33]. For the first day’s run, the conversion and hydrogen purity were around 0.8 and 99.95%, respectively. The performance of the membrane reactor was stable during the period of operation.

The application of membrane reactors with high-performance hydrogen-selective amorphous silica membranes to dehydrogenation of organic hydrides is one of the successful examples of hydrogen production.

6. DECOMPOSING HYDROGEN SULFIDE

Hydrogen sulfide is produced as a by-product of purification processes in petroleum refineries that remove sulfur by using an enormous amount of hydrogen, or it can be present in natural gas. For removing this, the conventional industrial treatment method is the Claus process. This process turns hydrogen sulfide into water and sulfur, which results in the consumption of a large amount of hydrogen. If hydrogen sulfide could be decomposed to sulfur and hydrogen as shown below, this hydrogen could be reused for the removal of more sulfur, or used for other purposes:

\[ \text{nH}_2\text{S} \rightarrow \text{nH}_2 + \text{S}_n \ (n = 1 - 8). \]

A membrane reactor to decompose hydrogen sulfide for producing hydrogen can achieve the goal. In addition, the equilibrium conversion for this decomposition reaction is extremely low, and, for example, the conversion cannot exceed 20% below 1200 K. This also makes membrane reactors superior to conventional reactors. Actually, the application of membrane reactors with hydrogen-selective membranes has been proposed for no less than 30 years; however, no membrane reactor for decomposing hydrogen sulfide has ever been developed successfully.
The kinetic diameter of hydrogen sulfide is almost the same as that of nitrogen. In addition, it was proved that the membrane reactor should be operated above 873 K to obtain higher conversion. Therefore, the use of TMOS-derived membranes for developing membrane reactors is preferable because these membranes can stably separate $\text{H}_2/\text{N}_2$ at such higher temperatures. Figure 10 shows the relationship between the residence time and conversion achieved by a membrane reactor with a TMOS-derived membrane at 873 K [35]. The equilibrium conversion at this reaction condition was only 7.8%. On the other hand, the membrane reactor exceeded equilibrium values at every residence time from 1 s to 7 s. The achieved conversion increased with the residence time, and it reached as high as 69% at a residence time of 7 s in the experimental conditions. This is of course because the TMOS-derived membrane had both sufficient permselectivity for hydrogen/hydrogen sulfide and a large permeance for hydrogen. This effect of hydrogen-selective extraction resulted in the shift of the equilibrium conversion. Figure 11 shows the hydrogen permeance before and after the decomposition reaction test. The hydrogen permeance changed only slightly: at 873 K, it changed from $1.4 \times 10^{-7}$ mol m$^{-2}$ s$^{-1}$ Pa$^{-1}$ to $1.0 \times 10^{-7}$ mol m$^{-2}$ s$^{-1}$ Pa$^{-1}$. This shows that the excellent performance of the membrane

**FIGURE 9** Course of the production of hydrogen from methylcyclohexane in the membrane reactor with the DMDPS-derived membrane. Changes in methylcyclohexane conversion, reaction pressure, hydrogen purity, production rate of hydrogen, and hydrogen permeance for the DMDPS-derived membrane during the dehydrogenation of methylcyclohexane are shown. (from [33]).
reactor is based on the high tolerance of the TMOS-derived membrane to sulfur-containing atmospheres and to the high temperature. Indeed, at this stage, no long-term durability test has been conducted yet. Therefore, the durability of the hydrogen-selective membrane or the membrane reactor under such a crude condition should be studied in detail.

7. METHANE STEAM REFORMING

Steam reforming of hydrocarbons is one of the best-established methods for industrial hydrogen production. Among them, the methane steam-reforming reaction has been used conventionally in large-scale hydrogen plants. This can be divided into the following two reactions:

$$\begin{align*}
\text{CH}_4 + \text{H}_2\text{O} & \leftrightarrow 3\text{H}_2 + \text{CO}, \\
\text{CO} + \text{H}_2\text{O} & \leftrightarrow \text{H}_2 + \text{CO}_2.
\end{align*}$$

The steam-reforming reaction is generally performed at around 1150 K because of the endothermic character of the first reaction, so the reaction temperature should be decreased to more moderate conditions. For this purpose, membrane reactors with hydrogen-selective membranes are promising. Another issue for applying silica membranes to the membrane reactors for steam reforming is the characteristics of steam durability of the membranes, as described in Section 3. Therefore, a membrane with steam stability is desirable.

Figure 12 shows the relationship between the reaction pressures and the conversions of methane at 773 K when a TMOS-derived membrane or an HMDS-derived membrane was used [30]. Just as with the reaction for dehydrogenating organic hydrides, a higher pressure is not preferable from the viewpoint of thermodynamic equilibrium. On the other hand, a membrane reactor
using the TMOS-derived membrane showed an equilibrium shift at every pressure condition because of hydrogen extraction. Furthermore, in the case of a membrane reactor using the HMDS-derived membrane, the conversion was much higher than that in the case of the TMOS-derived membrane. The order of the equilibrium shifts also coincided with the order of hydrogen permeance.

To study the stability of the membrane reactor with the HMDS-derived membrane, which showed a larger equilibrium shift, continuous operation of the membrane reactor was conducted to produce H₂ for 30 h. The pressure on the reaction side was just 0.1 MPa, and the S/C ratio was only 2.5. The time course of the conversion of methane is shown in Figure 13 [30]. A methane conversion as high as 0.8 was obtained, and this equilibrium shift was stable for 30 h. No sudden decrease in the performance of the membrane reactor was observed. This result showed that the HMDS-derived membrane stably extracted the H₂ and, at the same time, the catalyst also worked stably for this period of time.

8. SUMMARY

In this review, efficient high-purity hydrogen-production technology using membrane reactors was described as one of the promising technologies using ceramic membranes, in particular, using hydrogen-selective amorphous silica membranes. The membrane reactors consisting of amorphous silica membranes for hydrogen separation and catalysts for hydrogen production can extract hydrogen as soon as produced, which results in favorable equilibrium shifts when applied to equilibrium reactions such as dehydrogenation of organic hydrides, decomposition of hydrogen sulfide, or steam reforming of hydrocarbons. In addition, this efficient system also makes it possible to achieve higher conversions under lower reaction temperature conditions, compared with conventional reactors. To develop such membrane reactors, a high-performance hydrogen-selective membrane that can be used in the reaction atmospheres is one of the key factors, and silica membranes can meet this condition. Today, fundamental reports on developing membrane reactors with ceramic membranes are increasing in number. However, the study of long-term durability has been scarcely discussed. In addition, the technology to develop larger-size reactors, especially for large-area membranes, should also be developed for practical use. Consistent approaches to these issues will help “Hydrogen Production Technologies Using Ceramic Membranes” edge closer to practical use.

REFERENCES

Hydrogen-Production Technologies Using Amorphous Silica Membranes


All-Solid-State Li Battery for Future Energy Technology

Masashi Kotobuki, Hirokazu Munakata and Kiyoshi Kanamura

Department of Applied Chemistry, Graduate School of Urban Environmental Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

Chapter Outline

1. Introduction 343
2. Honeycomb-Type 3D Battery 343
3. All-Solid-State Battery With 3DOM Structure 346
4. Summary 350

1. INTRODUCTION

Rechargeable lithium ion batteries have been widely used as a power source for portable electronic devices such as mobile phones, laptop computers, and digital cameras because of their high operation voltage, high energy density, and no memory effect [1]. Recently, their application to electric vehicles has been accelerated; however, the safety reason caused by flammable organic liquid electrolyte is preventing them from the application. All-solid-state battery composed of nonflammable solid electrodes and solid electrolyte has been recognized as a battery with ultimate safety [2]. Hence, both polymer and ceramic solid electrolytes have been studied [3–5]. Particularly, ceramic electrolytes have been given much attention because of their high durability against high temperature. Among them, Li0.35La0.55TiO3 (LLT) with perovskite structure [6,7] and Li1+xAlxTi2−x (PO4)3 (LATP) with NASICON-type structure [8–12] have been given much attention owing to their high Li ion conductivity of approximately 10−3–10−4 S cm−1, which is acceptable for practical all-solid-state battery.

One of the critical issues of the all-solid-state battery is poor contact between solid electrode and solid electrolyte compared with the present liquid electrolyte (Figure 1). This poor contact leads to high internal resistance, resulting in low performance [13]. Hence, 3-dimensional (3D) battery has been considered to improve the performance [14–16]. Several prospective configurations of 3D batteries have been suggested (Figure 2) [17]. All of these configurations can provide not only large contact between solid electrode and solid electrolyte, but also high loading of active materials and short Li ion transport distance, which improves energy density and leads to high power density, respectively.

In this section, the 3D all-solid-state Li batteries with honeycomb-type and 3-dimensionally ordered macro-porous (3DOM) structures are described.

2. HONEYCOMB-TYPE 3D BATTERY

The honeycomb structure is constructed by microsized holes on both sides of membrane (Figure 3) [18]. The all-solid-state battery can be constructed by filling the holes with electrode materials. It seems that the all-solid-state battery with the honeycomb structure can have enough high capacity due to high loading of active material in the
honeycomb holes and high power density by short Li ion path in thin honeycomb wall. In this section, the fabrication of the all-solid-state battery using honeycomb-type Li$_{0.35}$La$_{0.55}$TiO$_3$ (LLT) electrolyte is introduced.

Figure 4 shows SEM images of honeycomb LLT electrolyte [19]. The hole size is 180 $\times$ 180 $\mu$m and thickness of walls between the holes is 80 $\mu$m. This configuration is limited by the present ceramics technique.

By injection of active material into the honeycomb holes, the all-solid-state battery can be fabricated. However, the injection is not easy; some specialized technique is required. Figure 5(a) displays a cross-sectional SEM image of honeycomb electrolyte after impregnation of ethanol suspension containing LiCoO$_2$ followed by calcination. This normal impregnation technique produced a large space between electrode and honeycomb electrolyte by shrinkage of the LiCoO$_2$ particle during the calcination. This space decreases the electrode–electrolyte contact area and causes an increase of internal resistance of the cell. The space can be reduced by application of the sol–gel method. Figure 5(b) shows a cross-sectional SEM image of honeycomb electrolyte after impregnation of a mixture of LiCoO$_2$ with its precursor sol. The contact between active material and electrolyte was much better. The impedance spectra of both LiCoO$_2$/honeycomb LLT cells are shown in Figure 6 [19]. It is clearly observed that the internal resistance was reduced by increase of electrode–electrolyte contact area. The charge and discharge curves of both LiCoO$_2$/honeycomb LLT half-cells are shown in Figure 7. The cell impregnated only LiCoO$_2$ particles showed a very low performance, whereas the cell prepared with precursor sol exhibited a much higher performance. As described, the sol-gel method is very useful for the formation of good electrode–electrolyte contact; however, the precursor sol composition must be decided carefully. Figure 8 shows cross-sectional SEM images of LiMn$_2$O$_4$/honeycomb LLT electrolyte prepared by acetate and nitrate sols [20]. The interface at LiMn$_2$O$_4$/honeycomb LLT is completely different depending on the sol composition. Some space is observed at the interface prepared by the acetate sol. The actual contact area, i.e., electrochemical active area, seems much smaller than the geometric one of the honeycomb hole. In the case of nitrate sol, the boundaries between LiMn$_2$O$_4$ and honeycomb LLT are quite indistinct. LiMn$_2$O$_4$ is coated on the whole wall of honeycomb holes and no space is found. In fact, smaller impedance and larger discharge capacity of LiMn$_2$O$_4$/honeycomb LLT prepared by the nitrate sol due to large contact area were observed (Figures 9 and 10).
Additionally, selection of appropriate cathode material is also critical for fabrication of the all-solid-state battery [21]. Figure 11 shows XRD patterns of honeycomb LLT electrolyte impregnated by LiCoO₂ and LiMn₂O₄ cathodes. An impurity phase of La₂Ti₂O₇ was produced at interface between LiCoO₂ and LLT. On the other hand, no impurity phase was observed at the interface of LiMn₂O₄/LLT. The impedance spectra of LiCoO₂/honeycomb LLT and LiMn₂O₄/honeycomb LLT cells are depicted in Figure 12. The internal resistance of LiMn₂O₄/honeycomb LLT was much smaller than that of LiCoO₂/honeycomb LLT because the impurity phase would act as resistance layer and block smooth Li ion transfer between electrode and electrolyte.
As mentioned above, the electrode—electrolyte interface governs battery performance. A choice of appropriate condition to provide the best contact between electrode and electrolyte is very important to fabricate all-solid-state battery.

3. ALL-SOLID-STATE BATTERY WITH 3DOM STRUCTURE

Three-dimensionally ordered macroporous (3DOM) structure is one of the ideal structures due to its high porosity and regularity providing large contact area between electrode and electrolyte and uniform current distribution, respectively. The 3DOM solid electrolyte is prepared and, then by injection of active material into the pores of 3DOM electrolyte, active material—3DOM composite electrode can be prepared. This active material—3DOM composite electrode can provide large contact area between solid electrode and solid electrolyte (Figure 13) [22]. Herein, the active material—3DOM composite electrode using LATP electrolyte is described.

The $\text{Li}_{1+x}\text{Al}_x\text{Ti}_{2-x} \text{(PO}_4)_3$ (LATP) with NASICON-type structure has been thought to be one of the promising solid electrolytes due to its high Li ion conductivity ($\sim 10^{-3}$ S cm$^{-1}$). The 3DOM solid electrolyte can be prepared by suspension filtration method. A
monodispersed polystyrene (PS) and LATP powder were suspended in water and then filtered under a small pressure difference of \( -5 \text{ kPa} \). Obtained composite membrane consisting of PS and LATP was heated to remove PS template and crystallize LATP. Figure 14 shows a cross-sectional SEM image of the obtained LATP membrane. The interconnecting window was observed on an entire part of the LATP membrane. The macropore size was 2 \( \mu \text{m} \), which was smaller than that of PS template (3 \( \mu \text{m} \)), indicating that the shrinkage of the 3DOM LATP took place during calcination. The suspension filtration is strongly affected by size of LATP powder and PS beads. These sizes must be controlled carefully.

The 3DOM LATP was immersed in the precursor sol for LiMn\(_2\)O\(_4\) under vacuum condition followed by calcination. By comparing the SEM image of the composite electrode system (Figure 15) with that of 3DOM LATP (Figure 14), it was found that LiMn\(_2\)O\(_4\) was injected into the pore of 3DOM LATP. However, some voids were still observed in the composite electrode. Figure 16 shows the charge and discharge curves of the composite electrode system. The potential plateau was observed at around 4.0 V. This plateau was due to the redox reaction of Mn\(^{3+/4+}\) in LiMn\(_2\)O\(_4\), indicating that Li intercalation/extraction into/from LiMn\(_2\)O\(_4\) occurred in the composite electrode system, which was produced only ceramics. The first discharge capacity was 64 mA h g\(^{-1}\).

There is another way to prepare 3DOM solid electrolyte, called “colloidal crystal templating method.” A preparation of 3DOM LLT solid electrolyte using the colloidal crystal templating method is introduced [23].
The LLT with perovskite structure also has been expected to be a good candidate of solid electrolyte for all-solid-state battery application. 3DOM LLT can be prepared by colloidal crystal templating method combined with the sol-gel method. The PS beads suspended in water were filtrated and interconnected by heat treatment. The precursor sol for LLT was injected into voids of the PS template using vacuum impregnation. After gelation, the composite of ordered PS template and precursor gel for LLT was calcined. During the calcination, the PS template was burned and 3DOM pores were formed. On the other hand, the precursor gel was converted to well-sintered LLT. A cross-sectional SEM image of 3DOM LLT is shown in Figure 17. 3DOM structure can be successfully prepared and the macropore size was ca. 2 μm. Interconnecting windows were also observed on an entire part of LLT membrane. LiMn$_2$O$_4$ was impregnated by “solvent substitution method.” At first, the pores of LLT membrane were filled with sodium dodecyl sulfate (SDS) solution. The solution can enter into the pores easily by low surface tension. Then, the 3DOM LLT filled with SDS solution was immersed in the precursor sol for LiMn$_2$O$_4$. The SDS solution was substituted by the precursor sol during immersion. Figure 18 shows a cross-sectional image of 3DOM LLT after injection and gelation of precursor sol for LiMn$_2$O$_4$. In Figure 18, the bright area with circle shape corresponds to 3DOM LLT and the gray color area is the gel. It can be seen that the pores of 3DOM LLT are almost completely filled up with the gel. The gel was converted to LiMn$_2$O$_4$ following calcination [23]. Figure 19 shows cross-sectional SEM images of composite electrode system between 3DOM LLT and LiMn$_2$O$_4$ after the first and fourth sol injection. By comparing the SEM image of composite with that of 3DOM LLT, it was found that LiMn$_2$O$_4$ was injected into the pores of 3DOM LLT. Although 3DOM LLT was filled up with the precursor sol, voids were still observed after heat treatment. This is due to the large amount of solvent in the precursor sol. The voids destroy electronic conduction path in the composite electrode; therefore, the sol composition should be determined prudently to minimize void formation. The filling ratio of the active material was 13% estimated from the weight of the 3DOM LLT membrane before and after immersion. In order to increase the filling ratio of active material, immersion and calcination were repeated four times and the filling ratio reached 83%.

Figure 20 demonstrates charge and discharge curves of the LiMn$_2$O$_4$—3DOM LLT composite electrode. The potential plateaus due to redox reaction of Mn$^{3+/4+}$ in LiMn$_2$O$_4$ were observed at 4.0 V vs. Li/Li$^+$. The result indicates that Li deintercalation/intercalation from/into LiMn$_2$O$_4$ takes place at the interface of the LiMn$_2$O$_4$—LLT solid electrolyte. The first discharge capacity was
220 mA h cm$^{-3}$, which was comparable to a volumetric capacity of LiMn$_2$O$_4$ electrode with liquid electrolyte.

A fabrication and electrochemical property of the active material—3DOM composite were introduced. However, only the 3DOM composite electrode cannot fabricate the all-solid-state battery. One of the promising structures using the 3DOM electrode is a 2-layered structure composed of 3DOM layer and dense layer (Figure 21). The dense layer works as both an electrolyte and a separator and the 3DOM layer is utilized as the composite electrode. This 2-layered configuration allows easy fabrication of the all-solid-state battery. In this section, fabrication and electrochemical property of the all-solid-state battery with the 2-layered structure using LLT solid electrolyte are described.

The 2-layered LLT was prepared by the suspension filtration method. An ethanol suspension composed of LLT powder and PS beads was filtered to prepare the porous layer and then LLT powder dispersed in ethanol was filtered for the dense layer. The deposition on the filter was calcined to obtain the 2-layered LLT. Cross-sectional SEM images of the 2-layered LLT are shown in Figure 22. Thicknesses of dense and porous layers are 17 and 111 μm, respectively. These thicknesses can be controlled by the amount of suspension. No space was observed at the interface between dense and porous layers, which allows the smooth movement of Li ion from the porous layer to the dense layer or vice versa. The porous layer has numerous pores interconnected to each other. By solvent substitution method, LiMn$_2$O$_4$ can be impregnated into the
3DOM pores. Figure 23 shows cross-sectional SEM images of the 2-layered LLT with LiMn₂O₄. LiMn₂O₄ particles in the 3DOM pores were clearly observed. However, as shown in the magnified image of the porous layer near the dense layer, some voids still existed in the bottom of the porous layer, while the porous layer near the surface was filled up well. A volume fraction of LiMn₂O₄ against pore volume was about 40%. Figure 24 exhibits charge and discharge curves of the 2-layered LLT with LiMn₂O₄. Charge and discharge behavior is distinctively confirmed, attesting that the LiMn₂O₄/2-layered LLT works as all-solid-state battery. Two potential plateaus were clearly observed at 4.0, 4.1 vs. Li/Li⁺, which corresponds to the intrinsic redox potential of Mn³⁺/⁴⁺ in LiMn₂O₄. No large degradation was observed during five cycles. The discharge capacity was 27 mA h g⁻¹ in the first cycle which is 19% of theoretical capacity. The reason for this low utilization is attributed to low filling ratio of active material which causes a collapse of electronic conduction path in the electrode.

4. SUMMARY

3D batteries with honeycomb structure and 3DOM structure were described. For practical use of all-solid-state battery, development of not only materials but also 3D configuration is needed. Therefore, development of all-solid-state battery requires a different concept from that of conventional battery. R&D of process technique would be more important. The numerous 3D battery configurations can be proposed. An appearance of novel 3D configuration, which can break the conventional battery concept, is anticipated.

REFERENCES


[10] Xu X, Wen Z, Gu Z, Xu Z, Lin Z. Lithium ion conductive glass ceramics in the system Li1.2Al0.4(Ge1-xTix)1.6(PO4)3 (x = 0–1.0). Solid State Ionics 2004;171:207–13.


Advanced Ceramics for Nuclear Applications

Toyohiko Yano* and Branko Matović**

*Research Laboratory for Nuclear Reactors, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan, **Vinca Institute of Nuclear Sciences, University of Belgrade, 11001 Belgrade, Serbia, Former Affiliate: Research Laboratory of Nuclear Reactor, Tokyo Institute of Technology

Chapter Outline

1. Introduction 353
2. Basics of Neutron Irradiation Effects with Matter 353
3. Radiation Damage of Advanced Ceramics 354
4. Advanced Ceramics for Fission Reactors and Fuel Cycle 362
5. Advanced Ceramics for Fusion Reactors 366
6. Conclusion 367
References 367

1. INTRODUCTION

Up to now more than 440 commercial nuclear power plants are in operation and more than 70 are under construction or planning to be built. Nuclear energy is supplying around 6% of primary energy and more than 15% of electricity in the world (as of 2008). Within many global claims including growing world population/food production, shortage of fossil resources, and global warming due to greenhouse gases, it is clear that energy and global environmental issues are deeply related with each other. We must find a better solution based on the idea for maintaining human security of all people on the Earth continuously for the future. How to supply the world’s growing energy demands with environmentally sustainable sources has focused renewed attention on nuclear power as an attractive option. Whereas miserable application of the atomic energy was assigned at the beginning of its history as wartime mission, it is unquestionable that nuclear energy can provide major possibilities to rescue humankind from the global claims mentioned above.

Nuclear power plants need various kinds of materials, including ceramics, to complete their mission such as electric power generation and high-temperature heat supply. From nuclear fuels to high-level nuclear waste confinements, various kinds of ceramic materials are necessary components of fission nuclear fuel cycle. In fusion reactors, wide variety of ceramics will be required to sustain fusion nuclear fuel cycle. In this chapter, basic material property change in the field of nuclear application is first mentioned and then some applications of advanced ceramics for fission and fusion reactors are described.

2. BASICS OF NEUTRON IRRADIATION EFFECTS WITH MATTER

Inside the pressure vessel of nuclear reactors, materials suffer from irradiation of neutrons which are generated by the fission reaction of nuclear fuels, mainly $^{235}\text{U}$ and $^{239}\text{Pu}$. Nuclear fission initiates by absorption of a neutron into a nucleus of fissionable isotope, and it sprits into fission fragments with mass around ~90 and ~140 amu (atomic mass unit). At the same time a couple of neutrons with high kinetic energy of around 2.0 MeV are also emitted (1 eV = $1.6 \times 10^{-19}$ J). These neutrons will interact with matters on their pathway, gradually losing their energy by collisions, and then a part of neutrons will be absorbed with another fissionable nuclei and induced fission chain reaction.

During the course of interactions, some neutrons are absorbed into nuclei, and excites its energy level up to an unstable level, leading gamma ray emission (n, $\gamma$) reaction, or another nuclear absorption reactions such as (n, p) and (n, $\alpha$) reactions. Neutron will change to proton (p), electron, and neutrino ($\nu$):

$$n \rightarrow p + e^- + \nu$$

Beta ray ($e^-$) emission reaction makes an increase in atomic number of the target atom. On the other hand, (n, p) reaction makes a decrease in atomic number of the target atom. These are kinds of transmutation of atoms. One of the neutron irradiation effects on matters is “transmutation” mentioned above. Emitted proton itself catches an electron and changes to a hydrogen atom at the same time. The (n, $\alpha$) reaction means absorption of neutron induces...
emission of an alpha particle, which is a nuclei of helium (atomic number 2) atom. Therefore, as a result of the \(\text{(n, } \alpha\text{)}\) reaction, the target atom will be transmuted into the new atom with protons less two numbers than the target atom, for example, boron transmuted into lithium. Alpha particles obtain electrons and will become helium atoms:

\[
\text{\text{^{10}B} + \frac{1}{2}\text{n} \rightarrow \text{^{7}Li} + \frac{1}{2}\text{He}}
\]

The possibility of nuclear reactions is measured for each isotope, and given as “cross section” which is evaluated and listed in the nuclear data libraries such as JENDL [1] or ENDF [2]. Cross sections of each isotope, not atom, depend on energy of neutrons. It is noted that “transmutation” effects are permanent effects on materials. Furthermore, many nuclear reactions have threshold energy where the reaction will take place above that neutron energy. Generally, \(\text{(n, p)}\) or \(\text{(n, } \alpha\text{)}\) reactions of many isotopes occur only by neutrons with energies more than 1 MeV. Application for fission reactors, neutron absorption reactions such as \(\text{(n, p)}\) or \(\text{(n, } \alpha\text{)}\) reactions are induced only for limited isotopes. In the case of application for fusion reactors, these reactions are induced for many isotopes due to high-energy of fusion neutrons, i.e., 14 MeV. Transmutation induced impurities in materials. Most of metallic impurities cause not so strong influences on the materials properties, except for electric and optic properties, since commercial materials are always not so pure and originally contain several kinds of impurities. Slight changes in amount and kind of impurities do not induce obvious effects on the other properties. Special caution should be paid on the formation of gaseous elements such as helium due to \(\text{(n, } \alpha\text{)}\) reactions. Helium easily coalesces into bubbles inside grains and along grain boundaries.

The other neutron irradiation effect on matters is “knock-on” of atoms. Neutrons with kinetic energy of more than the threshold displacement energy of atoms can knock-on atoms by collision. Since neutron has no electric charge and has heavy mass, more than a thousand of that of electron, probability of the collision with matter (cross section) is relatively small. On the other hand, energy transfer is bigger once it collides with an atom. Original kinetic energy of fission neutrons is around 2.0 MeV, and typical threshold displacement energy of solid matters is in the range of 10–150 eV. Based on the simple knock-on model, one energetic neutron can induce displacements in the range of a few hundred to thousand of atoms in solids, larger in target with lower mass number. In contrast to transmutation effects, all materials suffer from knock-on effects by neutrons with higher energy than threshold displacement energies without exception.

Displacement of atoms in crystalline solids, i.e., most of metals and ceramics, induces several kinds of crystalline defects. Primary defects are vacancies and interstitial atoms directly formed by the knock-on of lattice atoms. A pair of vacancy and interstitial atom is named the Frenkel defect. The presence of point defects causes deformation of crystalline lattice. Most of Frenkel defects are annihilated immediately, but a small number of each point defect survive longer. If the temperature of materials during neutron irradiation is relatively lower, both defects are not mobile and they are accumulated during irradiation. If neutron doses are exceeded by some critical value, crystals cannot keep crystallinity and then amorphize. Under medium temperature range, only interstitial atoms are able to migrate. Most of self-interstitials will recombine with vacancies to annihilate defects. Otherwise, some interstitials will come together at a specific place and form extra plane(s) of atoms between existed crystalline planes to reduce total energy of crystals. Shape of the extra plane is generally round plate-like. It is called interstitial dislocation loops. Under relatively higher temperature, both vacancies and interstitial atoms can migrate to form dislocation loops. In general, the number of single vacancies exceeds the number of single interstitial atoms at that condition; vacancies migrate and form voids. If there are gaseous atoms simultaneously, they will be trapped into vacancies and will migrate together with vacancies to form gas bubbles, both inside grains and along grain boundaries. Accumulation of crystalline defects is strongly related with irradiation condition, such as neutron flux, total neutron dose (fluence), or temperature during irradiation, since migration/diffusion capability of initial defects depends strongly on temperature, faster at higher temperature. As a result in general, the amount of accumulated point defects should be less under high-temperature irradiation condition. On the other hand, defect cluster formation is progressed.

Both crystalline defects and transmutation atoms cause permanent property changes of materials. In many cases, material properties will be degraded; therefore, it is called “radiation damage”.

3. RADIATION DAMAGE OF ADVANCED CERAMICS

Due to accumulation of crystalline defects and transmuted gaseous atoms, several properties of almost all ceramics are degraded beyond some doses, depending mainly on irradiation temperature.

Tables 1 and 2 are examples of property changes of typical advanced ceramics caused by severe fast neutron irradiation, corresponding to the fluence of core materials inside fast reactors for relatively long operation period [3]. Amount of linear swelling of \(\text{Al}_2\text{O}_3\) and \(\text{AlN}\) ceramics are relatively bigger than those of \(\text{SiC}\) and \(\text{Si}_3\text{N}_4\) ceramics. Linear swelling, more than 1%, indicates microcrack formation or void-swelling, so that mechanical integrity of these materials degrades severely. Covalent-bonded \(\text{SiC}\)
and Si$_3$N$_4$ ceramics show excellent tolerance as compared to the compounds with more ionic bonding nature. Table 2 indicates change in thermal diffusivity of the same materials. It is observed that degradation of thermal diffusivity is more severe in ceramics with high thermal diffusivity before irradiation, and final values are almost in the same range, but slightly higher in the case of SiC or Si$_3$N$_4$ ceramics. It is known that the effect of neutron irradiation on thermal conductivity of ceramics appears from very low neutron doses, so that quick degradation of thermal conductivity should be carefully monitored.

### 3.1. Silicon Carbide

One of the polytypes of silicon carbide (SiC) is cubic 3C, by Ramsdell’s notation, belonging to the zinc-blende structure. Many other polytypes of SiC belong to hexagonal (2H, 4H, 6H, and so on) or rhombohedral system (15R, 21R and so on). Anisotropy of properties is not so obvious in non-cubic polytypes, since basic component of SiC is SiC$_4$ polyhedron and stacking sequence perpendicular to the sheet plane is only modified in each polytype. Difference in radiation effects on silicon carbides by polytypes is not significantly recognized up to the present.

Dimensional change due to neutron irradiation of SiC can be categorized into 3 regimes depending on irradiation temperatures if the fast neutron dose is rather high, more than $1 \times 10^{24}$ n/m$^2$. At low irradiation temperature ($T_{irr} < \sim 100$ °C), crystalline SiC can be amorphized by neutron irradiation greater than a few dpa ($\sim 2 \times 10^{25}$ n/m$^2$) [4]. In the intermediate temperature range (100–1050 °C), swelling induced by the strain due to irradiation-induced point defects and their small complexes saturates less than one dpa, as shown in Figure 1 [5]. Saturated amount of swelling is inversely related to irradiation temperature, as shown in Figure 2, decreasing with increasing irradiation temperature [6]. In this temperature range, the formation speed of displaced atoms by knock-on should be balanced by recombination speed of defects. Recombination of point defects is mainly caused by the migration of interstitial atoms. At the highest temperature range studied (1050–1500 °C), migration of both irradiation-induced interstitials and vacancies is possible to form defects clusters, such as voids, and leads to swelling that progresses with both temperature and irradiation fluence. Figure 3 indicates an example of void formation inside SiC crystal of the specimen irradiated at 1050–1460 °C up to $1.5–9.6 \times 10^{25}$ n/m$^2$ ($E > 0.1$ MeV) [7].

Clusters of interstitial atoms also formed in SiC in the intermediate and highest temperature ranges. The size of such defect clusters depends on both fluence and irradiation temperature; higher fluence and increase in temperature promote formation of extended defect clusters, as shown in Figure 4 [6]. Yano et al. first observed the structure of these clusters and clarified that it is an interstitial dislocation loop using high-resolution electron microscopy [8]. One

<table>
<thead>
<tr>
<th>TABLE 1 Amount of Linear Swelling of Typical Engineering Ceramics Irradiated Concurrently by Fast Neutrons [3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutron fluence (n/m$^2$) ($E_n &gt; 0.1$ MeV)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>$3.9 \times 10^{26}$</td>
</tr>
<tr>
<td>$4.2 \times 10^{26}$</td>
</tr>
<tr>
<td>$6.9 \times 10^{26}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2 Thermal Diffusivity of Typical Engineering Ceramics Irradiated Concurrently by Fast Neutrons [3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutron fluence (n/m$^2$) ($E_n &gt; 0.1$ MeV)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Unirradiated</td>
</tr>
<tr>
<td>$0.5 \times 10^{26}$</td>
</tr>
<tr>
<td>$1.4 \times 10^{26}$</td>
</tr>
<tr>
<td>$0.4 \times 10^{26}$</td>
</tr>
</tbody>
</table>
tetrahedral unit sheet is inserted into {111} planes, as shown with a high-resolution electron micrograph in Figure 5. Precise analysis clarified that one rotated SiC₄ tetrahedral layer is inserted into {111} planes, and Burger’s vector is \(1/3\langle 111\rangle\). Structural model of defected part is shown in Figure 6.

Effect of neutron irradiation on the room temperature thermal conductivity of SiC is presented in Figure 7 [6]. It can be observed that strong correlation exists between amount of swelling and thermal conductivity at room temperature, indicating that introduction of crystalline defects which can deform a crystalline lattice also disturbs conduction of heat significantly.

3.2. Silicon Nitride

Silicon nitride (Si₃N₄) usually crystallizes into two kinds of phases, \(\alpha\) and \(\beta\). After higher temperature sintering (more than 1600 °C), most crystals are transformed into \(\beta\)-phase, although \(\alpha\)-phase powder is used as a raw material. The structure of \(\beta\)-Si₃N₄ is based on the packing of SiN₄ tetrahedra. Each nitrogen atom is shared by three tetrahedra. The structure contains six-membered, corner-shared tetrahedral rings. Along the \(a\)-axis or \(b\)-axis, one unit cell consists of three layers of SiN₄ tetrahedra. Sheets consisting of tetrahedral rings stack along the \(c\)-axis without any relative displacement, so that relatively wide open channels penetrate the (0001) plane along the \(c\)-axis.
Research on neutron irradiation effects on Si$_3$N$_4$ is limited up to present. Superior retention of flexural strength and thermal diffusivity to oxide ceramics such as alumina or spinel is reported for Si$_3$N$_4$ and SiAlON ceramics irradiated up to $1.0 \times 10^{24}$ n/m$^2$ at 150 °C [9]. Small swelling and relatively large drop of thermal diffusivity after irradiation up to $2.8 \times 10^{25}$ n/m$^2$ ($E > 0.1$ MeV) at 740 °C are also reported [10]. Dimensional stability of silicon nitride after concurrent neutron irradiation up to the order of $10^{26}$ n/m$^2$ was superior than that of SiC, and was kept relatively higher thermal diffusivity same as that of SiC, as shown in Tables 1 and 2, respectively [3,11,12].

After high-dose neutron irradiation on Si$_3$N$_4$, dislocation loops are densely formed mostly parallel to the [0001] axis. From electron micrographs of the as-irradiated specimens, three different microstructures are observed [13]. In the specimen irradiated up to $0.5 \times 10^{26}$ n/m$^2$ at 380 °C, no clear loop formation was identified, but small strain contrasts were densely observed throughout grains. If the irradiation temperature increased to 540 °C, tiny loop-like contrasts

FIGURE 3 Cavity microstructures in β-SiC irradiated at (a) 1050 °C, 4.9 dpa, (b) 1300 °C, 9.3 dpa, (c) 1400 °C, 9.4 dpa, and (d) 1460 °C, 5.6 dpa [7].

FIGURE 4 Microstructural development during neutron and self-ion irradiation [6].
with distorted strain contrast were observed. After higher fluence at higher irradiation temperature, extended defects are induced. Figure 8 shows representative HREM images of neutron-irradiated Si3N4 under different conditions [14]. Dislocation loops were frequently observed as differently contrasted portions compared with the normal hexagonal atomic configuration in the specimens irradiated up to $2.8 \times 10^{26} \text{n/m}^2$ at 520 °C (c, d, and g) and a smaller number in the specimen irradiated to $0.4 \times 10^{26} \text{n/m}^2$ at 540 °C (b). Defects were looked-like line defects lying along the $<1120>$ or $<10\bar{1}0>$ directions if it was observed along the $[0001]$ axis. These were mostly dislocation loops lying on the {1010} or {1120} planes. In the specimen irradiated to $0.5 \times 10^{26} \text{n/m}^2$ at 380 °C, there were small areas where atomic arrangement did not clearly indicate the formation of loops but the contrast was different from perfect crystal and showed strain contrast, as indicated by circles in (e) and (f).

Typical dislocation loops that lie on the {1010} or {1120} planes are named type-I and type-II dislocation loops, respectively. Each configuration can be created by rearrangement of SiN4 tetrahedra. In the type-I dislocation, one SiN4 tetrahedral layer is inserted into {1010} planes, as shown in Figure 9 [13]. From the high-resolution image, and the measured excess thickness of the planar defects of about one-third of the normal unit cell spacing, it is
FIGURE 8  High-resolution electron micrographs of as-irradiated Si₃N₄ specimens. (a, e, f) $0.5 \times 10^{26}$ n/m$^2$, 380 °C, (b) $0.4 \times 10^{26}$ n/m$^2$, 540 °C, (c, g) $2.8 \times 10^{26}$ n/m$^2$, 520 °C, (d) $3.7 \times 10^{26}$ n/m$^2$, 730 °C [14].

FIGURE 9  A high-resolution electron micrograph of an interstitial dislocation loop (type-I) induced by neutron irradiation up to $2.8 \times 10^{26}$ n/m$^2$ at 520 °C [13].
indicated that single SiN$_4$ layer should be inserted between {10\ 10\ 0} planes, resulting in interstitial loop formation. The inset is the simulated image based on the structural model of type-I dislocation, as shown in Figure 10.

Type-II dislocation is also analyzed, and one SiN$_4$ tetrahedral layer is inserted into {11\ 2\ 0} planes [15]. All observed dislocations are created on the planes parallel to the $c$-axis, and not on (0001) basal plane. Type-I or type-II dislocations are forbidden to intersect other dislocations, but they can connect to different type dislocations. Atomic configuration of connected portion was also proposed [16]. These dislocation loops are relatively stable, and do not annihilate by thermal annealing up to 1500 °C.

### 3.3. Alumina

High-purity alumina (Al$_2$O$_3$) is the most typical insulator material and also a candidate insulator for fusion reactors. Alumina is mostly ionic-bonded crystal with corundum structure and hexagonal symmetry. A large number of neutron irradiation experiments on several kinds of alumina ceramics and single crystals have examined swelling, mechanical properties, and microstructure.

Volume change due to neutron irradiation is presented in Figure 11, based on the review by Pells [17] and Yano et al. [18], indicated that the amount of swelling depends on the irradiation temperature that is higher than 850 K or lower than 650 K. In the higher irradiation temperature, voids array along the $c$-axis is formed inside grains, as shown in Figure 12. In the lower irradiation temperature, mobility of vacancies is not so significant to form voids, and point defects are accumulated independently or form dislocations.

Interstitial dislocation loops are formed both on {10\ T\ 0} prismatic planes and on (0001) basal plane, with Burger’s vector of $1/3<10\ 10\ 0>$ [19]. Loops in alumina are frequently extended to form dislocation network. Furthermore, difference in change of lattice parameters due to irradiation is significant in the case of the higher temperature irradiation range, expansion of the $c$-axis is greater than that of the $a$-axis, and anisotropy increased with increasing irradiation temperature [20]. The anisotropy may be caused by the preferential formation of dislocation loops on the basal plane. If the anisotropy of lattice swelling is significant, grain boundary microcracks are induced into ceramics.

**FIGURE 10** Defect model for the interstitial dislocation loop formed in β-Si$_3$N$_4$ crystal shown in Figure 9.

**FIGURE 11** Macroscopic volume swelling of alumina as a function of neutron fluence [18].
resulting in drastic decrease in mechanical properties. Example of microcrack formation is presented in Figure 13. On the other hand, in the lower irradiation temperature range, mostly isotropic expansion is reported.

Furthermore, in situ electrical property measurements have also been conducted to clarify radiation-induced electrical conductivity (RIC), and results are shown in Figure 14 [21]. It is generally agreed that RIC is caused by free electrons excited by ionizing irradiations. However, detailed knowledge concerning mobility, mean free path, and the average lifetime of electrons is needed to understand RIC thoroughly. From an engineering point of view, the RIC effect could be described only by the term of an ionizing dose rate, irrespective of radiation sources.

3.4. Aluminum Nitride

Aluminum nitride (AlN) ceramics with very high thermal conductivity and high electrical resistivity have been developed recently; therefore, neutron irradiation data on AlN are very limited compared with those of alumina, particularly at high fluence. The crystalline phase of AlN is 2H-type, hexagonal wurtzite structure. The structure is presented as a sequential stacking of AlN₄ tetrahedral sheets, parallel to the (0001) plane. From limited data, it appears that the anisotropic feature of the irradiation response is similar to that of alumina, except for void formation only in alumina.

According to the increase of fast neutron fluence, roughly more than 5 × 10²¹ n/m², swelling of the a-axis and c-axis due to the irradiation changed from isotropic to anisotropic in manner. Figure 15 indicates change in the ratio of the c-axis length/a-axis length as a function of neutron fluence [22]. Beyond 5 × 10²¹ n/m², the ratio increases markedly. In anisotropically changed specimens, dense tiny interstitial dislocation loops are observed on the (0001) basal plane [23], as shown later. With further increase in fluence and/or higher irradiation temperature, finally the strain caused by anisotropic swelling of grains induces microcracks along the grain boundary, as shown in Figure 16. It is also clearly indicated from the departure of macroscopic volume change and unit cell volume change, as shown in Figure 17 [18].

A high-resolution electron micrograph taken along the [1120] incident beam direction of AlN specimen neutron-irradiated to 2.4 × 10²⁴ n/m² is shown in Figure 18 [23]. This projection shows both the (0001) basal plane and the {1100} prism plane on edge view. A two-layer repeat of 0.25-nm-spaced fringes along the [0001] is clearly observed. One array of black dots parallel to the (0001) plane corresponds to one tetrahedral layer. The array of dot parallel to the (0001) planes is waved at the center part labeled X or Y. In this case, single extra layer parallel to the
(0001) plane is inserted between the original stacking layers. The stacking sequence perpendicular to the loop plane is estimated to be $AB_0/A_0C_0/B_0A$ as a result of an $A$ layer plus an inserted layer creating $A_0C_0$ layers. The atomic arrangement of this stacking sequence is modeled in Figure 19. Another possible variant of the stacking sequence is $AB_0/A_0C_0/B_0A'$ as a result of $B_0$ layer plus an inserted layer creating $BC$ layers. This model gives the opposite stacking direction of sphalerite layers. The loop labeled $Y$ in Figure 18 coincides with this stacking sequence model. The Burgers vector of the loops is $(1/2)[0001]$.

4. ADVANCED CERAMICS FOR FISSION REACTORS AND FUEL CYCLE

4.1. Advanced Materials for High-Temperature Gas-Cooled Reactors

As one of the advanced fission reactors, high-temperature gas-cooled reactors are promising since very high operation temperature leads to high heat/electricity conversion efficiency. Maximum temperature of outlet He coolant may be close to 1000 °C. Not only electricity generation but also direct use of output heat for steel making plants or other chemical plants are considered, including hydrogen production and desalination of sea water.

Metallic fuel cladding pipes and core structure materials cannot suffer from deformation and chemical interactions under such a high-temperature; therefore, entirely another concept of core structure with refractive materials is indispensable. For the structural materials incorporated with a function of neutron moderator (that slows down speed of neutrons desirable for the uranium fission reaction), the so-called “nuclear (grade) graphite” is applied.

As same as other crystalline materials, graphite also received radiation damage under fast neutron irradiation, as mentioned in Section 2. The graphite crystal is stacking of covalently bonded strong graphene sheets that are weakly bonded together along the $c$-axis through the van der Waals
force. Due to the neutron irradiation, the displaced carbon atoms come to rest as interstitial complexes between existing planes. Mobilities of both interstitial atoms and vacancies depend on irradiation temperature, and then formation rate of each defect cluster is varied in a complex manner. Irradiation causes defects to form between basal planes resulting in the \( c \)-axis swelling. On the other hand, collapse of the basal plane vacancy clusters results in the \( a \)-axis shrinkage, as shown in Figure 20 [5]. As a result, the graphite crystal undergoes in an approximately volume-conserving fashion, i.e., dimensional change of graphite is relatively small. Besides its very high-temperature stability under inert atmosphere, graphite is a very attractive material for high-temperature gas-cooled reactors, because of the large manufacturing capability, machinability, coupled with neutron moderation ability, and relatively good radiation-tolerance mentioned above. To obtain isotropic
feature as a bulk material, microstructure control is important. Suitable graphite for high-temperature reactors should have a relatively good isotropy \((\alpha_{xx}/\alpha_{yy} < 1.15)\), a density higher than 1.75 g/cm³, appropriate mechanical strengths (compressive strength before irradiation greater than 65 MPa), a high resistance to crack propagation, and an impurity level as low as possible, particularly neutron absorbing elements. Under irradiation, the dimensional changes should be as small as possible and the thermal conductivity as high as possible.

Confinement of fission products from fuels is another critical issue in high-temperature gas-cooled reactors, since temperature of fuel itself reached more than 1000 °C in normal operation and higher during loss of coolant events. Instead of metal cladding pipes in the case of water-cooled reactors, refractory materials are necessary. For this purpose, a coated particle fuel concept is established. Uranium dioxide (UO₂) or uranium oxycarbide (UCO) particles (fuel kernel) of diameter around 800 μm are coated by three layers of materials, termed TRISO (tri-structural isotropic) fuel. In TRISO-coated fuel, a SiC layer is sandwiched between pyrolytic carbon layers, and these coatings provide thermo-mechanical strength to the fuel kernel and containment of fission products. The fissile fuel kernel is first surrounded by a porous carbon (buffer) layer, which will accommodate volumetric swelling of fuel, attenuate fission recoils, and retain released fission gases. Next is a dense carbon layer named inner pyrolytic carbon layer which protects the kernel from reactive chlorine compounds produced during the SiC-coating deposition process and acts as diffusion barrier for gaseous fission products. Next is a dense SiC layer which provides mechanical strength of the particle and acts as diffusion barrier for metallic fission products. The outermost is a dense carbon layer named outer pyrolytic carbon layer which protects the particles during succeeding fuel-element process and also acts as diffusion barrier for gaseous fission products. The inner and outer pyrolytic carbon layers are deposited from mixture of acetylene, propylene, and Ar between 1200 and 1400 °C. The SiC layer is deposited from a mixture of H₂ and methyltrichlorosilane between 1500 and 1600 °C. There are two designs of fuel compacts, one is pebble bed design and the other is prismatic design, as shown in Figure 21 [24]. In reactors with pebble bed or prismatic fuels, the core of the reactor consists of billions of these TRISO-coated fuels distributed within fuel elements in the shape of circular cylinders (12.5 mm in diameter and 50 mm in length) called compacts or spheres called pebbles (60 mm in diameter).

![Diagram of TRISO fuel](image_url)

**FIGURE 21** Illustration of fabrication process of TRISO fuel into particles, compacts, and pebbles [24]. For color version of this figure, the reader is referred to the online version of this book.
Recently, a ZrC coating is examined instead of a SiC-coating of TRISO-coated particles to improve high-temperature mechanical integrity and to reduce chemical reaction with fission product Pd, since SiC gradually loses strength due to neutron irradiation and mechanical integrity at high-temperatures, particularly above 1700 °C, by thermal dissociation. The ZrC coating layer is deposited with the pyrolytic reaction of ZrBr4, CH4, and H2 at about 1500 °C in a fluidized bed [25,26].

4.2. Advanced Ceramics for Minor Actinide Transmutation

After some operation periods, nuclear fuels are replaced with new fuels, and spent fuels are reprocessed to retrieve uranium containing residual 235U and fissile Pu for use as recycled fuels. During reprocessing process, high radioactivity nuclear waste containing many kinds of fission products is separated. The high-level waste is immobilized into borosilicate glass and contained into steel bottles, and finally it is planned to be deposited into deep underground with multiple barriers. Not only the radioactivity of the waste is very high but also they generate heat for long period more than tens of thousands years. To reduce risks of leakage of active waste into human environments and to reduce deposit cost, it is proposed that large parts of harmful nuclides, such as minor actinides (MAs) or long-lived fission products (LLFPs), should be separated and transmuted into more safe (short-lived or stable) nuclides using nuclear transmutation reactions in nuclear reactors or accelerators adopting nuclear transmutation reactions. For this purpose, materials to contain and immobilize MA and LLFP nuclides are necessary. These materials are named “inert matrix”. Requirements for the inert matrix are high-temperature chemical stability, high thermal conductivity, low neutron absorption, and appropriate mechanical integrity with proper radiation-tolerance up to relatively high doses, because inert matrix containing MA, so-called inert matrix fuel, will be irradiated for long time at the core of fission reactors, such as fast breeder reactors.

ZrO2, Y2O3, MgO, MgAl2O4, ZrN, SiC, and other ceramics are candidates of inert matrix. Silico nitride ceramics is a candidate since it is tolerant to heavy neutron irradiation and preserves relatively high thermal conductivity after severe irradiation compared with these of many ceramics, as shown in Section 3.2. For the inert matrix fuel, several kinds of microstructures are studied. Homogeneous microstructure, for example, solid solution such as MA/LLFP elements dissolved into a compound such as zirconia-yttria or ceria, is a typical one. The other is heterogeneous microstructure such as composite or mixture. MA/LLFP-containing compounds dispersed into a ceramic inert matrix of MgAl2O4 or MgO, or dispersed into a metallic matrix such as Mo, are typical candidates. Nitride fuels such as AmN are also one kind of target fuels, but in this case no inert matrix is necessary. In the case of Si3N4, the crystal itself cannot contain other elements so much, particularly large-sized cations such as transuranium elements. Otherwise, Si3N4 ceramics can be densified via liquid-phase-assisted sintering. In other words, if MA/LLFP elements can form liquid-phase at a relatively low temperature, they can be immobilized as grain boundary phases in Si3N4 ceramics. In this case, dispersant particles are an inert material, and continuous phase (matrix) contains MA/LLFP; thus, it is a new type of inert matrix fuel. The same idea can be applied for ceramics using SiC as an inert material.

Figure 22 shows microstructure of Si3N4 ceramics containing large amount of CeO2 as a raw material, as a simulant of minor actinides. Since MA compounds are easy to evaporate at higher temperature, decrease in densification temperature is important. Incorporated with silica and magnesia, dense Si3N4 ceramics can be obtained. Grain boundary glassy phases contain Ce element. Thermal conductivity of the Si3N4 ceramics sintered at 1550 °C is
about 20 W/m K and increases with sintering temperature. Only grain boundary phase can be dissolved using hot nitric acid for further reprocessing [27].

5. ADVANCED CERAMICS FOR FUSION REACTORS

Silicon carbide fiber-reinforced silicon carbide composite (SiC/SiCf) is one of the candidate structural materials for future fusion reactors because of its low induced radioactivity, quick decay of activity, low heat evolution, low atomic number, good fracture resistance, excellent high-temperature properties, and corrosion resistance. Fundamentally, SiC shows good resistance for high-energy neutron irradiation up to very high neutron fluences, as mentioned in Section 3.1.

There are several processes to fabricate SiC/SiCf composites, for example, chemical vapor infiltration (CVI) method and polymer infiltration and pyrolysis (PIP) method. These processes have some advantages, such as a relatively low processing temperature and formation of pure matrix. However, these processes require a long manufacturing time that leads to high processing cost. Furthermore, the SiC/SiCf composites fabricated by these processes usually contain 10–30 vol.% of voids, resulting in low mechanical and thermal properties. In order to simplify the fabrication process and to obtain dense SiC/SiCf composites with good mechanical and thermal properties, a new fabrication process using hot-pressing is developed [28,29]. SiC/SiCf composite can be fabricated using a green sheet of SiC with proper sintering additives to decrease the densification temperature and polycarbosilane (PSC)-impregnated two-dimensionally woven SiC (Hi-Nicalon) cloth with BN-coating by hot pressing at 1650–1750 °C [30].

Figure 23 shows the load—displacement curves of the SiC/SiCf composites measured at room temperature. These composites exhibited nonbrittle fracture behavior. After the elastic region, the nonlinear region, which is caused by fiber bridging and sliding after debonding, appeared. The SiC/SiCf composite hot-pressed at 1700 °C showed higher maximum strength (240 MPa) than the composites hot-pressed at 1650 °C or 1750 °C. Maximum strength and fracture mode are mostly maintained at test temperature of 1200 °C [30]. Microstructure of this SiC/SiCf composite is shown in Figure 24.

Based on the hot-pressing technique, further improvement in coating and infiltration using electrophoretic deposition method [31], reaction-sintering [32], or infiltration of nanometer-sized SiC particles [33] is progressed recently.

Another critical component of fusion reactors is a diverter plate of very high heat-transfer property. The heat received at the surface of fusion diverter will be expected up to 10 MW/m². For this component, a fiber-reinforced carbon composite is a primary candidate. In this composite, basal plane of graphite crystals is highly arraigned perpendicular to the surface of the component and brazed on the high-conductivity copper base. The prototype of diverter is shown in Figure 25.

![Figure 23](image1)

**Figure 23** Load—displacement curves of the SiC/SiCf composites tested at room temperature. The thickness of the BN-coating was 0.4 μm. Fiber volume fraction of the composites was about 52 vol.% [30]. For color version of this figure, the reader is referred to the online version of this book.

![Figure 24](image2)

**Figure 24** Microstructure of SiC/SiCf composite hot-pressed at 1700 °C (left: polished cross section) or 1700 °C (right: fracture surface) [30].
For the neutral beam injector for fusion devices, a five-stage insulator composed of large bore ceramic rings has been required to sustain 1 MV. Since many conductors penetrate inside the ceramic ring maintaining insulation of 1 MV, a dimension of the ceramic ring of 1.56 m in diameter, 29 cm in height, and 5 cm in thickness is successfully fabricated by cold isostatic press and sintering. This high-purity alumina ring is world’s largest one in size. These ceramic rings are joined with metal rings with high vacuum tightness by brazing [34].

In present experimental fusion facilities, radiofrequency heating of the plasma is a necessary support heating system. In that system, ceramic windows with very low tan δ and very high heat-transfer properties are necessary. Artificial diamond seems to be the best material for that window and was actually applied to the JT-60 fusion experimental device in Japan [35], due to its very high thermal conductivity on the order of $<1000 \text{ W/m K}$ and low tan δ on the order of $<1.3 \times 10^{-4}$ at 170 GHz before neutron irradiation [36]. Neutron irradiation also induces point defects into diamond and breaks sp³ configuration [37,38], so that application of a diamond window close to the vessel of fusion reactors should be avoided.

6. CONCLUSION

In this chapter, basic radiation effects in ceramics particularly by neutron irradiation were mentioned first, and then examples of neutron irradiation damage on advanced ceramics were reviewed. Finally, some applications of advanced ceramics for fission and fusion reactors were briefly introduced. Important nuclear fuel-related ceramics, UO$_2$–PuO$_2$ ceramics [24], and control rod materials such as B$_4$C [39] were not included in this chapter due to page limitation. Various kinds of ceramics will be needed particularly for fusion reactors, but not mentioned in this chapter [40]. The reader is referred the proceedings series of international conference for fusion reactor materials for more information, for example, that of the 14th conference [41]. For configuration and system of several kinds of nuclear power plants and basics of nuclear engineering, the reader is referred to textbooks [42–45].

For development of advanced nuclear reactors, many advanced ceramics should be applied. As requirements for applications in nuclear fields, special caution should be paid for property degradation by intense radiation. On the other hand, highly controlled materials will be developed for nuclear application; these materials should be observed with superior performance in non-nuclear fields.

REFERENCES


FIGURE 25 Photograph of prototype diverter components of the ITER fusion device. Black colored potion with slits is made of a C/C composite. For color version of this figure, the reader is referred to the online version of this book.


Chapter 5.1

Glass-Ceramics

Wolfram Höland* and George H. Beall†

*Research and Development, Ivoclar Vivadent AG, Bendererstrasse 2, 9494 Schaan, Liechtenstein, †Corning Inc., SP-FR-05-1, Corning, New York 14831, United States

Chapter Outline

1. Introduction 371
2. Glass-Ceramics With Minimal Thermal Expansion 372
3. Glass-Ceramics With Special Optical Properties 373
4. Machinable Glass-Ceramics 375
5. High-strength and High-toughness Glass-ceramics 375
6. Biomaterials 376
7. Products With Special Electrical and Magnetic Properties 378
8. Applications in Energy Technology 379
References 379

1. INTRODUCTION

The term glass-ceramics refers to a material that combines two types of materials to form a product that is in a class of its own between glasses and polycrystalline ceramics. Glass-ceramics are special in that they can be tailored to demonstrate both the properties of glasses and those of polycrystalline materials. The fact that they can be regulated in this way is due to their composition and microstructure and their synthesis and manufacture. The processing of glass-ceramics, therefore, involves a variety of controlling mechanisms. The definition of glass-ceramics below reveals that a number of controlling processes are necessary in the development and manufacture of these materials:

Glass-ceramics consist of one or more glass phases and one or more crystalline phases. The crystalline phases are produced by the controlled nucleation and crystallization of a base glass.


The mechanisms of controlled nucleation constitute the prerequisites for the development of glass-ceramics. The fundamentals of nucleation and crystallization in amorphous materials systems, especially glasses, are described in numerous technical publications and textbooks [28,45] as well as in review papers [10,23,57,77]. Two mechanisms, both of which exhibit specific variations, are of considerable importance in the fabrication of glass-ceramics. The first involves controlled internal/volume nucleation and crystallization. This mechanism can be subdivided into homogeneous and heterogeneous processes. In many cases, the mechanisms of heterogeneous reactions together with the principles of epitaxy or amorphous glass-in-glass phase separation are used. The second type of reaction mechanism used in the fabrication of glass-ceramics is that of surface crystallization. In this case, nuclei form on the surface of the glass and initiate the crystallization process. If glass grains or glass powders are used in this process, the surface of the glass particles crystallizes. During the simultaneous sintering process, a powder compact is formed. Höland and Beall (2002, 2012) [35,36] have summarized the details of these two mechanisms, that is, of internal nucleation and surface nucleation and crystallization, as well as their relationship to the fundamental theories of nucleation and crystallization in their textbook.
In order to fit the requirements of a reference book, such as the present “Handbook of Advanced Materials”, the authors present an overview of the available glass-ceramic materials and discuss them in an order dictated by their special properties. In the different sections of the book, the properties of the glass-ceramics are presented according to the title of the section. Subsequently, the chemical and materials-specific aspects are explained (chemical system, mechanisms of nucleation and crystallization, microstructure). Finally, products (marked with ® or ™) are presented and their preferred applications are shown. The basis for the present publication is the textbook entitled “Glass-ceramic technology”, authored by Höland and Beall (2002, 2012) [35,36] and the references cited in it. All the information which is only briefly outlined or summarized in the current publication can be studied in detail in this textbook.

2. GLASS-CERAMICS WITH MINIMAL THERMAL EXPANSION

The largest group of materials among the glass-ceramics in terms of quantity and economic significance is comprised of products with minimal thermal expansion. These products exhibit practically no physical expansion over a wide temperature range. It is difficult to imagine that a material can show only minimal expansion between room temperature (or even lower) and 700 °C. All other materials and metals and ceramics in particular are known to expand as temperatures rise. Beall (1971), Petzoldt and Pannhorst (1991) [4,60] and Tashiro and Wada (1963) [72] developed this type of glass-ceramic by the controlled precipitation of crystals with low and even negative thermal expansion (in other words, shrinkage). Two very special crystal phases have given these ceramics their respective names: a) β-spodumene solid solution and b) β-quartz solid solution glass-ceramics.

2.1. β-Spodumene Solid-Solution Glass-Ceramics

The crystals of the spodumene type possess the chemical composition LiAlSi2O6–SiO2 s.s. In addition to the chemical composition of the crystals, their crystallographic modifications are of significance to the product. The aim of development efforts is to form the high-temperature modification (β-phase), which exhibits the special thermal expansion properties. This modification of the crystal phase can be present in metastable form at room temperature because its transformation to stable α-spodumene is extremely sluggish, being reconstructive and occurring below 500 °C. Also, it is usually a mixed crystal with variable silica, in which case it would require a breakdown to two stable crystalline phases, a feat even more difficult at low temperature. Further details about these crystallographic aspects are described by Höland and Beall (2002, 2012) [35,36].

Glass-ceramics were developed in the SiO2–Al2O3–Li2O–MgO–ZnO–Na2O–K2O–TiO2–(ZrO2) system. The melted and cooled glasses were heat-treated in order to promote nucleation and crystallization (in situ crystallization). The nucleation and crystallization mechanisms were characterized by heterogeneous nucleation with the TiO2 nucleating agent, which initiates glass-in-glass phase separation (Stookey, 1959 [70]; Beall, 1971). In addition, Tashiro and Wada (1963) [72] described the nucleation effect of ZrO2. Further thermal treatment following the nucleation process resulted in the formation of the crystalline β-quartz solid-solution precursor. Finally, between 900 and 1000 °C the β-quartz solid solution was irreversibly transformed into β-spodumene solid solution and the secondary phase of rutile (TiO2).

The crystal type as well as the microstructure of the glass-ceramic shown in Figure 1 contribute to the minimal expansion. As visible in the SEM, the diameter of the β-spodumene solid-solution crystals measures between approx. 800 nm and 1 μm. The appearance of the glass-ceramic is white and very opaque, due primarily to the scattering power of rutile with its high birefringence and index of refraction.

This type of material was one of the first glass-ceramics to be developed by Stookey (1959) [70]. It was registered under the name of Pyroceram® 9608 or Corning Ware® 9608 (Corning Inc., USA). The coefficient of thermal expansion (CTE) of this product is between 0.7 and 1.2 × 10⁻⁶ K⁻¹ (⁰–500 °C). Because of this minimal expansion coefficient, this glass-ceramic has a very high resistance to thermal shock. Further modifications of the mentioned chemical system resulted in the development of Cercor® (Corning, USA) and Neoceram™N-11 (Nippon Electric

FIGURE 1 Microstructure of Pyroceram®/Corning Ware® 9608 showing β-spodumene solid solution crystals. (Reproduced from Höland and Beall, 2002 [35]; with permission of The American Ceramic Society, www.ceramics.org. Copyright 2002. All rights reserved.)
Glass, Japan). The special properties of this type of glass-ceramics allow these materials to be used in the manufacture of consumer goods, such as kitchen ware. Figure 2 shows EuroKera/C210 dishes on a glass-ceramic cooktop panel. The glass-ceramic hob is presented in Section 2.

2.2. β-Quartz Solid-Solution Glass-Ceramics

If the heterogeneous internal double nucleation mechanism of TiO₂ and ZrO₂ is used in the SiO₂–Al₂O₃–Li₂O–MgO–ZnO–Na₂O–K₂O–TiO₂–ZrO₂ system, a foreign substrate of highly dispersed ZrTiO₄ crystals is formed by solid-state reactions in the nucleation process [7,53]. In the thermal process of crystallization, the main crystal phase of β-quartz solid solution is formed by epitaxy on these nano-sized crystals. The existence of this metastable high quartz modification at room temperature is made possible by the incorporation of Li⁺, Mg²⁺, Zn²⁺, Al³⁺, or even (PO₄)³⁻ ion into the crystals. These compositions are called stuffed derivatives of β-quartz. The microstructure of this glass-ceramic is very special: The diameter of the β-quartz solid-solution crystals is smaller than 100 nm and the heterogeneous nuclei measure only a few nanometers in size. Compared with the main crystal phase, an excessive number of nuclei are formed. Only one nucleus initiates the formation of β-quartz s.s., while the remaining heterogeneous crystals agglomerate around the main phase. The microstructure has the appearance of a cellular membrane.

Due to the small size of the crystals and their isolated precipitation in the microstructure of the glass-ceramic, these products exhibit not only minimal thermal expansion (practically no expansion at all) and high thermal shock resistance but also high transmission in the visible range (see Section 3).

The most important products and their applications are presented below (this list contains only a small selection of examples):

- Vision® (Corning Inc., USA): domestic products (Figure 3)
- Keraglas®, EuroKera®, Keralite® and Eclair® (Corning Inc./St. Gobain, France): cooktop panels and safety-rated material
- Ceran® (Schott AG, Germany): cooktop panels [3]
- Zerodur® (Schott AG, Germany): telescope mirror (Figure 4) (Petzoldt and Pannhorst, 1991)
- Neoceram® N-0, Nurami® (Nippon Electric Glass, Japan): cooktop panels, domestic products, optical connectors/ferrules [64,65]

3. GLASS-CERAMICS WITH SPECIAL OPTICAL PROPERTIES

Apart from the glass-ceramics of the β-spodumene s.s., and β-quartz s.s. types, the following materials deserve particular mention: glass-ceramics with a main crystal phase of β-eucryptite (LiAlSiO₄), lithium metasilicate (Li₂SiO₃), Cr-doped mullite (Al₆Si₂O₁₀), oxyfluoride, and Zn-spinel.

3.1. β-Quartz s.s., β-spodumene s.s., and β-eucryptite s.s. Glass-Ceramics

β-Quartz s.s. glass-ceramics were discussed in Section 2. They are also mentioned in the present section, because of their exceptional optical properties (in addition to their minimal thermal expansion), which are requisite in products such as Vision® (Corning Incorporated, USA) and Zerodur® (Schott AG, Germany).

β-Spodumene s.s. glass-ceramics were also discussed in Section 2. Because of their specific optical characteristics,
they are used in the fabrication of ferrules as connectors in telecommunications technology [64].

β-Eucryptite (LiAlSiO₄) s.s. crystals were developed in glass-ceramics of the SiO₂–Al₂O₃–Li₂O system by Beall et al. (1998) [9]. This type of glass-ceramic exhibits a microcrack network in the microstructure. This characteristic is exploited to produce optical components for glass fiber-based telecommunications, e.g. fiber Bragg grating packages (Corning Incorporated, USA).

3.2. Lithium Metasilicate

In contrast to the crystal phase of lithium disilicate (Li₂Si₂O₅), that of lithium metasilicate (Li₂SiO₃) is chemically very unstable. Borelli (1985) [16] took advantage of this characteristic to develop a very special product. Thin films (20 μm) made of a base glass of the SiO₂–Al₂O₃–K₂O–Na₂O–AgO–AuO–Ce₂O₃ system (with additions of SnO₂ and Sb₂O₃) were masked with a pattern and irradiated with UV light. The pattern prevented circular areas on the glass sheet from being irradiated. In the irradiated part of the glass, the following reactions occurred:

\[
\text{Ce}^{3+} + h\nu(312 \text{ nm}) \rightarrow \text{Ce}^{4+} + e^- \\
\text{Ag}^+ + e^- \rightarrow \text{Ag}^0
\]

The metallic silver nuclei (Ag⁰) formed in this process act as a nucleating agent for the crystallization of Li₂SiO₃ at 600 °C. As these crystals form, they cause the circular glassy regions to be squeezed into microglass lenses with a diameter of 400 μm and thickness of 20 μm (see Figure 5). The lens shape is created as a result of the fact that the crystals are denser than the glass. The product and the process in which it is produced are called Fotoform® (Corning, Inc., USA) [8].

3.3. Cr-Doped Mullite, Oxyfluoride, and Zn-Spinel Glass Ceramics

A very important branch of glass-ceramics development is concentrated on producing high-performance optical components. As a result, three additional glass-ceramics need to be discussed here: Cr-doped mullite, oxyfluoride, and Zn-spinel glass-ceramics.

Beall (1992, 1993) [7,8] was able to precipitate a main crystal phase of mullite (3Al₂O₃·2SiO₂) in the base glass system of Al₂O₃–SiO₂ by doping it with Cr₂O₃. These products allowed the characteristic transfer of electrons from 4T₂ to 4A₂ at an absorption band of 730 nm. This transfer represents a fluorescence process. Likewise, phosphorescence was established at the absorption band of 695 nm, due to the electron transfer from 2E to 4A₂. These products are used in solar and laser technology.

Tick (1995) [73] developed oxyfluoride glass-ceramics in the SiO₂–Al₂O₃–CdF₂–PbF₂–ZnF₂–YF₃–ErF₃–PrF₃ system with the main crystal phase of (Pb,Cd,Re)F₂·3, where Re represents the rare earth elements of the applied system. Furthermore, Dejneka (1998) [22] was able to demonstrate the incorporation of Re elements into the LaF₃ crystals. Beall (2000) [12] demonstrated a transparent forsterite (Mg₂SiO₄) glass-ceramic, which, when doped with Cr⁴⁺, displayed a wide (~400 nm) emission band centered near 1100 nm in the infrared. These types of glass-ceramics can be used for amplification purposes in telecommunications.
Pinckney and Beall (1997) [61] and Beall and Pinckney (1999) [11] managed to precipitate nanoscale crystals of the spinel type (ZnAl$_2$O$_4$) in the SiO$_2$–Al$_2$O$_3$–ZnO–MgO system. The crystal size was controlled below 100 nm, and the crystals were evenly precipitated in the end product due to the addition of the nucleating component TiO$_2$. As the crystals are embedded in the glass-ceramic in isolated form, the glass-ceramic is transparent. This type of material can be used in the electronics industry to produce photovoltaic substrates.

4. MACHINABLE GLASS-CERAMICS

Glass-ceramics are popular because they can be machined with carbide tools rather than with diamonds. The following example clearly shows how the design of glass-ceramics can be utilized to obtain these properties.

4.1. Straight Mica Platelets

Beall (1971) [4] precipitated a mica of the fluorophlogophite type (Na,K)[Mg$_3$(Al,B)Si$_3$O$_{10}$]F$_2$ in the glass-ceramic. In the multicomponent SiO$_2$–Al$_2$O$_3$–B$_2$O$_3$–MgO–Na$_2$O–K$_2$O–F glass-forming system, glass-in-glass phase separation was used in such a way that controlled nucleation of a crystalline preliminary phase serving as a fluorine depot occurred. In the subsequent solid body reaction at temperatures between 900 and 1000 °C, phlogophite formed from the crystalline preliminary phase and the glass matrix. The microstructural development was controlled such that the platelet-shaped crystals grew until they touched each other. A microstructure with the appearance of a house of cards was formed (Figure 6). Mica with a similar morphology can be produced in the SiO$_2$–MgO–Na$_2$O–K$_2$O–F system. In this case, a tetrasilicic mica (KMg$_2.5$Si$_4$O$_{10}$F$_2$) grows [27].

Since the mica crystals in the house-of-cards microstructure touch each other, the characteristic cleavage properties of mica can be exploited to yield a machinable glass-ceramic. Cracks propagate within the crystals and the cracks can propagate from one crystal to the next.

Apart from their machinability, these glass-ceramics exhibit outstanding electrical insulation properties and strengths of up to 180 MPa. The most important products of this materials group and their applications are listed below:

- Macor® (Corning Inc., USA): products for mechanical engineering applications, engineering of specialized products for cosmos research and space technology
- DICOR® (Corning Inc, USA/Dentsply USA): biomaterials (see Section 6)

4.2. Curved Mica Platelets

The formation of this type of mica crystal is very unusual. Therefore, their development was eagerly anticipated (Höland and Beall 2002 [35]; Vogel and Höland 1988). In the SiO$_2$–Al$_2$O$_3$–MgO–Na$_2$O–K$_2$O–F system, considerably reduced glass-in-glass phase separation is observed compared with glass-ceramics containing straight mica. A crystalline preliminary phase does not form; however, the mica grows like the leaves of cabbage heads. This type of glass-ceramic is also very easy to machine (Bioverit® II, Vitron, Germany).

5. HIGH-STRENGTH AND HIGH-TOUGHNESS GLASS-CERAMICS

5.1. Enstatite

The chain silicate enstatite (MgSiO$_3$) was formed as the main crystal phase in the SiO$_2$–MgO–ZrO$_2$ glass system by Beall (1991) [6]. The complex nature of the different solid-state reactions, which have to be controlled, as well as further types of enstatite glass-ceramics are discussed in the aforementioned textbook [36]. The microstructure is characterized by a high, lamellar twinned crystal content. This
microstructure yields a $K_{IC}$ toughness of 5 MPa·m$^{0.5}$. As a result, toughness values similar to those of sintered ceramics of the ZrO$_2$ type (e.g. 3Y$_2$O$_3$–ZrO$_2$) can now be achieved. Apart from its high toughness, the material is highly refractory.

5.2. Canasite

By incorporating fluorine into the SiO$_2$–CaO–Na$_2$O–K$_2$O system, Beall (1991) [6] developed a canasite (fluorocanasite) (Na$_{3-x}$K$_x$Si$_2$O$_6$) glass-ceramic. The phase-formation process of the crystals is characterized by a preliminary reaction leading to the precipitation of CaF$_2$. The main crystal phase of the canasite type is formed with a high content of interlocking crystalline sperulites. This microstructure, like that of the previously mentioned enstatite system, enables the production of high-toughness ceramics. High $K_{IC}$ toughnesses of up to 5 MPa·m$^{0.5}$ were also established for this product.

5.3. Lithium Disilicate

The first glass-ceramic that paved the way for this new group of materials containing the layer silicate Li$_2$Si$_2$O$_5$ [70]. Subsequently, numerous research groups developed lithium disilicate glass-ceramics in the stoichiometric range of 2 SiO$_2$–Li$_2$O glass [24,25,55,77,78] and in multicomponent glasses [4,13,17]. The most significant advance was achieved by Beall et al. (1992) [7] with the development of chemically durable materials. Further research into this multicomponent glass-ceramics system led to the development of the SiO$_2$–Al$_2$O$_3$–Li$_2$O–K$_2$O–P$_2$O$_5$ system [37] and the SiO$_2$–Al$_2$O$_3$–Li$_2$O–K$_2$O–P$_2$O$_5$–ZrO$_2$ system [2,39]. The internal nucleation and crystallization initiated by P$_2$O$_5$ was discovered to be the driving force of crystal phase formation. This was in contrast to the investigations of Headley and Loehman (1984) [29]; which showed that epitaxy on crystalline Li$_3$PO$_4$ was the driving force in their glass system. In the above-mentioned system, the interface between an amorphous phosphate phase and the glass matrix served as the basis for the nucleation of Li$_2$SiO$_3$ and Li$_2$Si$_2$O$_5$ (see Höland and Beall, 2012 [36]). In a multistep heat-treatment process, an interlocking microstructure of the glass-ceramic forms (Figure 7). This type of microstructure, which is characterized by a crystal content of 40–60 vol%, is requisite for achieving flexural strengths between 400 and approx. 700 MPa (biaxial method) [63,68] and $K_{IC}$ fracture toughness values of 2.3–2.9 MPa·m$^{0.5}$ (according to the SEVNB method). Lithium disilicate glass-ceramics are used for the manufacture of substrates for magnetic disks (Goto, 1995) and for biomaterials (see Section 6.3).

![FIGURE 7 SEM of an interlocking microstructure of lithium disilicate glass-ceramic (Ivoclar Vivadent AG, Liechtenstein). (Reproduced from Höland and Beall, 2012 [36]; with permission of Wiley)](image)

6. BIOMATERIALS

Glass-ceramics have been developed in the form of both bio compatible and bioactive materials (e.g. for the replacement of bone in human medicine and that of dental hard tissue in restorative dentistry). A wide range of different biomaterials has been described. The largest group of bioactive materials is composed of phosphate-containing glass-ceramics [20,43,44,49,50]. Some of the restorative materials used in dentistry also contain phosphate; however, these contain a high concentration of additions that enhance their chemical durability.

6.1. Apatite-Containing Glass-ceramics

The development of inorganic products as biomaterials is based on the fundamental work of Hench et al. (1991) [30] in the glass system of SiO$_2$–Na$_2$O–CaO–P$_2$O$_5$. The bioactive product of their research is called BIOGLASS®. This specialized glass is mainly used as an additive in tooth pastes [31]. Further research led to the development of special scaffolds [42] and composites [14,18] used as bone replacements in human medicine. Kokubo (1991) [49] developed a strong, bioactive glass-ceramic by controlling the nucleation and crystallization processes of two phases: fluoroapatite (Ca$_5$(PO$_4$)$_3$F) and wollastonite (CaSiO$_3$). This glass-ceramic, by the name of Cerabone® A-W, was successfully used in the form of implants in human patients in more than 10 000 clinical cases between 1991 and 2000 (predominantly as artificial spacers, vertebra and fillers).

Apart from the development of bioactive biomaterials, research has gone into the development of biomaterials for restorative dental applications. These products do not show any bioactivity. However, they demonstrate very good chemical durability (particularly durability in oral conditions). Amounts of 5–15 vol% of the main crystal phase of fluoroapatite are contained in the glass-ceramics
of the SiO$_2$–Al$_2$O$_3$–Na$_2$O–K$_2$O–CaO–P$_2$O$_5$–F system. Crystallization occurs internally. The special characteristic of this material is its needle-like morphology, as crystals grow along the crystallographic c-axis \[40\] (Figure 8). Similar crystallization phenomena of apatite have been described by Clifford and Hill et al. (1996) \[21\] and Moisescu et al. (1999) \[56\]. This glass-ceramic with its needle-like apatite crystals is used as a veneering material for high-strength ZrO$_2$ frameworks or lithium disilicate glass-ceramics in restorative dentistry (IPS e.max® Ceram, Ivoclar Vivadent AG, Liechtenstein) \[66\]. The veneering process is carried out by sintering or molding the material in question. Lithium disilicate frameworks are usually used in the fabrication of dental crowns and small three-unit bridges. Frameworks made with ZrO$_2$ are suitable for the manufacture of crowns and long-span bridges.

### 6.2. Leucite-containing Glass-ceramics

The framework silicate (KAlSi$_2$O$_6$) is produced according to the mechanisms of surface nucleation and surface crystallization in glasses from the SiO$_2$–Al$_2$O$_3$–Na$_2$O–K$_2$O chemical system. These products are used to produce powder compacts. The crystals demonstrate polysynthetic twinning \[36\] and residual stresses are introduced into the glass matrix due to their high coefficient of thermal expansion. As a result, flexural strengths of 120–180 MPa (biaxial strength) are achieved. The glass-ceramic can be produced as a powder compact and sintered as a veneering product. However, molding represents a special processing technique, as it makes use of the viscous flow properties of glass-ceramics.

Important products from this category of materials include biomaterials for restorative dental applications, such as products from the IPS Empress® range (Ivoclar Vivadent AG, Leichtenstein). Furthermore, mass-produced dental inlays made of these materials are suitable for replacing amalgam fillings. Crowns represent an additional important application.

### 6.3. Lithium Disilicate Glass-ceramics

The high-strength and high-toughness glass-ceramic described in Section 5.3 is used in restorative dentistry to fabricate dental frameworks that are subsequently veneered with apatite glass-ceramics. They are also suitable for producing self-supporting metal-free structures. Two processing techniques have shown to be particularly successful: molding (IPS e.max® Press) and CAD/CAM machining (IPS e.max® CAD). Both of the products mentioned are manufactured by Ivoclar Vivadent AG, Liechtenstein \[15,41,62,68\]. Molding allows the lithium disilicate glass-ceramic to be processed at 920 °C by taking advantage of the viscous flow phenomenon. The complicated shapes of dental restorations are reproduced in detail with this dental laboratory method. The machining technique, however, is not restricted to the dental lab. It can also be used chairside in the dental clinic (apparatus: e.g. Cerec®, Sirona, Germany). In both cases, the process is carried out in two steps. First, a preliminary product made of a lithium metasilicate glass-ceramic (see blue ceramic in Figure 9) is machined. After having been shaped and then fitted in the mouth of the patient, the restoration is heat-treated at 850 °C. At this temperature, the lithium metasilicate is transformed into lithium disilicate and the desired dental shade is obtained.

### 6.4. ZrO$_2$-Containing Glass-Ceramics

The crystal phase of Li$_2$ZrSi$_6$O$_{15}$ was formed in the glass-ceramic system of SiO$_2$–Al$_2$O$_3$–Li$_2$O–P$_2$O$_5$–ZrO$_2$. This

---

**FIGURE 8** SEM of a fluoroapatite glass-ceramic (Ivoclar Vivadent AG, Liechtenstein). (Reproduced from Holland and Beall, 2002 \[35\] with permission of The American Ceramic Society, www.ceramics.org. Copyright 2002. All rights reserved.)

**FIGURE 9** Glass-ceramics for dental restoration: (a) dental crown of lithium metasilicate glass-ceramic as precursor product (blue ceramic) and (b) lithium disilicate glass-ceramic as final dental crown, Dentist: A. Peschke (IPS e.max®, Ivoclar Vivadent AG, Liechtenstein). For color version of this figure, the reader is referred to the online version of this book. (Reproduced from Holland and Beall, 2012 \[36\]; with permission of Wiley)
glass-ceramic exhibits a strength of 160 MPa (biaxial strength) and it can be molded and fused to polycrystalline ZrO₂ ceramics. As a result, a composite material featuring high-strength sintered ceramics and glass-ceramics is produced. These materials are used [67] in the form of dental posts for building up devitalized teeth after root canal treatments (IPS Empress® Cosmo, Ivoclar Vivadent AG).

6.5. Glass-Ceramics With Two Crystal Phases

Apart from the apatite–wollastonite glass-ceramics presented in Section 6.1, two further glass-ceramics must be highlighted. They include glass-ceramics with mica and apatite and those with leucite and apatite. In the latter glass-ceramic, the apatite crystals may show chemically different crystal types.

Based on the successful use of glass-ceramics with curved mica crystals (BIOVERIT® II, Vitron, Germany) in the field of ENT medicine, the next challenge was to achieve controlled twofold crystallization. Mica and apatite glass-ceramics are produced by this type of twofold controlled nucleation and crystallization. The base glass consists of three glass phases: an SiO₂-rich glass matrix, an alkali and F-rich small droplet (ball-shaped) phase, and a larger CaO–P₂O₅–F-rich droplet phase. Apatite forms homogeneously in the CaO–P₂O₅–F-phase and mica forms as a result of solid-state reactions between the glass matrix and the precipitated small droplet phase [75].

Leucite–apatite glass-ceramics were developed in the SiO₂–Al₂O₃–Na₂O–K₂O–CaO–P₂O₅–F system. This base system is similar to those of the mentioned apatite ceramics. However, the composition ranges are considerably different (details see Höland and Beall, 2012 [36]). The phase-formation mechanism is characterized by twofold nucleation and crystallization via surface and internal nucleation. For this purpose, granulated glass was used to produce powder compacts. The framework silicate leucite (KAl₂Si₂O₆) is precipitated on the glass grains by surface nucleation and crystallization, and fluoroapatite (Ca₅(PO₄)₃F) develops by internal nucleation and crystallization. The solid-state reaction of sintering takes place in parallel to the crystallization process. As a result, the powder compacts are formed. A typical microstructure of a leucite–apatite glass-ceramic is shown in Figure 10. In order to reveal the small needle-like apatite crystals in SEM analysis, the samples are prepared with hydrofluoric acid. In the process, the SiO₂-rich glass matrix is etched and the needle-like apatite becomes visible. The interface around the leucite crystals is also attacked.

The two-phase crystallization of leucite and apatite allows the favorable properties of the glass-ceramics to be combined. Leucite contributes to a high CTE of $11 - 2.5 \times 10^{-6} \text{ K}^{-1}$ and apatite controls the light absorption to produce a translucent appearance. As a result of these excellent characteristics, this material has been used to develop a dental biomaterial for porcelain-fused-to-metal restorations (IPS d.SIGN® , Ivoclar Vivadent AG).

7. PRODUCTS WITH SPECIAL ELECTRICAL AND MAGNETIC PROPERTIES

7.1. Cordierite

Glass-ceramics with the ring silicate cordierite (Mg₂Al₄Si₁₀O₁₈) have been developed in the multicomponent non-stoichiometric SiO₂-Al₂O₃-MgO-TiO₂ system by Beall (1993) [8] and McDowell and Beall (1990) [51]. The nucleating agent TiO₂ enabled the controlled crystallization of a preliminary phase of the MgTi₂O₅ type and the subsequent crystallization of cordierite. This glass-ceramic exhibits a KIC toughness of 2 MPa·m⁰.⁵ and a thermal conductivity of 3.75 W m⁻¹ K⁻¹ at a CTE of $4.4 \times 10^{-6} \text{ K}^{-1}$.

The product name of this glass-ceramic is Corning 9606® (Corning, Inc., USA). It is used to produce radomes

![Figure 10](image-url)
in the aerospace industry and electronic packages in the electronics industry.

Further cordierite glass-ceramics were discovered by Budd (1993) [19] and Partridge et al. (1989) [58]. This type of material is used by GEC Alsthom, UK, for electronic packaging. Glass-ceramics with the stoichiometric composition of cordierite in combination with ZrO₂ are used to produce inorganic–inorganic composites [69]. These products from Schott, AG, Germany, are used in computer technology.

7.2. Ferrite and Mica + Ferrite Glass-ceramics

Glass-ceramics with ferrite and mica crystals have been developed in the SiO₂–Al₂O₃–Fe₂O₃–FeO–MgO–Na₂O–K₂O–F system. Depending on the composition of these base glasses, they may be supersaturated with iron ions. As a result, spontaneous crystallization of iron oxides may occur while the glass melts cool. However, ferrite (e.g., MgFe₂O₄) can also be produced in parallel to the formation of the mica phase. In the microstructure of this glass-ceramic, the ferrite crystal phases are embedded between the mica crystals in the form of cubic crystals. Therefore, a magnetically soft material is formed with a characteristic hysteresis curve.

7.3. Perovskite-type Glass-ceramics

Perovskite glass-ceramics have been developed in the SiO₂–Nb₂O₅–Na₂O, SiO₂–Al₂O₃–TiO₂–PbO and SiO₂–Al₂O₃–K₂O–Ta₂O₅–Nb₂O₅ systems (for details, see Höland and Beall 2012 [36]). Beall (1971) [4] developed a very special glass-ceramic in the SiO₂–Al₂O₃–Li₂O–Ta₂O₅ system. The main crystal phase is LiTaO₃, which actually has a structure closer to ilmenite (FeTiO₃) than perovskite (CaTiO₃). The most important characteristic of this glass-ceramic, apart from its high dielectric constant (above 40), is its transparency.

Controlled crystallization of perovskites of the KNbO₃ type was achieved by Kioka et al. (2010) [47] by laser excitation. This mechanism of initiating crystallization has also been used by Komatsu et al. (2007) [48] in other glass-ceramic systems.

8. APPLICATIONS IN ENERGY TECHNOLOGY

Sakamoto and Yamamoto (2010) [65] have presented ways of developing glass-ceramics for use in energy technology. The authors emphasized the importance of the P₂O₅–Li₂O–Fe₂O₃–Nb₂O₅ phosphate system for this purpose. The main crystal phase of this glass-ceramic was first described by Hirose et al. (2007) [32] as the orthophosphate phase LiFePO₄. The possibility of Li⁺ ion transport should enable the material to be used in the production of rechargeable batteries.

Glass-ceramics in the SiO₂–B₂O₃–MgO–BaO–ZnO and SiO₂–CaO–SrO–ZnO–B₂O₃–TiO₂ systems contain a main crystal phase of barium silicate, barium magnesium silicate, or magnesium silicate [59]. These materials are intended for use as solid oxide fuel cell components.

REFERENCES


Chapter 5.1 Glass-Ceramics


New Glasses for Photonics

Shunsuke Murai, Koji Fujita and Katsuhisa Tanaka
Department of Material Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo, Kyoto, Kyoto 615-8510, Japan

1. INTRODUCTION

Advance in semiconductor technology including integrated circuit (IC) and large-scale integration (LSI) allows the emergence of information technology society in the late 20th century. For example, a computer, which was so huge that it occupied the entire room at its infancy, now becomes very compact to be a laptop and a mobile. Considering the success of electronics in promoting the progress in information technology, the late 20th century is just an era of electronics.

The integration density of the electronic circuit has increased in the last 40 years following Moore’s law; the number of transistors that can be placed on an IC tip has doubled approximately every two years. The information content an IC can process has increased accordingly. This trend has been maintained based on the progress in nano-fabrication technology that allows the downsizing of electronics devices such as transistors. However, it is predicted that the trend does not continue anymore as it has done, because the size of transistors reaches a lower limit below which the quantum tunneling effect of electrons prevents the accurate calculation, i.e., there exist limits in the way and speed of information processing by using electrons alone as carriers of information.

One possible solution to overcome these limitations is an incorporation of photons into electronics to process information (referred to as optoelectronics or photonics). One advantage in making use of a photon as a carrier of information lies in its speed. Let us take a look at the characteristics of light (photons) and electrons as a signal. In Figure 1 shown is the typical timescales of response of the bio-, electronic, and photonic systems [1]. Bio-systems, for example, eyes of human, can respond as fast as $10^{-9}$ s, whereas the response in electronic system can be in a range of $10^{-12}$ s, i.e., $10^{10}$ times faster than the human eye. The response in photonic system is in principle much faster than the electronic systems, the response time being on the scale of $10^{-15}$ s. Also, photons have an advantage that they can be utilized in parallel processing, which is impossible for electronic systems. Let us imagine a situation in Figure 2, where you have two routes from A (a present position) to B (a destination): mountainous and...
In an electric circuit, this corresponds to the situation that you have two resistances in parallel, and one has a higher resistivity than the other. In a photonic circuit, the mountainous path means an optical path with a narrow slit or a nearly closed shutter. Then, what will happen if electrons and photons travel in this circuit? The electric circuit is split at a bifurcation point in accordance with the resistances in the routes. In this sense, the electrons look at the whole circuit and select the route to take. Thus, the fluxes of electrons (\(\equiv\) currents) in these two routes are not equal to each other unless the resistances are the same. In contrast, in the case of photonic circuits, the fluxes of photons in the two optical paths depend only on the configuration at the bifurcation point. For example, the flux of light is split equally when you put a splitter with a 1:1 separation ratio. This means that various kinds of processes can be operated for the same light or the same signal, which is called parallel proceeding of the light.

Also, the light can carry a great deal of information in a multiplexing way via its domains of wavelength, phase, and polarization. Optical fiber communication, which is presently an inevitable infrastructure of the society, is the technology that makes use of the advantage of light as a carrier of information. For the achievements concerning the light transmission in fibers for optical communications, Charles Kao received the Nobel prize in physics in 2009.

The processing of light signals is necessary in photonics just as in the case of electronics. Here, the processing includes a wavelength conversion, a light amplification, a memory, a separation, and a switch. The element technologies of photonics are summarized in Figure 3. One milestone for photonics may be the integration of these elements into one package, i.e., IC and LSI of photons. It will lead to the development of photonic computers, which can do calculation much faster and deal with more information than the present computers; the development of optical computer will progress by replacing electronic devices with photonic counterparts.

For the purpose of developing the devices for photonics, many materials have been synthesized and examined thus far. One group is inorganic crystalline materials including single crystals, nanoparticles, and thin films. From a point of view of properties, optical functionalities have been studied for many semiconductors, dielectrics, magnets, and phosphors for practical applications. Examples include a semiconductor laser diode (LD) for the pick-up of optical storages, a magneto-optical crystal for an isolator in optical communication, and phosphors for color television and illumination.

In comparison with crystalline materials, amorphous materials such as inorganic glass and organic polymers have several practical advantages. They are mostly transparent in the visible wavelengths and can be easily fabricated into any aimed shapes (e.g., fibers). Also, they can dissolve various optical functional species such as rare-earth ions. Inorganic glass is thermally and chemically stable so that it is widely applicable to photonics. It is well known that inorganic glasses have been practically utilized in various fields of optics such as lenses, filters, and prisms. Some of glass materials are also used as lasers, optical fibers, and optical memories.

In this chapter, we focus on the inorganic glasses as photonics materials and describe their optical functionalities [2]. We mainly mention nonlinear optical glasses and magneto-optical glasses based on our recent works. Other important glass materials for photonics including birefringent glasses, upconversion glasses, and optical memory glasses are also described.

2. NONLINEAR OPTICAL GLASS

Glass is a prototype of an optically isotropic material and possesses inversion symmetry from a macroscopic point of view. Thus, glasses are generally considered not to show second-order nonlinear optical effects. However, second-harmonic generation (SHG), one of the typical second-order nonlinear optical effects, was observed for a GeO\(_2\)-doped silica glass fiber after an irradiation with a laser. It was clarified that the optimal amount of GeO\(_2\) doping results in efficient SHG. The origin of this effect is considered to be an internal DC electric field generated by charged defects, which are periodically arranged due to the interference between fundamental and second-harmonic waves as described below in detail. Also, a silica glass plate after application of DC electric field at high temperatures (the process is known as thermal poling) shows SHG. In the case of thermal poling, formation of an internal electric field by the migration of cations is one possible origin for SHG.

Third-order optical nonlinearity can be observed in any materials with any symmetry. The origin of the third-order nonlinearity is largely classified into two
2.1. Fundamentals of Nonlinear Optics

Nonlinear optics treats a discrepancy from proportional relation between the intensity of electric field of light incident on a material and the polarization of the material induced by the light irradiation. Because light is an electromagnetic wave, irradiation of a material with light exerts electric field \( E \) on the material. Electric flux density \( D \) is expressed by using polarization \( P \) as
\[
D = \varepsilon_0 E + \mathbf{P},
\]
where \( \varepsilon_0 \) is the dielectric constant in vacuum. \( P \) and \( E \) are usually related to each other as
\[
P = \chi^{(1)} \varepsilon_0 E.
\]

By coupling Eqns (1) and (2), we obtain
\[
D = \varepsilon_0 E + \chi^{(1)} \varepsilon_0 E = \varepsilon E,
\]
where \( \chi^{(1)} \) is an electric susceptibility. The equations indicate that the polarization is proportional to the electric field. However, this situation is not always correct, especially when the incident light intensity is high enough. Taking this nonlinear effect into account, Eqn (2) is modified to
\[
P = \chi^{(1)} \varepsilon_0 E + \mathbf{P}^{NL},
\]
where \( \mathbf{P}^{NL} \) is a nonlinear polarization, and is expanded as
\[
\mathbf{P}^{NL} = \chi^{(2)} \varepsilon_0 E \cdot E + \chi^{(3)} \varepsilon_0 E \cdot E \cdot E + \cdots,
\]
where \( \chi^{(2)} \) and \( \chi^{(3)} \) are referred to as second- and third-order nonlinear susceptibilities, respectively.

2.2. Second-order Nonlinear Optics

Let us consider the situation when two electromagnetic waves having angular frequencies of \( \omega_1 \) and \( \omega_2 \) are incident on a material with a nonzero \( \chi^{(2)} \). Electric components of the electromagnetic waves are written as follows:
\[
E_1(\omega_1) = \frac{E_1^1 \exp[i(k_1 \cdot r - \omega_1 t)] + E_1^2 \exp[-i(k_1 \cdot r - \omega_1 t)]}{2},
\]
\[
E_2(\omega_2) = \frac{E_2^1 \exp[i(k_2 \cdot r - \omega_2 t)] + E_2^2 \exp[-i(k_2 \cdot r - \omega_2 t)]}{2}.
\]

Here, \( E_1^\ast \) and \( E_2^\ast \) mean the complex conjugates of \( E_1 \) and \( E_2 \), respectively. Second-order nonlinear polarization \( \mathbf{P}^{(2)NL} \) contains terms such as \( \chi^{(2)} \varepsilon_0 E_1(\omega_1) E_2(\omega_2) \), which yields
\[
\chi^{(2)} \varepsilon_0 E_1(\omega_1) \cdot E_2(\omega_2) = \frac{1}{2} \chi^{(2)} \varepsilon_0 \left\{ \frac{E_1^1 E_2^2 \exp[i(k_1 + k_2) \cdot r - i(\omega_1 - \omega_2)t]}{2} + \frac{E_1^2 E_2^1 \exp[-i(k_1 + k_2) \cdot r + i(\omega_1 + \omega_2)t]}{2} \right\}
\]

This equation contains the electromagnetic waves having angular frequencies of \( \omega = \omega_1 + \omega_2 \) and \( \omega = \omega_1 - \omega_2 \). The former is called sum frequency generation and the latter difference frequency generation, respectively, especially when \( \omega_1 = \omega_2, \omega = 2\omega_1 = 2\omega_2, \) and \( \omega = \omega_1 - \omega_2 = 0 \). The former is denoted as SHG and the latter optical rectification.

In general, \( \chi \) is a tensor. For example, \( \chi^{(1)} \) is given by
\[
\chi^{(1)} = \begin{pmatrix}
\chi_{11}^{(1)} & \chi_{12}^{(1)} & \chi_{13}^{(1)} \\
\chi_{21}^{(1)} & \chi_{22}^{(1)} & \chi_{23}^{(1)} \\
\chi_{31}^{(1)} & \chi_{32}^{(1)} & \chi_{33}^{(1)}
\end{pmatrix} = \left( \chi_{ij}^{(1)} \right)
\]

Likewise, \( \chi^{(2)} \) and \( \chi^{(3)} \) are given by \( \chi_{ijk}^{(2)} \) and \( \chi_{ijkl}^{(3)} \), respectively. Generally, \( \chi^{(2)} \) is zero for materials with a macroscopic inversion symmetry. The reason is explained.
as follows. First, second-order nonlinear polarization can be written as

$$P^{(2)NL}_i = \sum_{j,k} \chi^{(2)}_{ijk} \varepsilon_0 E_j E_k \quad (i = 1, 2, 3).$$  \hspace{1cm} (10)

When you rotate the object around the origin, $E_j$ and $E_k$ become $-E_j$ and $-E_k$, respectively, and $P^{(2)NL}_i$ is converted into $-P^{(2)NL}_i$. Moreover, $(\chi^{(2)}_{ijk})$ is changed into $(\chi^{(2)}_{ijk})$:

$$-P^{(2)NL}_i = \sum_{j,k} \chi^{(2)}_{ijk} \varepsilon_0 E_j E_k \quad (i = 1, 2, 3).$$  \hspace{1cm} (11)

When the object possesses the center of inversion, $(\chi^{(2)}_{ijk}) = \chi^{(2)}_{ijk}$. On the other hand,

$$\chi^{(2)}_{ijk} = -\chi^{(2)}_{ijk}. \hspace{1cm} (12)$$

from Eqs (10) and (11). This means that $\chi^{(2)} = 0$ for materials with macroscopic inversion symmetry.

To obtain intense SHG output, we should satisfy the condition of phase matching. The intensity of SHG light traveling a distance $l$ in a material, $I_{2\omega}(l)$, is given by

$$I_{2\omega}(l) = \frac{\omega^2}{8\varepsilon_0 \sqrt{\varepsilon_0}} \chi^{(2)}(\omega) E(\omega) E(\omega) \left[ \frac{2\sin(\Delta k l/2)}{\Delta k} \right]^2,$$

where $c$ is the velocity of light and

$$\Delta k \equiv 2k(\omega) - k(2\omega). \hspace{1cm} (14)$$

Here, $k(\omega)$ and $k(2\omega)$ are the wavevectors along $l$ direction of the incoming light with a angular frequency of $\omega$ and SHG light with $2\omega$. Generally, $2k(\omega)$ is not equal to $k(2\omega)$ because of the dispersion of refractive index in the material. $\Delta k$ is called a phase mismatching. According to Eqn (13), $I_{2\omega}(l)$ is decreased with an increment of $\Delta k$, while

$$\left[ \frac{2\sin(\Delta k l/2)}{\Delta k} \right]^2$$

becomes $l^2$ when $\Delta k = 0$. Thus, an intense SHG can be obtained under the condition where $\chi^{(2)}$ is large and $\Delta k = 0$. It is known from the rigorous research on organic materials that a large $\chi^{(2)}$ is obtained when the electric dipole moment induced on a molecule is large and the induced moments are aligned along one direction. On the other hand, $\Delta k = 0$ can be satisfied by using ordinary and extraordinary light; it is possible to find the condition of phase matching by making use of the difference in refractive indices for ordinary and extraordinary light.

### 2.3. Nonlinear Optical Effect in Glass Materials

The nonlinear optical effect was considered to be detrimental to high-output laser systems. This is because intense electric field inside the laser cavity modifies the refractive index of the materials, which eventually alters the resonant condition of the cavity. The relation between the cavity length $L_c$ and the wavelength, $\lambda$, of a standing wave is described as

$$L_c = \frac{a\lambda}{2} = \frac{ac}{2\nu n},$$

where $a$, $c$, $\nu$, and $n$ are a positive integer, the velocity of light, the frequency of light, and the refractive index, respectively. This equation means that $\lambda$ depends on $n$, which varies with the intensity of light due to the nonlinear optical effect. To overcome this problem, glasses having low nonlinear optical effects have been pursued as a material for laser rods. On the other hand, considerable attention has been paid to glasses showing a large optical nonlinear effect after it was recognized that the nonlinear optical effects can be utilized to process the light signals. Also, it was revealed in the 1980s that some glasses show second-order optical nonlinearity under certain conditions. The glasses showing second-order optical nonlinearity include silica and aluminosilicate glasses containing GeO$_2$, Eu$_2$O$_3$, and CeO$_2$ [3,4]. It was also reported that SiO$_2$, TeO$_2$-based, and P$_2$O$_5$-based bulk glasses and amorphous thin films show second-order optical nonlinearity after thermal poling, i.e., application of DC electric field at high temperatures [5–19].

### 2.4. Second-Harmonic Generation in Fibers

In 1986, Österberg and Margulis irradiated the Ge–P-doped SiO$_2$ glass fiber (1 m long) with a Nd:YAG laser (fundamental wave, the wavelength of which is 1064 nm), and found that green light with a wavelength of 532 nm, corresponding to the second harmonic of 1064 nm, was emitted from the fiber [3]. The intensity of 532 nm output was increased with the irradiation time of fundamental wave, and the conversion efficiency reached 5% after 10 h passed. Later, Stolen and Tom reported that the intensity of second-harmonic wave was saturated after 10 min when both of the 532 nm light and the fundamental wave were simultaneously incident [20]. It was also shown that no SHG was observed for the SiO$_2$ glass fibers without GeO$_2$, indicating that the origin of SHG is related to the presence of Ge. Tsai et al. [21] investigated the relation between electron spin resonance (ESR) spectrum and SHG intensity of the Ge-doped SiO$_2$ glass, and revealed that the SHG intensity increased with an increase in the concentration of Ge$^+$ center, a threefold-coordinated Ge defect with an unpaired electron, i.e., a singly occupied dangling bond [Figure 4(a)]. The formation mechanism of Ge$^+$ center in the glass has been intensively studied [22]. Ge-doped SiO$_2$ glass shows an absorption band at 5 eV (248 nm) due to two defects of Ge as shown in Figures 4(b) and 4(c). The defect shown in Figure 4(b) is called monovacancy where an oxygen atom is missing to form Ge-Ge bond, and the defect shown in Figure 4(c) is called divacancy where Ge is
two-coordinated. The monovacancy can be excited by a 5 eV light to yield a GeE' center and an electron. The GeE' center absorbs a 5.7 eV (218 nm) light and does not contribute to the absorption around 5 eV. An electron generated by this process is subsequently trapped by four-coordinated Ge. On the other hand, the divacancy absorbs the 5 eV light and decays radiatively, showing an intense blue emission. In a typical fiber, the amount of divacancy is 90% of the total amount of Ge-related defects, and that of monovacancy is 10%. It is believed that when the glass is irradiated with a visible laser, the monovacancy yields a GeE' center, a three-coordinated Ge with a hole, and a four-coordinated Ge with an electron through the multi-photon absorption, and such a hole—electron pair can be the origin of SHG. The concentration of GeE' center estimated from the ESR measurements is about $2 \times 10^{-4}$ mol% [21].

On the other hand, SHG is also observed when Al$_2$O$_3$–SiO$_2$ glasses containing rare-earth ions such as Eu and Ce are irradiated with a 488 nm light from an Ar$^+$ laser. The origin of SHG is considered to be photoionization of these ions due to two-photon absorption. For example, Ce$^{3+}$ shows absorption at around 4.2 eV (292 nm) and is photoionized to Ce$^{4+}$. The electron that released by this ionization makes a hole—electron pair to contribute to SHG.

Two models are proposed as the mechanism by which the hole—electron pair leads to the SHG. One is that the electric dipole induces the internal DC electric field and yields $\chi^{(3)}$ through the interaction with third-order nonlinear susceptibility $\chi^{(3)}$ as follows [23]:

$$\chi^{(3)} = 3\chi^{(3)}(0, \omega, \omega) E_{dc}. \quad (16)$$

To induce $\chi^{(3)} = 5.3 \times 10^{-12}$ esu that is a typical value observed for optical fibers, $E_{dc}$ of $10^4$ V/cm is required given that $\chi^{(3)}$ of SiO$_2$ glass is $2.8 \times 10^{-14}$ esu. The other mechanism proposed is that the inversion symmetry of the glass is lost by the presence of electric dipole and the periodic alignment of dipoles results in the SHG [24]. This means that the $\chi^{(3)}$ is indeed nonzero in the glass.

As mentioned above, the phase-matching condition should be satisfied in order to observe SHG. As shown in Eqn (13), the intensity of second-harmonic wave, $I_{2\omega}(l)$, is expressed as

$$I_{2\omega}(l) \propto \left| \chi^{(2)}E(\omega)E(\omega) \right|^2 \frac{2 \sin(\Delta kl/2)}{\Delta k}^2 \quad (17)$$

The $l_c = 2\pi/\Delta k$ of the glass fiber here is of an order of 30 μm, meaning that SHG is observed only in a short fiber (length $\approx l_c$) when the phase-matching condition is not satisfied. However, the SHG is actually observed even in fibers longer than one meter. This means that some mechanisms exist to satisfy the condition of phase matching. This can be explained based on the formation process of defects, i.e., the origin of SHG. The intensity of electric field is high at the position where the fundamental and second-harmonic light are in phase and interference constructively. At such positions, multiple photon absorption and defect formation take place preferably, and thus the orientation of electric dipole occurs in the period of $l_c$, resulting in the spontaneous phase matching to enhance the SHG output.

Before moving to the other types of glasses showing SHG, we mention another important phenomenon that occurs in the fiber due to the defect formation. In 1978, Hill et al. revealed that when an optical fiber was irradiated with a 488 nm light from an Ar$^+$ laser, the reflectivity gradually increased to be 50% after several minutes passed [25]. The reflection bandwidth was as narrow as 200 MHz. The origin of this reflection is closely related to the origin of SHG in an optical fiber. When the Ge-doped SiO$_2$ glass is irradiated with a visible laser, the two-photon absorption process takes place at 5 eV. This reduces the concentration of monovacancy, resulting in a decrease of the absorption intensity at 5 eV. On the other hand, photoreaction of monovacancy yields GeE' center and gives rise to a new absorption band at 6.3 eV. This change in absorption results in a variation of real part of refractive index through the Kramers–Kronig relation. Interference of incoming light with a reflected light from the edge of the fiber makes a standing wave and the refractive index varies periodically in accordance with the intensity distribution [26]. This provides the increasing reflection as the elapse of irradiation time. This periodic distribution of refractive index induced in the optical fiber is called Hill’s gratings.

### 2.5. Second-Harmonic Generation in Glasses after Poling

In 1991, Myers et al. found that the SHG arose in a SiO$_2$ glass plate when a DC electric field was applied to the
glass at 250–300 °C and then the glass was cooled to room temperature while the electric field being applied [5]. The process is called poling (especially thermal poling). Okada et al. performed the poling on SiO₂ glass thin film using a corona discharge and confirmed the occurrence of SHG [6,7]. The mechanism of the SHG is still an open question, although several mechanisms have been proposed, including the migration of Na⁺ ion and the formation of SiE’ center.

Nasu et al. carried out the following experiment focusing on OH⁻ and H⁺ ions as migration ions [8]. They prepared SiO₂ glass samples via sol–gel method and measured the SHG intensity by using the Maker fringe method [27,28]. SHG was observed in the SiO₂ glasses after thermal poling. The analysis of the Maker fringe pattern revealed that the thickness of the region which contributes to the SHG is equal to the thickness of the glass sample, so that they thought that the region poled effectively is extended throughout the sample. They found that the intensity of SHG and the concentration of OH⁻ are related to each other as follows:

\[ I_{2\omega} \propto [\text{OH}^-]^m, \]  

where \( m \) is a constant. On the other hand, no correlation was observed between \( I_{2\omega} \) and the concentration of paramagnetic defects estimated from the ESR measurements. This implies that the SHG is not ascribable to E’ center (SiE’ in this case) but due to the orientation of electric dipoles related to OH⁻ and the conduction of H⁺.

Fujiwara et al. used UV laser instead of heat in the poling process to increase the second-order optical nonlinearity induced in a glass [29]. They irradiated the glass having a composition of 15.7GeO₂·84.3SiO₂ with a 193 nm light from an ArF excimer laser while applying a DC voltage of 30 kV/mm. The value of second-order nonlinear susceptibility thus obtained is close to the \( \chi^{(2)} \) value of LiNbO₃, a representative optical second-order nonlinear crystal. It is believed that Ge-related photoreaction occurs by the irradiation of the UV laser, and electric dipoles formed by the photoreaction are aligned by the application of external electric field.

As an example of optical second-order nonlinearity in glasses other than SiO₂ glass, we mention here the SHG in TeO₂-based (tellurite) glasses after poling. The tellurite glass is known to possess a large dielectric constant [30], and a high transmission in the infrared regime [31], and a high third-order nonlinear susceptibility. Thus far, many glass systems including ZnO–TeO₂, WO₃–TeO₂, Na₂O–TeO₂, Li₂O–Nb₂O₅–TeO₂, and Na₂O–Al₂O₃–TeO₂ [32] have been demonstrated to show SHG after thermal poling process. Among them, ZnO–TeO₂ glass shows a relatively stable second-order nonlinearity as indicated by a long relaxation time [33]. Also, WO₃–TeO₂ glass exhibits a relatively large \( \chi^{(2)} \) value [34]. For the tellurite glass with alkaline and divalent cations, it has been argued that the migration of mobile charged ions during the thermal poling process causes the formation of an internal electrostatic field, which contributes to the second-order nonlinearity via the mechanism discussed in Eqn (16). In most of the thermal poling experiments, a thin charge depletion layer is created near the anode-side glass surface since the mobile charged ions are often cations such as Na⁺.

2.6. Phenomenon Related to Thermal Poling

In addition to the optical second-order nonlinearity, the thermal poling process can bring about interesting phenomena into glasses. Kazansky et al. reported that bleaching took place in a silicate glass containing Ag nanoparticles when thermal poling was applied [35,36]. In their experiment, with an application of a DC electric field, Ag nanoparticles in the vicinity of anode-side glass surface are selectively oxidized into Ag⁺, converting a region under the electrode from dark green to transparent. The mechanism of bleaching can be interpreted in terms of electromigration, or electrolysis of Ag metal in the vicinity of the anode and subsequent migration of Ag⁺ [36]; due to the intense electric field in the depletion layer underneath the anode, Ag particles release electrons to become Ag⁺. This technique can be applied to the formation of nano- and microstructures on the surface of glass. Murai et al. performed the thermal poling for a tellurite glass containing Ag⁺ and found that Ag nanoparticles were selectively precipitated in the vicinity of anode-side glass surface [37]. A typical result is displayed in Figure 5. Although this phenomenon appears to be similar to the above-mentioned poling-assisted bleaching in respect that the reaction relevant to Ag occurs beneath the anode-side glass surface, the physical background is very different; Ag⁺ is reduced into Ag metal in the case of tellurite glass, whereas Ag metal is oxidized into Ag⁺ in the case of silicate. It is surprising that the reduction takes place at the anode side, where oxidation is usually observed. In the thermal poling process for the tellurite glass, alkali-containing cover glasses were inserted in between the tellurite glass sample and electrodes to avoid an undesired discharge between the electrodes. XPS measurements indicate that Na⁺ in the cover glass is incorporated into the anodic surface of tellurite glass sample during the thermal poling. As a result, the glass composition is modified in the vicinity of anode-side surface to cause thermal reduction of Ag⁺ into metallic Ag. The phenomenon observed is analogous to the ion implantations via vapor or liquid phases, where charged species are introduced in a solid, and could be referred to as “poling-assisted ion implantation”.

MAGNETO-OPTICAL GLASS

Oxide and fluoride glasses containing a large amount of rare-earth ions and heavy metal ions have been intensively studied as glass materials that show a large Faraday rotation in the visible region. The Faraday effect is one of the magneto-optical effects, and hence, from such a point of view, these oxide and fluoride glasses are one kind of amorphous magnets. Thus far, research on amorphous magnets has been mainly restricted within the field of amorphous alloys, while there are not so many investigations on magnetic properties of oxide and fluoride glasses. This is mainly because the magnetization is weaker and the magnetic transition temperature is lower for oxide and fluoride glasses than for their crystalline counterparts. Moreover, the amorphous oxide and fluoride magnets have found few practical applications. Due to the disordered structure of glass, it is difficult to construct a long-range-ordered magnetic phase. However, under an external magnetic field, the transparent oxide and fluoride glasses show an excellent magneto-optical effect in the visible and ultraviolet regions so that they can be applied to optical isolators for shorter wavelength lasers.

3.1. Magnetic Properties of Amorphous Solids

Amorphous materials showing ferromagnetism and ferrimagnetism include alloys of a transition metal and a metalloid, and of a transition metal and a rare earth. In these materials, a long-range order exists in the ensemble of magnetic dipole moments although the arrangement of constituting atoms is disordered. Thus far, many amorphous alloys have been prepared and their magnetic properties and structures have been examined. According to the previous studies, the ferromagnetic amorphous alloys possess well-defined Curie temperatures, and the second-order phase transition between ferromagnetic and paramagnetic phases is observed, where the critical exponents can be defined [38]. The ferromagnetic states and the magnetic transition are intrinsically the same as those of ferromagnetic crystalline materials. The ferromagnetic amorphous alloys can be used as a magnetic core and a magnetic head because of their excellent soft magnetic properties.

On the other hand, antiferromagnetic interaction is usually dominant between magnetic moments in amorphous ionic materials such as oxide and fluoride glasses. Namely, one magnetic moment is oriented antiparallel to its adjacent moments. Moreover, one magnetic ion indirectly interacts with surrounding magnetic ions via nonmagnetic ions such as oxide and fluoride ions; it is called superexchange interaction. The antiferromagnetic superexchange interactions among the magnetic moments dispersed in a disordered structure inevitably lead to so-called magnetic frustration [39]. This means that strong magnetization and high magnetic transition temperature are not achieved in the amorphous oxides in contrast to the case of ferro- or ferrimagnetic crystalline oxides. This is the main reason why the magnetic properties of amorphous ionic materials are far from application. However, they are interesting as random magnetic materials like a spin glass.

3.2. Magnetism of Amorphous Ionic Solids

Magnetic properties have been examined for the amorphous ionic solids (mainly oxide and fluoride glasses) containing a high concentration of transition metal ions, and some models for the magnetic transition have been proposed [40–62]. In Figure 6, a typical experimental result of temperature dependence of magnetic susceptibility $\chi(T)$ (for a Fe$_2$O$_3$–TeO$_2$ glass[63]) is illustrated. Here, zero-field cooling (ZFC) and field cooling (FC) refer to two different protocols for measuring $\chi(T)$. In ZFC protocol, the sample is cooled to the lowest temperature of...
measurement without applying an external magnetic field, and $\chi(T)$ is measured during the heating process after the application of external magnetic field. In contrast, in FC protocol the external magnetic field is applied at a high temperature, and $\chi(T)$ is measured during the cooling process. For the ZFC process in Figure 6, magnetic susceptibility increases with a decrease in temperature, reaches a maximum at a certain temperature, and then decreases with a further decrease in temperature. This behavior is typical of either spin glass or micromagnetism [64]. A typical example of the system showing spin glass behavior is Au–Fe alloy [65]. Here, spin glass means the state in which isolated magnetic moments of extrinsic magnetic atoms are embedded in nonmagnetic matrix. In the spin glass, the individual magnetic moments are frozen at low temperatures so that the random orientation of magnetic moments is achieved. In contrast, micromagnetism is the state where the concentration of magnetic atoms is large enough to make clusters in nonmagnetic matrix.

Many ionic glasses are known to show the temperature dependence of magnetic susceptibility as shown in Figure 6. Examples include CoO–Al$_2$O$_3$–SiO$_2$ [45], Fe$_2$O$_3$–Bi$_2$O$_3$–B$_2$O$_3$ [66], and MnO–TeO$_2$ [67] systems. The common behavior of magnetic transition is schematically illustrated in Figure 7. In Figure 7(a), a temperature of maximum magnetic susceptibility is referred to as the spin freezing temperature, $T_f$. Figure 7(b) shows the reciprocal of $\chi(T)$. At temperatures well above $T_f$, the reciprocal magnetic susceptibility is proportional to the temperature, following the Curie–Weiss law:

$$\chi = \frac{NM_B^2\mu_B^2}{3\mu_0k(T-\theta)},$$

(19)

where $\mu_0$ is the magnetic permeability of vacuum, $k$ is the Boltzmann constant, $\theta$ is the paramagnetic Weiss temperature, $\mu_B$ is the Bohr magneton, $M_B$ is the effective number of Bohr magnetons, and $N$ is the number of atoms or ions per unit volume. For almost all transition metal oxide and fluoride glasses, $\theta$ is negative, i.e., the magnetic interaction is antiferromagnetic in these materials. Values of $T_f$ and $\theta$ observed for several oxide and fluoride glasses are summarized in Table 1 [45–49]. $T_f$ lies in a range of several tens of kelvins, whereas the magnitude of $\theta$ is of an order of 100 K. This means that the local magnetic interaction is strong even in the disordered oxides and fluorides.

The temperature dependence of magnetic susceptibility starts to deviate from the Curie–Weiss law as the temperature is approaching $T_f$ from the higher temperature side, as shown in Figure 7(b). In other words, the magnetic susceptibility is larger than that estimated from eq. (19), i.e., the apparent magnetic moment becomes larger than that of an individual magnetic ion.

$T_f$ varies with a timescale of observation. This is one characteristic of the magnetism of ionic glasses. When an AC magnetic susceptibility is measured, $T_f$ increases as an increment of applying AC frequency. Also, in Mössbauer effect measurements, the magnetic transition is observed.
at a temperature higher than that observed in AC susceptibility measurements. Mössbauer effect refers to the recoil-free emission and resonant absorption of γ-ray by a nucleus of an atom bound in a lattice of crystalline and amorphous solids. The spectra bring about information on the magnetism and chemical state in the materials within the timescale of nuclear Larmor precession. The frequency dependence of $T_f$ is sometimes described by the empirical Fogel–Fulcher low:

$$\nu = \nu_0 \exp \left\{ - \frac{E_A}{k(T - T_0)} \right\}, \quad (20)$$

where $\nu$ is the frequency, and $T_0$ and $E_A$ are empirical parameters. As described below, for cluster spin glasses or interacting superparamagnetic particles, $T_0$ and $E_A$ denote the intercluster interaction and the activation energy for relaxation process of magnetic clusters, respectively. Figure 8 shows frequency dependence of $T_f$ of the glass with a composition of 51.8FeO·5.8Al$_2$O$_3$·42.4SiO$_2$ [50]. The line in the figure was drawn using Eqn (20) with $T_0 = 8.0$ K and $E_A/k = 172$ K. An agreement between the line and experimental data is reasonably good.

In addition, magnetism in ionic glasses has unique features such that $\chi(T)$ depends on the measurement protocol, i.e., $\chi(T)$ is different between ZFC and FC, and a remanent magnetization varies with the time $t$ following the $\log(t)$ relation [51]. One model proposed by Verhelst et al. to explain these observations is based on superparamagnetism [45]. In this model, the presence of clusters, the regions where magnetic ions exist in a high concentration, is supposed and these clusters are considered to behave superparamagnetically. The orientation of magnetization of

---

**TABLE 1 Spin Freezing Temperatures $T_f$ and Paramagnetic Weiss Temperatures $\theta$ of Several Oxide and Fluoride Glasses**

<table>
<thead>
<tr>
<th>Glass compositions</th>
<th>$T_f$ (K)</th>
<th>Measurement method of $T_f$</th>
<th>$\theta$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.0CoO·40.0Al$_2$O$_3$·20.0SiO$_2$</td>
<td>2.5</td>
<td>AC magnetic field, 500 Hz</td>
<td>−95</td>
</tr>
<tr>
<td>50.0MnO·5.0Al$_2$O$_3$·45.0SiO$_2$</td>
<td>5.55</td>
<td>AC magnetic field, 500 Hz</td>
<td>−245</td>
</tr>
<tr>
<td>47.2BaO·31.5Fe$_2$O$_3$·21.3B$_2$O$_3$</td>
<td>12</td>
<td>DC magnetic field</td>
<td>−95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mössbauer</td>
<td></td>
</tr>
<tr>
<td>45.2Li$_2$O·11.9Fe$_2$O$_3$·42.9B$_2$O$_3$</td>
<td>3.25</td>
<td>AC magnetic field, 17.3 Hz</td>
<td>−115$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>197.6 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2103 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mössbauer</td>
</tr>
<tr>
<td>Ba$<em>{0.3}$Th$</em>{0.35}$Mn$<em>{0.55}$F$</em>{2.7}$</td>
<td>1.32</td>
<td>AC magnetic field, 75 Hz</td>
<td>−30</td>
</tr>
<tr>
<td>PbMnFeF$_2$</td>
<td>11.77</td>
<td>AC magnetic field, 75 Hz</td>
<td>−190</td>
</tr>
<tr>
<td>Pb$<em>2$MnFe$</em>{0.6}$Ga$_{0.4}$F$_8$</td>
<td>2.75</td>
<td>AC magnetic field, 20 Hz</td>
<td>−100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>750 Hz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1250 Hz</td>
</tr>
</tbody>
</table>

$^a\theta$ of 43.0Li$_2$O·17.5Fe$_2$O$_3$·39.5B$_2$O$_3$ glass

---

**FIGURE 8** Dependence of spin freezing temperature on frequency of AC magnetic field for 51.8FeO·5.8Al$_2$O$_3$·42.4SiO$_2$ glass.
clusters is frozen when the transition time of magnetization between energy minima, which is determined by the magnetic anisotropy and external magnetic field, becomes longer than that of the timescale of observation. At higher temperatures, the magnetization of localized clusters acquires the thermal energy large enough to overcome the activation energy between the configurations with energy minima. In this case, the polarization can be switched freely. On the other hand, the probability of transition decreases when the temperature is decreased. In other words, the relaxation time is prolonged at low temperatures so that the orientation of magnetization is apparently fixed. The net magnetization is decreased upon the decrement of temperature since the random orientation of magnetization of each cluster is frozen. As a result, a maximum is observed in the plots of magnetic susceptibility as a function of temperature. Also, at just the higher temperature side of $T_N$, the apparent magnetic susceptibility becomes larger than the value expected for one magnetic ion because the magnetic clusters grow and manifest superparamagnetic behavior in the temperature region.

Among amorphous ionic solids, there are unique materials that show a positive exchange interaction between the magnetic moments. Amorphous FeF$_2$ obtained by vapor-phase synthesis possesses a positive $\theta$ of 21 K [56]. The fact that amorphous FeF$_2$ shows a ferromagnetic interaction is very interesting, considering that the crystalline FeF$_2$ is an antiferromagnet with a Néel temperature of 78 K. Nonetheless, the amorphous FeF$_2$ exhibits superparamagnetism as well. The blocking temperature, or $T_B$ in the case of cluster spin glass, was reported to be 15±2 K. Some amorphous materials other than amorphous FeF$_2$ were found to show magnetic interactions that are different from their crystalline counterparts. Schoenes et al. reported that a positive $\theta = +1$ K was obtained for binary europium silicate glass with a composition of Eu$^{2+}_{0.12}$Eu$^{3+}_{0.02}$Si$_{31}$O$_{55}$, although no magnetic transition was observed down to 1.5 K [68]. Akamatsu et al. studied EuO$-$B$_2$O$_3$, EuO$-$Al$_2$O$_3$-$B_2$O$_3$, and EuO$-$Al$_2$O$_3$-$SiO$_2$-$B_2$O$_3$ glass systems, and revealed that the interaction between Eu$^{2+}$ ions was positive, i.e., $\theta$ was positive irrespective of the glass composition [69]. In particular, 60.0EuO·11.0Al$_2$O$_3$·19.5B$_2$O$_3$·9.5SiO$_2$ glass shows behavior of reentrant spin glass; it exhibits both paramagnetic–ferromagnetic and ferromagnetic-spin glass transitions at very low temperatures. They also prepared amorphous thin films of $x$EuO·$(1-x)$TiO$_2$ system by using a pulsed laser deposition technique and demonstrated that amorphous EuTiO$_3$ shows ferromagnetism although the crystalline EuTiO$_3$ is an antiferromagnet with a Néel temperature of 5 K [70]. Also, the transition temperature becomes higher by amorphization of crystalline Eu$_2$TiO$_4$; amorphous Eu$_2$TiO$_4$ is a ferromagnet with a Curie temperature of 14 K, while crystalline Eu$_2$TiO$_4$ shows ferromagnetism below 7 K. Ferromagnetic interactions between Eu$^{2+}$ ions were also reported for amorphous EuO$-$ZrO$_2$ system [71].

### 3.3. Magneto-optical Effect

The modulation of optical properties of materials by the interaction of light with magnetic field or magnetic moment is referred to as magneto-optical effect. Magneto-optical effects include the Faraday and Voigt (Cotton–Mouton) effects that modulate the transmitted light. In contrast, magneto-optic Kerr effect is related to the reflection light. Especially, the Faraday effect is applicable to optical switches, isolators, circulators, and magnetic and current sensors. On the other hand, magneto-optical Kerr effect is applicable to magneto-optical memories as demonstrated for metallic thin films.

The Faraday effect refers to the phenomenon that occurs when linearly polarized light is incident on a material where magnetic field is applied or magnetization is present, the orientation of which is parallel to the propagation direction of light. Linearly polarized light is a superposition of right- and left-handed circularly polarized light. Supposing the linearly polarized light is incident on a material parallel to the magnetic field or magnetization, the refractive index of the material for the right-hand circularly polarized light is different from that for the left-handed circularly polarized light. Consequently, the polarized plane of the transmitted light is rotated relative to the plane of the incident light.

This effect can be explained as follows from a microscopic point of view. We assume here that the material is either a ferromagnet or a paramagnet. As mentioned above, the linearly polarized light is a sum of the right- and left-handed circularly polarized light. Based on the conservation law of momentum, the change in orbital momentum of electron, $\Delta L$, must be $+1$ and $-1$ for the electronic transition accompanied by the absorption of right- and left-handed circularly polarized light, respectively. For example, as depicted in Figure 9, when the excited state is split into two energy levels corresponding to $\Delta L = \pm 1$ because of the spin–orbit interaction, the shape of the resonance spectrum depends on the handedness of circularly polarized light, i.e., either right- or left-handed. As a result, the phase and amplitude of the transmitted left-handed circularly polarized light is different from those of the right-handed circularly polarized light, and the net transmitted light, which corresponds to a superposition of these two lightwaves, becomes ellipsoidal with the plane of polarization being tilted with respect to the incoming light. When the number of up- and down-spins that are excited is the same as each other, the phase and amplitude of the transmitted left-handed circularly polarized light are just the same as those of the transmitted right-handed one because the energy corresponding to $\Delta L = +1$
transition for the up-spin is the same as that corresponding to $\Delta L = -1$ transition for the down-spin. This means that the Faraday effect can be observed only when the number of the up- and down-spins is different from each other due to the spontaneous magnetization or application of external magnetic field.

The Faraday rotation angle is proportional to the magnitude of external magnetic field or magnetization (in the case of ferromagnets) and the length through which the light propagates inside the material. The factor of proportionality is called the Verdet constant in the case of a paramagnet. Namely, the Faraday rotation angle $\theta_F$ is expressed by the product of external field $H$, light path $l$, and Verdet constant $V$ as follows:

$$\theta_F = VHL.$$  \hspace{1cm} (21)

The relation is also valid for a diamagnet. For a ferromagnet, the Faraday effect is evaluated by the rotation angle divided by the length of light path. For the application of the materials to devices utilizing the Faraday effect, low optical absorption is required in addition to the large rotation angle. Thus, the value of the Faraday rotation angle or the Verdet constant divided by the absorbance, $\alpha$, is used as a figure of merit for the evaluation of the materials.

Here, we explain the principle of an optical isolator, which is one of the main devices using the Faraday effect. The schematic setup is shown in Figure 10, wherein light transmits only in one direction (from left to right in Figure 10). A transparent magneto-optical material is set in the optical path, and the length of the material is adjusted so that the rotation angle of polarization plane of the transmitted light is exactly $45^\circ$ with respect to the incident
linearly polarized light. Two polarizers are placed in front (referred to as P-1) and back (P-2) of the magneto-optical material. The plane of polarizer P-2 is rotated by 45° with respect to that of P-1. In this configuration, the linearly polarized light transmitted through the magneto-optical material can transmit the polarizer P-2. On the other hand, a reflected light which propagates in the opposite direction is rotated by 90° with respect to the incoming light after it passes through the magneto-optical material. Thus, the light cannot transmit through the polarizer P-1 and cannot travel further in the opposite direction. In this sense, the light propagates only in one direction. This optical isolator is very important for stable operation of optical systems and has been used in lasers and telecommunication systems.

3.4. Glasses Containing a Large Amount of Rare-earth Ions

The superexchange interactions between rare-earth ions in oxides are very small compared with those working between transition metal ions. Thus, it is generally reasonable to consider the oxide glasses containing a high concentration of rare-earth ions as paramagnets. Also, most of them are transparent in the visible region unlike the glasses containing transition metal ions. As described above, the figure of merit of the Faraday effect is the rotation angle divided by the optical absorption, so that the rare-earth-containing glasses with a high Verdet constant can be applied to magneto-optical devices in the visible window of the frequency. Berger and Rubinstein measured the Faraday effect for phosphate and borate glasses containing 20 mol% of rare-earth ions to evaluate factors determining the magnitude of the Verdet constant. The Verdet constant for a paramagnet is negative by definition, so it is known that the magnitude is large for Ce³⁺, Pr³⁺, Tb³⁺, and Dy³⁺ [72,73]. The following equation deduced in the Van Vleck-Hebb theory [74] is often used for the description of the Verdet constant of rare-earth containing glasses:

\[
V = \frac{4\pi^2\chi}{g\mu_B c} \sum C_i \left(1 - \frac{\lambda_i^2}{\lambda_i^2}ight)^{-1}
\]

(22)

where \(g\) is the Landé factor, \(\mu_B\) is the Bohr magneton, \(c\) is the velocity of light, \(h\) is the Planck constant, \(C_i\) is related to the transition probability, and \(\lambda_i\) is the transition wavelength. The subscript \(i\) denotes the electronic transition which contributes to the Faraday effect. The approximation taking into account only single electronic transition has been used to explain the wavelength dependence of \(V\) for rare-earth-containing glasses [75,76]:

\[
V = A\chi C_i \left(1 - \frac{\lambda_i^2}{\lambda_i^2}ight)^{-1}.
\]

(23)

Here \(A = \frac{4\pi^2}{g\mu_B c}h\) is a constant, and \(C_i\) and \(\lambda_i\) are the effective transition probability and effective transition wavelength, respectively. According to the equations, the magnitude of the Verdet constant becomes large for the glass with a large magnetization, a large transition probability, and a long transition wavelength. A large magnetization counts for much in the case of Tb³⁺ and Dy³⁺, while a large transition probability has great influence on the Verdet constant for Ce³⁺ and Pr³⁺ [77]. Furthermore, for the glass containing Eu²⁺, the magnitude of the Verdet constant is very high because of the large magnetization and long transition wavelength [78]. A comparison is shown in Table 2. In most of the rare-earth-containing glasses, the 4f–5d transitions contribute to the Faraday effect. In the case of Eu²⁺-containing glasses, the effective transition wavelength varies according to the split of the 5d levels of Eu²⁺, which is modified by the electron density of oxide ions around the Eu²⁺. This means that one can tune the transition wavelength by controlling the composition of glass to obtain a large Verdet constant at an aimed wavelength. The crystal field splitting becomes large when the

<table>
<thead>
<tr>
<th>Rare-earth ion</th>
<th>(V) (min/cm · Oe)</th>
<th>(N) (\times 10^{21}) cm⁻³</th>
<th>(C_i) (\times 10^{-45}) J cm⁻³</th>
<th>(gJ(J + 1))</th>
<th>(\lambda_i) (nm)</th>
<th>((1 - \lambda_i^2/\lambda_i^2)^{-1}) at (\lambda = 500) nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eu²⁺</td>
<td>−0.30</td>
<td>4.1</td>
<td>3.0</td>
<td>31.5</td>
<td>384</td>
<td>−1.5</td>
</tr>
<tr>
<td>Ce³⁺</td>
<td>−0.33</td>
<td>6.0</td>
<td>28</td>
<td>7.5</td>
<td>289</td>
<td>−0.50</td>
</tr>
<tr>
<td>Pr³⁺</td>
<td>−0.26</td>
<td>5.3</td>
<td>28</td>
<td>16.0</td>
<td>210</td>
<td>−0.21</td>
</tr>
<tr>
<td>Tb³⁺</td>
<td>−0.32</td>
<td>5.4</td>
<td>7.8</td>
<td>63.0</td>
<td>215</td>
<td>−0.23</td>
</tr>
<tr>
<td>Dy³⁺</td>
<td>−0.33</td>
<td>6.2</td>
<td>8.5</td>
<td>85.0</td>
<td>175</td>
<td>−0.14</td>
</tr>
</tbody>
</table>
electron density of oxide ions is large. Thus, the magnitude of the Verdet constant becomes large when the concentration of alkaline metal ions is increased to raise the basicity of the glass [79,80]. Also it is known that Pr$^{3+}$-containing glass shows a high figure of merit in blue to UV region [81,82].

3.5. Glasses Containing a Large Amount of Transition Metal Ions

In contrast to the glasses containing rare-earth ions, glasses containing transition metal ions generally show a large absorbance in the visible region because of the d–d transitions of transition metal ions. This indicates that the glass is not suitable for the magneto-optical material in the visible range. In spite of this common belief, Akamatsu et al. prepared iron phosphate glasses that exhibited fairly high transmittance in the visible range [83]. These glasses show a large Faraday effect at a wavelength of about 400 nm. Glasses with xFeO·(100−x)P$_2$O$_5$ (mol%) ($x = 50.0, 54.0, 57.1$) compositions were prepared by a two-step melt-quenching method. The starting powders were first melted in a Pt crucible and then the glass thus obtained was remelted in a glassy carbon crucible to completely reduce Fe$^{3+}$ to Fe$^{2+}$. The valence state of iron ions was evaluated by using the Mössbauer spectroscopy, and it was confirmed that almost all the iron ions were present as Fe$^{2+}$. Intense optical absorption bands are observed in the ultraviolet and infrared wavelength ranges due to the charge transfer transition from O$^2$− to Fe$^{2+}$ and the intra-atomic d–d transition, respectively as shown in Figure 11(a). In contrast, the glass is fairly transparent in the visible range due to the absence of Fe$^{3+}$.

The wavelength dependence of the Verdet constant for the glasses is illustrated in Figure 11(b). The absolute value of $V$ monotonically increases as the wavelength of incident light, $\lambda$, becomes shorter. The relation between $V$ and $\lambda$ for the glasses was analyzed in terms of eq. (22). Here, the dependence of $V$ on $\lambda$ is described well by considering two optical transitions, i.e., the charge transfer transition and the crystal field transition in the ultraviolet and infrared regions, respectively. For instance, an analysis using Eq. (22) yields $C_1 = 8.6 \times 10^{-45}$ J cm$^3$, $\lambda_1 = 214$ nm, $C_2 = 1.1 \times 10^{-40}$ J cm$^3$, and $\lambda_2 = 1100$ nm for 57.1FeO·42.9P$_2$O$_5$ glass. The value of $C_1$ corresponding to the charge transfer transition is much larger than that of $C_2$ corresponding to the d–d transition. This fact indicates that the charge-transfer transition mainly contributes to the Faraday effect due to its large oscillator strength. Also shown in Figure 11(b) is the magneto-optical figure of merit defined as $\vert \theta_F \vert / \alpha$, where $\theta_F$ is the Faraday rotation angle and $\alpha$ is the absorbance. For all the glasses, $\vert \theta_F \vert / \alpha$ exhibits a maximum at around 380 nm. For instance, $\vert \theta_F \vert / \alpha$ of 50.0FeO·50.0P$_2$O$_5$ glass is 13° at 380 nm under $H = 15$ kOe. Since the wavelength like 380 nm corresponds to the emission of blue laser diodes, the present glasses are expected to be a candidate for optical devices such as an optical isolator in such a short wavelength region, but the figure of merit is rather low so that it must be improved by changing glass systems, compositions, and so forth. The amplitude of $V$ for the present glasses is somewhat smaller than or comparable to those for the rare-earth-containing glasses; for instance, at 405 nm, $V_{Fe} = -0.204$ and $-0.560$ min/cm·Oe for 57.1FeO·42.9P$_2$O$_5$ and 25.4Tb$_2$O$_3·74.6$P$_2$O$_5$ glasses, respectively, both of which are phosphate glasses and have almost the same effective transition wavelength [72]. One advantage of the present glasses lies in the fact that both iron oxide and phosphoric acid, i.e., the starting materials for the glasses, are much more inexpensive and available when compared with the rare-earth oxides which have been practically utilized as a component of a Faraday rotator. Considering that Fe is an abundant element in the crust of the earth,
the achievement of FeO-based glass materials which show high transmittance as well as a large Faraday effect in the near-ultraviolet to blue region is of great importance.

3.6. Diamagnetic Glasses

Glass containing neither rare-earth nor transition metal ions also shows the Faraday effect due to its diamagnetic nature. Especially, glasses containing Ti$^{4+}$, Pb$^{2+}$, and Bi$^{3+}$ exhibit a strong absorption in the visible and near UV regions, and also these elements can lead to a large diamagnetic magnetization due to their large atomic numbers and ionic radii, resulting in a large Faraday rotation angle in the visible to UV region. Moreover, the Faraday rotation angle does not depend on the temperature in the case of the diamagnetic glass, and the rotation angle can become very large by making fibers from these glasses. These points are advantageous in the sensing application.

4. OTHER IMPORTANT GLASSES FOR PHOTONICS

4.1. Birefringence Glass

As mentioned in the introduction of optical second-order nonlinearity, glass is a prototype of optically isotropic material so that in principle it does not show any phenomena that arise from optical anisotropy, such as optical birefringence as well as optical second-order nonlinearity. We have already seen that this is not always the case; for example, the application of both heat and electric field (thermal poling) gives rise to SHG. Optical birefringence, which refers to the difference in refractive index for different polarization states of light traveling through a material [84], is also observed in some cases. The phenomenon is important in photonics because it can be utilized to control the polarization state of light.

It is possible to induce optical birefringence in glasses extrinsically via geometrical and/or external-field effects [85]. The geometrical effect, which is commonly referred to as form birefringence [86], is achieved by anisotropic nanostructures such as sub-wavelength gratings. The external-field effect takes place when the glass is placed under external fields, such as stress, electric field, and magnetic field. The stress-optical effect, or photoelasticity [87], appears when a material is subjected to mechanical stress. The stress deforms the isotropic arrangement of atoms in the glass, leading to a polarization-sensitive response to the incoming light. Electro-optical effect occurs when the application of external electric field brings about ionic polarization and deformation of electron clouds of constituent ions to cause an optical anisotropy in the glass.

A metal nanostructure can act as an origin to modulate the polarization states of light in an enhanced way through the strong coupling between photons and surface plasmons [88]. Because the coupling strength depends on tiny variations in the local electronic states, an anisotropic metal nanostructure has significant influence on the frequency and polarization states of incoming light. In the case of glass-based nano-composites where metal nanoparticles are precipitated in the glass matrix, optical anisotropy can be induced by heat treatment under an external stress (uniaxial tensile), which allows the metal nanoparticles to elongate and orient in one direction. The composite glass thus obtained can be used as a polarizer that works selectively at the wavelength of localized surface plasmon resonance (LSPR).

In the case described above, it is a nanoparticle that possesses optical anisotropy. Alternatively, it is also possible to make birefringent composite materials by embedding the metal nanoparticles in an optically anisotropic matrix. Murai et al. fabricated tellurite glasses containing Ag$_2$O, Na$_2$O, and ZnO using the conventional melt-quenching method [89]. The glass was heat-treated to convert the Ag ions into metallic Ag. The Ag nanoparticles were spherical, and an ordered orientation of the Ag nanoparticles was not observed. Optical birefringence was evaluated from polarization rotation of linearly polarized light. Figure 12(a) shows a typical result of the measurement of polarization rotation. The variation in optical rotation angle, $\xi$, is plotted as a function of the wavelength. No obvious polarization rotation is observed ($\xi = 0^\circ$) for the as-prepared glass. In contrast, for the heat-treated glass,

![FIGURE 12](image-url)
a polarization rotation appears at around the frequency region corresponding to the LSPR of the Ag nanoparticles. Figure 12(b) illustrates the variation in $\xi$ at $\lambda = 494$ nm as a function of sample azimuth $\Phi$. This figure shows that $\xi$ periodically changes with $\Phi$. The azimuth dependence of the polarization rotation indicates that the period of oscillation is 180°, suggesting that the phenomenon is attributable to the optical birefringence. They ascribed the optical birefringence to an optical anisotropy of the matrix, which was enhanced by the presence of Ag nanoparticles. It was thought that an internal stress was presumably generated during the quenching of the melt, although no stress was applied externally. This causes small birefringence of the matrix via photoelasticity, although it is not clearly detected as shown in Figure 12(a) (see the result of as-prepared sample). Because of the birefringence of the matrix, the wavelength of LSPR shifts depending on the polarization state of the incident light, which in turn induces the polarization dependence of the real part of refractive index via the Kramers–Kronig relation. To verify this speculation, they have experimentally realized a situation that optical anisotropy is induced along a predefined direction in the glass matrix, where spherical Ag nanoparticles are dispersed isotropically. They have adopted the thermal poling [37] to achieve both the introduction of anisotropy and the precipitation of Ag nanoparticles simultaneously, and confirmed that the optical axes are parallel to the direction of applied electric field [90].

4.2. Upconversion Glass

Upconversion fluorescence is a process that converts a longer wavelength to a shorter one through multi-photon (usually two-photon) excitation of electrons. This phenomenon was reported for many rare-earth-containing glass systems. The principle of upconversion is schematically illustrated in Figure 13, where a two-photon process is shown [91]. Figure 13(a) depicts the process that an electron in an excited state is further excited to a higher energy level by absorbing a photon. This process is called excited state absorption. Figure 13(b) shows another process involving the energy transfer from an electron at an excited state to another excited electron. In both processes, two photons are absorbed and one photon is emitted, so that the emission wavelength is shorter than the wavelength of excitation. Upconversion fluorescence in oxide glasses was first observed by Tanabe et al. by using Er$^{3+}$ doped tellurite and gallate glasses [92]. The study of upconversion glass was triggered by the invention of fluoride glasses, which show a high fluorescent yield because of their low phonon energy.

One application of upconversion fluorescence is laser devices with shorter wavelengths. In optical memory devices, a spot of laser beam incident on the surface of compact disk corresponds to one bit, a unit of information. Thus, the size of spot governs the density of storage. Numerical aperture, NA, of lens used to focus the light with a wavelength of $\lambda$ is related to the diameter of spot, $\Delta D$, as

$$\Delta D = 1.22\lambda/\text{NA}. \quad (24)$$

This means that for the increment of optical memory density by reducing the spot size, it is necessary to realize a laser which can emit light at a shorter wavelength. A laser diode based on GaN has been developed as a short-wavelength laser. For example, Nakamura et al.
achieved CW laser oscillation at room temperature by using 420 nm emission from a GaN-based semiconductor system [93,94]. Also, a material with large optical nonlinearity such as large SHG efficiency is promising for the development of a short-wavelength laser. Compared with other materials, a short-wavelength laser based on upconversion fluorescence has following advantages. First, it does not require phase matching in contrast to the case of SHG. Second, a fluorescence spectrum with a broad linewidth usually observed in glass materials is beneficial for the fabrication of a wavelength-tunable laser. The upconversion laser at room temperature was realized by using fibers based on zirconium fluoride glasses containing Ho$^{3+}$ and Er$^{3+}$ [95,96]. Silicate glasses containing nanosized fluoride crystals are also promising as short-wavelength laser materials from a point of view of high conversion efficiency, excellent chemical durability, high mechanical strength, and possibility to form fibers.

Other applications include sensing of the infrared light, which can be converted into the visible light by the upconversion process, three-dimensional display by using upconversion fluorescence inside transparent glass, and so forth. For the infrared sensing, the fact that Er$^{3+}$ converts 800 nm light to 550 nm green light can be utilized. The conversion efficiency is important because it determines the detection limit of the infrared light. Wang and Ohwaki fabricated an aluminosilicate glass containing (Pb, Cd)F$_2$:Er$^{3+}$, Yb$^{3+}$ nanocrystals, and succeeded in increasing the upconversion efficiency [97]. Yb$^{3+}$ was added as a sensitizer which improved the upconversion efficiency of Er$^{3+}$ via energy transfer. The transparent nanostructured glass-ceramic material was excited by using the 972 nm line from a GaAs-based laser diode. The emission intensity is 100 times stronger for the material, i.e., the transparent glass embedded with nanocrystals, than for aluminosilicate precursor glass and 2 times stronger than for fluoride glass.

### 4.3. Optical Memory Glass Based on Spectral Hole Burning

In conventional optical memories such as CD and DVD, a two-dimensional space is used for the storage of information. In contrast, optical memory glass based on spectral hole burning effect makes use of the wavelength of light as an additional domain of information storage, leading to a drastic increase in the density of optical storage. In inorganic crystalline and amorphous materials doped with rare-earth and transition metal ions as well as organic polymers containing dye molecules, a site distribution of rare-earth ions, transition metal ions, and dye molecules causes the broadening of their absorption spectra. When such a material is irradiated with a narrow band laser having a wavelength in the spectrum range of absorption, only a part of the photoactive species such as rare-earth ions and dye molecules that show a resonance with the incoming light is excited. Due to subsequent capture of an excited ion by a certain trap site and/or change in the site configuration around the excited species, the number of the species that show resonance with the incident laser light is decreased. This results in a decrease of absorption intensity at this wavelength, and a dip or hole comes to be observed in the absorption spectrum. This phenomenon is referred to as spectral hole burning. It is called photochemical hole burning especially when the chemical reactions are involved in the process of hole production. Detailed description on the spectral hole burning can be found in literature [98,99]. Figure 14 schematically illustrates the optical memory based on the spectral hole burning. A great deal of information can be written and read in one spot, and the density of storage is enhanced as the number of wavelength domain that is used for the memory is increased. For glass materials, room-temperature spectral hole burning was attained for oxide glasses doped with Sm$^{2+}$ [100,101] and Eu$^{3+}$ [102]. One characteristic of glass materials is a broad absorption spectrum corresponding to a broad site distribution compared with their crystalline counterparts, allowing an increase in the number of wavelength domain that can be used.

### 5. SUMMARY

In this chapter, we have looked through the amorphous materials that can contribute to the advance in
optoelectronics and photonics. These materials include nonlinear optical glass, magneto-optical glass, birefringent glass, upconversion glass, and optical memory glass based on spectral hole burning. Glass has a long history since the dawn of civilization and has been one of the key materials that support our daily life and society, because of their easy fabrication and shaping, high chemical durability, abundance of ingredient, as well as authentic beauty of their color. The amorphous materials overviewed in this chapter and will be supporting our society as important components of photonics. Lastly, it is worth noting that recent progress in nanofabrication technique will facilitate the development of glass materials for photonics.

REFERENCES


1. INTRODUCTION

Encapsulation of light in small volumes, in cavity structures, has been an essential requirement for optical devices. Optical fibers, fiber lasers, fiber amplifiers, and semiconductor lasers are successful examples of the light encapsulation in one- and two-dimensional cavities. Optical fibers, consisting of relatively high index core surrounded by low-index cladding glass, show long-length optical resonators. Rare-earth ion (Nd$^{3+}$, Er$^{3+}$)-doped silica glass fibers have been developed in the 1970s-1980s and now they are well known as conventional fiber lasers and amplifiers [1,2]. Similarly, Raman fiber lasers have also been investigated since the 1970s, but Raman gain is not so large that high pumping power of watt order was needed in the early trial [3-5]. Planar waveguide lasers, which have also been investigated since 1970s-1990s [6,7], were not used widely compared with the fiber lasers and the amplifiers.

Of all optical cavities, micrometer-sized spherical resonators show superior ability in storing optical energy within small volumes: in fact, it is the light encapsulation in three-dimensional cavity [8]. In the spheres, light forms whispering-gallery modes (WGMs) by total internal reflection along the curved boundary between the sphere and the surrounding medium. Spherical resonators have potential uses, such as a multiwavelength laser, a low threshold laser, and optical signal processors. A brief history of spherical cavity lasers has been summarized in our previous report [9].

This review describes briefly fiber lasers and fiber amplifiers as an introduction for one-dimensional resonators in the first section, and then the recent progress in spherical cavity glass lasers including the reports in my laboratories. In the section of spherical cavity laser, the report focuses on two optical coupling techniques for glass spheres developed in my laboratories: (1) light entrance portion formed on high-refractive index glass spheres (terrace microspheres) [10] and (2) a half-polished bending fiber coupler [11]. We tried to pump several tens of micrometer-sized high-index glass spheres ($n_D = 1.93$) for laser demonstrations.

2. FIBER LASERS AND FIBER AMPLIFIERS

Historically speaking, the light transmission along transparent cylindrical materials by internal reflection was surprisingly old. The idea of light transmission was first appeared in the image transmission along an aligned bundle of glass fibers of few micrometers in diameter, which is known as the fiberscope in medical inspection instrument, in the 1950s [12]. The first big stimulus to optical fiber researchers appeared in 1966, with estimation and proposal of optical fibers as the transmission medium, in the paper by C. Kao and G. Hockham [13]. The next spectacular advance was reported in 1970 by Corning Glass Works, which was the fabrication of optical fiber with a transmission loss of 20 dB/km [14]. Researchers devoted much effort to develop and improve the following fabrication techniques for optical fibers with a low-optical loss: modified chemical vapor deposition (MCVD) (Bell Labs, USA) [15], outside vapor deposition on a mandrel (Corning, USA), vapor-phase axial deposition (VAD)
Fiber laser was first reported in the early 1960s, with an Nd³⁺-doped barium crown glass rod at 1.06 μm wavelength \[18\], and amplification in a fiber laser of Nd³⁺-doped glass fiber \[19\]. At the early stage, there is little attention to such fiber lasers. Nowadays however, with greatly improved fiber fabrication including doping techniques, rare-earth ion-doped fibers lasers were enormously investigated and now cover the wide range of wavelength from the visible to the near-infrared region. Basic configuration of Fabry–Perot-type fiber lasers is shown in Figure 1. A fiber laser consists of a rare-earth ion-doped fiber with mirrors at both ends. Pumping laser light is introduced into the small core of a fiber amplifier through an optical coupler and the output laser light is radiated from the fiber end through an isolator. Since pumping light is confined into the core of several micrometers in diameter, its intensity is very high and thus the internal pumping efficiency is remarkable. In rare-earth ion-doped fibers, pumping threshold can be as low as a few hundred microwatts. Moreover, in 1990s cw (continuous wave) rare-earth ion-doped fibers lasers, pumping threshold can be as low as a few hundred microwatts. Moreover, in 1990s cw output power from several watts \[20\] to 100 kW \[21\] and femtosecond (fs) optical pulse generation \[22\] were achieved in fiber lasers. Raman fiber lasers operate on the basis of stimulated Raman scattering process. They have also been investigated since the 1970s, but Raman gain is not so large that high pumping power of watt order was needed in the early trial \[4,5\]. On the other hand, fiber lasers are flexible; thus, they can be coiled up and applied to ring fiber lasers such as Raman oscillator \[23\]. Configuration of the ring fiber laser is shown in Figure 2. Raman ring lasers have advantages such as long interaction lengths in low-loss optical fibers in relatively small coil diameter and broad Raman bandwidth in glasses. Pumping (input) and output were carried out with optical couplers, and the circulation of light is limited to one way using isolators. As shown in Figures 1 and 2, most of these fiber lasers have been realized in fiber-optic configurations using fiber couplers. Integrated optical circuit was proposed and tried to make silica glass waveguide system \[24\]: silica-based planar lightwave circuits (PLCs). Silica glass waveguides can be formed on substrates by the similar deposition techniques for optical fiber fabrication such as vapor-phase axial deposition (VAD). At early stage, PLC technology was used as passive optical devices such as filters, cross-connected switches, arrayed-waveguide grating multiplexer, and so on \[25\]. Planar waveguide lasers as active devices, which have also been investigated \[6,7\], and ring resonators of few millimeters in diameter with two stacked layers of input and output waveguides were demonstrated \[26\].

Optical fibers have been used remarkably efficiently to carry optical information from one point to another. Any optical signals, however, have to decrease after long length transmission in optical fibers. In early days, information system was required to involve high-cost electrical amplifiers every several ten kilometers (fault location), especially expensive in underwater cables. Er³⁺-doped fiber amplifiers changed the situation \[2\]. The optical loss minimum of silica-glass-based optical fibers is found to be at the wavelength of 1.55 μm \[17\], and fortunately erbium-doped optical fibers (EDFA) act as an optical amplifier in the wavelength region. In rare-earth-doped fiber amplifiers, typical fibers make use of Er³⁺, Pr³⁺, Tm³⁺, Nd³⁺, Yb³⁺, and Ho³⁺ for covering in the near infrared wavelength region. Host glasses for fiber amplifiers are different: silica glasses for Er³⁺ (1.5–1.6 μm wavelength band), and fluoride glasses for Pr³⁺ (1.3 μm wavelength band) \[27\] and Tm³⁺ (1.47 μm band) \[28\]. Yb³⁺ and Ho³⁺ were used as co-dopants with Er³⁺ and Tm³⁺. Raman fiber lasers have also been investigated from early years since the 1970s \[5\], because of the potential usefulness of glass fibers as broadband amplifiers and wavelength tunable oscillators; moreover, Raman gain is a very fast process compared with EDFA. The Raman scattering is the result of nonlinear effect; thus, its gain is so small that high pumping power of watt order was needed. Therefore without high-power lasers the Raman amplifier could not operate sufficiently in the 1970s. After 1995, however, various types of high-power lasers containing laser diodes and fiber lasers were developed; fiber Raman amplifiers were highlighted again \[29–31\].
3. SPHERICAL CAVITY GLASS LASERS

The previous section describes a brief history of the fiber resonators used for fiber lasers and amplifiers. Sizes of the resonators became smaller, from fiber laser, ring fiber lasers of meter length to planar waveguide, and ring resonators of few millimeters. Capability of light encapsulation increases with downsizing and from one—two-dimensional to three-dimensional structure resonators. Therefore, the spherical cavity of several tens of micrometers is the ultimate style of the optical resonator. This section describes the investigation of the micrometer-size spherical lasers.

A lasing action from a glass sphere is illustrated in Figure 3. In a sphere, activated light by a laser pumping forms whispering-gallery modes (WGMs) by total internal reflection along the curved boundary between the sphere (refractive index \( n_1 \)) and the surrounding medium (index \( n_2 \)). Quality factor \( Q \) of WGM depends on the diameter and relative refractive index \( (n_r = n_1/n_2) \) of the sphere. High \( n_r \) is inevitable to accomplish the spherical optical cavity structure in micrometer-sized spheres. WGM enables us to perform a strong interaction between laser light and materials and results in superior effects in a small volume. Using the WGM resonance, lasing can be achieved at much lower excitation intensity than within the corresponding bulk materials [8], and recently, spherical cavity-based Raman lasers were performed with ultrahigh-Q silica microspheres using fiber-taper couplers [32].

The first trial of laser action in spherical resonator was reported in Sm\(^{2+}\)-doped CaF\(_2\) spheres of millimeter size in 1961 [33]. Early works have already shown the ability of spherical resonators through Raman excitation in microdroplets of 60 \( \mu \)m in diameter [34]. To excite high-Q microspheres, pumping laser light should be coupled into spheres effectively. Various techniques have been tried using prism couplers [35], half-polished fiber couplers [36], fiber tapers [37], and so on. In Figure 4, schematic illustrations of the optical coupling techniques are shown: (a) prism coupling, (b) half-polished fiber coupler, and (c) fiber-taper coupler. In common techniques such as the prism coupling, evanescent field of a pumping laser beam overlaps with the WGMs of the spheres, resulting in sufficient optical coupling. High-purity silica glass spheres of high quality \( (Q = 10^{10}) \) [38] are expected to achieve a strong interaction of light and materials. The taper-fiber
coupling technique enabled us to perform a ultra-low threshold spherical Raman lasers [32]. In Figure 5, fabrication process of a taper fiber is illustrated. Taper fibers are made from a commercially available silica-glass-based single-mode optical fiber by heating and subsequently slowly pulling until suitable waist diameter is formed. A transmission mode of optical fibers is converted into an air-guided taper mode by passing through the waist region. Phase matching between taper-fiber modes and WGMs of the spheres was achieved by changing the taper waist diameter. The taper fiber is highly efficient for optical coupling in laboratory-scale experiments, but it requires delicately drawn optical fibers into few-micrometer-waisted taper fiber, and a crucial point is that tapered region is very fragile and easy to deteriorate.

3.1. Terrace-Microspheres for Laser Emission

Commercially available high-refractive-index ($n_D = 1.93$) glass microspheres of 30–40 μm in diameter were used for the experiments. Typical composition of the glass sphere is 38.0BaO–11.5SiO$_2$–38.5TiO$_2$–6.7CaO–5.3ZnO containing a few ppm of Nd$^{3+}$ [39]. The glass microspheres were made by the flame spray technique, where small pieces of glass cullet were melted in flame and glass spheres of smooth surface were formed by their surface tension.

For the fabrication of terrace microspheres, organic–inorganic hybrid sol was prepared by sol–gel technique using 3-methacryloxypropyltrimethoxysilane (MOPS) and tetramethoxysilane (TMOS) as starting materials. Reagents of MOPS and TMOS were hydrolyzed with hydrochloric acid as a catalyst and stirred for 3 days at room temperature. Suitable molar ratio of MOPS/TMOS was chosen to form the terrace structure on glass spheres. A microcapillary supplier controlled with a micromanipulator was used for the terrace formation. Using glass microcapillaries of a tip of about a 1 μm in inner diameter, picoliter volume of sol (25MOPS–75TMOS composition) was supplied into extremely narrow space between a sphere and Teflon sheet. After gelling, terrace microspheres were dried at room temperature and subsequently heated at 100 °C for 1 hour.

Typical SEM images of glass spheres and a terrace microspheres are shown in Figures 6 and 7. Surface of the
glass spheres is smooth and as shown in Figure 7, they have good-looking terrace shapes. Refractive index of the terrace portion was $n_D = 1.45$ after curing. Thus, the relative refractive index between high-index glass spheres and terrace portion $n_r = n_{glass}/n_{terrace}$ was about 1.3.

Emission spectra of the terrace microspheres were measured by a micro-Raman scattering spectrometer. Pumping wavelength was 514.5 nm (CW Ar$^+$ laser), with spectral resolution of 1 cm$^{-1}$. The incident laser beam was collimated to about 1 μm-diameter spot. Emission spectra of the terrace microspheres were also measured with a tunable CW Ti:sapphire laser ($\lambda = 700$ nm-850 nm) for choosing the suitable pumping wavelengths to WGMs of the microspheres. The incident laser beam after passing band-pass filter ($\lambda = 750$–820 nm) was collimated to about 2 μm-diameter spot by an objective lens (×100, NA). Terrace microspheres were fixed on the edge of a slide glass substrate by an electrostatic force. Emission from a terrace microsphere ($\lambda > 835$ nm wavelength) was sent to a monochromator through an optical fiber and analyzed by a CCD detector.

Figure 8 shows the resonant Raman spectra from terrace microsphere (of the BaO–SiO$_2$–TiO$_2$ glass) with various irradiation points. Corresponding laser pumping spots (from A to E) are illustrated in the inset figure. For color version of this figure, the reader is referred to the online version of this book.

![Resonant Raman spectra from terrace microsphere with various irradiation points. Corresponding laser pumping spots (from A to B) are shown in the inset figure. For color version of this figure, the reader is referred to the online version of this book.](image1)

![Emission spectrum of a terrace microsphere pumped by Ti:sapphire laser. For color version of this figure, the reader is referred to the online version of this book.](image2)
line), the threshold of stimulated Raman scattering and the change point in the Nd$^{3+}$ fluorescence intensity was correlated with each other. The two physical meanings should be noted from Figures 9 and 10. First, stimulated Raman scattering (SRS) was observed and the intensities of the emission peaks showed the maximum at the overlap region of spontaneous Raman scattering band with fluorescence band due to Nd$^{3+}$. It seemed that the SRS enhancement effect caused by the fluorescence was induced in the solid-state glass spheres. The fluorescence intensity showed the saturation of intensity against pumping power and subsequently SRS emission started in Figure 10; The energy transfer from fluorescence to Raman scattering support the enhancement effect in the glass sphere [40]. Second, the threshold of the Nd$^{3+}$-stimulated emission itself was of remarkably low value, less than 1 mW. These results reflect the strong interaction between laser light and glass matrix in spherical-cavity glass microspheres.

### 3.2. Pumping Demonstration Using a Half-Polished Fiber Coupler

The setup for the pumping experiments of the high-index glass microspheres is shown in Figure 11. A glass microsphere was placed on a coated half-polished fiber coupler. A tunable CW Ti:sapphire laser was used as a pumping light source. Emission from a glass microsphere was collimated by a lens with the similar technique described above in the terrace-microsphere experiment. The fiber coupler was made from a single-mode optical fiber with a core diameter of 3.7 μm (refractive index of the core glass is $n_D = 1.470$, cut-off wavelength $\lambda_c = 820$ nm) and outer diameter of 125 μm (refractive index of the cladding glass, $n_D = 1.461$). The bending fiber with a certain curvature was buried into plastic resin and was polished from surface to center of the fiber core. To improve optical coupling efficiency, high-refractive-index film ($n_D = 1.73$, 500 nm in thickness) made from organic—inorganic hybrid material was coated on the polished core.

Typical emission spectra from high-index glass microspheres are shown in Figure 12. In the upper and lower spectra, the sphere diameters were 36 μm and 33 μm, respectively. The similar spectra were obtained with those of Figure 9. Using the uncoated fiber coupler, only the resonant emissions in 880—940 nm wavelength regions were observed (not shown in the figure). Emission was originated from Raman scattering ($\lambda = 840$—880 nm) and Nd$^{3+}$ fluorescence ($\lambda = 880$—940 nm), and the calculated mode spacing agreed well with the measured values, for example, 4.35 nm vs. 4.39
nm in the upper spectrum. Emission intensities of 860.1 nm (due to Raman scattering) and 902.1 nm wavelength (due to fluorescence) are plotted as a function of pumping power in Figure 13. With increasing pumping intensity, the resonant emission due to fluorescence was observed at early stage (below 9.5 mW), and then that of Raman scattering was evolved at relatively high power pumping (above 12.5 mW). The thresholds of Nd3+ emission and Raman were 12 and 13 mW, respectively.

Glass superspheres of 0.5 mol% Nd3+-doped 15.5Na2O–12.8CaO–71.7SiO2 were also prepared by the surface tension mold (StM) technique as standard samples, which was previously reported by our groups [41]. Pumping experiments of the Nd3+-doped glass superspheres were carried out using a half-polished fiber coupler. Emission spectra in the 1060 nm wavelength region are shown in Figure 14. Laser emission due to well-known Nd3+ active ion was observed at three wavelengths originated from WGMs. The threshold intensities were estimated as 6.5 mW (1060.0 nm wavelength), 9 mW (1062.4 nm), and 12 mW (1063.3 nm). Additional pumping experiments, direct laser pumping in free space of Nd3+-doped glass superspheres was carried out, resulting in threshold intensities of 40–100 mW [41].

Many authors have investigated spherical resonators with WGMs. Relatively early investigations from theoretical and experimental views were published in the literatures [8,42]. Furthermore, transmission characteristics of optical fibers including taper fibers have also investigated theoretically [43]. Silica glass microspheres were successfully pumped by silica-glass fiber tapers and showed high-Q factors (quality factors) >109, where the matching between WGMs of microspheres and transmission modes of fiber-tapers was inevitable for laser
performance [44]. Nowadays, [fiber tapers]/[silica-glass spheres] system was well established experimentally in a laboratory scale and theoretical background was understood satisfactory. As pointed out in the literature [45], however, from the theoretical point of views, silica glass microspheres and optical coupling devices such as a fiber taper should have nearly equal refractive index for satisfying the mode matching. Moreover, fiber tapers, unclad and unsupported waveguides, are very fragile [44]. Therefore, if we have plans to open up a way to use the wide possibilities of glass microspheres of various compositions, development of the pumping techniques that is conventional and applicable for commercial uses is an urgent issue. Similarly, tailoring the glass composition is requested for obtaining desirable physical and chemical properties from their wide-spreading properties. Our works on the high-index glass spheres aim at giving one of the answers to the issues: two pumping techniques are applicable to the practical uses and the multicomponent high-index glass is chosen to explore the possibilities. In terrace microspheres, for obtaining high resonant emission, refractive index difference between the sphere and the terrace portion \( n_{\text{sphere}}/n_{\text{terrace}} = n_r > 1.3 \) is inevitable [45]. The pumping experiment of high-index spheres using a half-polisher fiber coupler showed the improvement of optical coupling by high-refractive-index film \( (n_D = 1.73) \) coating on the polished core [11]. These results were not explained by the above-mentioned theoretical consideration of the [fiber tapers]/[silica-glass spheres] system. Theoretical support for our new experimental techniques is the next important issue to make advance in practical applications.

To overview the efficiency of various resonators discussed in these sections, thresholds in various resonators such as fiber lasers, waveguide lasers, and micrometer-size spherical lasers are summarized in Figure 15. Designation (R) means the thresholds of Raman laser action, and \((\text{Nd}^{3+})\) means the thresholds of \(\text{Nd}^{3+}\)-doped waveguide and spherical lasers. The data in direct pumping \((\text{Nd}^{3+})\), fiber couplers \((\text{R and Nd}^{3+})\), and terrace \((\text{R})\) were obtained in my laboratory. An arrow in the figure shows the historical progress of small-size resonators along with optical coupling techniques, resulting in lowering the thresholds of laser emission.

Historically speaking, there were a number of earlier contributions of \(\text{Nd}^{3+}\)-doped fiber lasers during the 1960s—1970s [18,46], and the 1980s is the era of the rapid rise of \(\text{Er}^{3+}\)-doped fiber lasers and fiber amplifiers [47]. Raman lasers were also investigated in the 1970s, but the gain of Raman is smaller than that of rare-earth ion-doped fibers [48]; 1—10 W power was needed for laser action at an early stage. Waveguide-type lasers made from \(\text{Nd}^{3+}\) and \(\text{Er}^{3+}\)-doped glasses have been investigated since the early 1970s; several to several tens of mW power was

![FIGURE 14](image1.png) Laser emission spectra of 0.5 mol% \(\text{Nd}^{3+}\)-doped soda-lime silica glass spheres at around 1060 nm wavelength.

![FIGURE 15](image2.png) Thresholds of laser emission in various small-sized resonators: fiber lasers (R); [4,5] waveguide lasers \((\text{Nd}^{3+})\); [7] glass-sphere lasers pumped by direct pumping [41], half-polished fiber coupler \((\text{R and Nd}^{3+})\) [11], terrace \((\text{R})\) [9,45] and taper fibers \((\text{R})\) [32]. \((\text{R})\) and \((\text{Nd}^{3+})\) represent stimulated Raman emission and laser emission due to \(\text{Nd}^{3+}\). For color version of this figure, the reader is referred to the online version of this book.
necessary for laser demonstration [7] and is not still so widely used in the industrial-scale applications. Investigation of spherical lasers is in the midway from the stage of ‘conventional fiber and waveguide lasers’ to the ultimate stage of ‘[High-Q silica glass spheres]/[a taper fiber]’. Theoretically and experimentally, high-Q silica glass spheres pumped by taper fiber showed microwatt (μW)-level threshold for Raman laser action [32]. Such a microcavity Raman laser is highly attractive for extending the available wavelength range of conventional laser sources. Therefore, the fragile situation of taper fibers should be improved or replaced by other optical coupling techniques to achieve the commercial-scale applicability.

4. CONCLUDING REMARKS

In this review, I have tried to cover briefly the development of fiber lasers, amplifier, and spherical lasers as small-sized resonators. As the spherical cavity glass lasers are one of the promising optical devices for the next optical system, I described them in detail with including the data in my laboratory. Although the fiber lasers and amplifiers have been developed rapidly, the spherical cavity lasers are still under investigation to overcome the optical coupling issues to satisfy both the high efficiency and the sufficient mechanical strength for the practical applications. We described two coupling techniques developed in our laboratory for pumping multicomponent glass spheres. The techniques show the possibility of milliwatt order thresholds for laser demonstration, along with multiwavelength emissions.

REFERENCES


Multi-layered Ceramic Capacitors

Takaaki Tsurumi and Takuya Hoshina
Graduate School of Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

Chapter Outline

1. High-Capacitance MLCCs With Nickel Internal Electrodes 415
2. Nonlinear Dielectricity of MLCCs 417
3. Capacitance Aging in MLCCs 418
4. Size Effect of BaTiO₃ Ceramics 421
5. Reliability of MLCCs — Lifetime in HALT 423
6. Computer Simulation and Further Prospective of MLCC Technology 425
References 426

1. HIGH-CAPACITANCE MLCCs WITH NICKEL INTERNAL ELECTRODES

Chip-type capacitors are the main products of electronic ceramic components due to their superior frequency performance and volume efficiency, which meets the requirement from the advanced electronic devices. A schematic view of multilayer ceramic capacitor (MLCC) is shown in Figure 1. The capacitor is composed of many thin plate capacitors in parallel connection. The capacitance of MLCC with electrode area $S$, dielectric layer thickness $t$, and the number of dielectric layers $n$ is given by

$$ C = \frac{n \varepsilon_r S}{t} = \frac{\varepsilon_r S T}{t^2} $$

where $\varepsilon_r$ is dielectric constant and $T$ is the chip thickness (Figure 1). The above relationship indicates that the reduction of dielectric layer thickness $t$ is very effective to increase the volume efficiency. The increase of capacitance volume—density of MLCCs has been therefore realized by reducing dielectric layer thickness; the latest commercialized products have the layer thickness less than 1 $\mu$m as shown in Figure 2 [1].

Since drastic increase of price during 1995—1997, MLCC manufacturers tried to replace silver—palladium inner electrodes with nickel electrodes. As nickel is easily oxidized in air at elevated temperatures, MLCCs with nickel inner electrodes should be fired in an atmosphere with a low oxygen partial pressure. However, the problem was that properties of BaTiO₃-based dielectrics were easily reduced in such sintering conditions. Figure 3 shows a famous result of Sakabe at Murata Manufact., indicating the possibility to produce BaTiO₃ ceramics with anti-reducing properties [2,3]. Resistivity $\rho$ of BaTiO₃ ceramics sintered in reducing atmosphere (oxygen partial pressure of $2 \times 10^{-12}$ MPa) was not degraded by doping CaO when A/B ratio $m$ was over unity. Essential points in the composition designing are 1) control of the molecular ratio ($Ba + Ca$)(Ti Zr)O₂ larger than unity, 2) incorporation of Ca ions into Ti lattice site (B site) of BaTiO₃, and 3) doping acceptor as Mn ions. Mg ions have the same effect as Ca ions. The formation of oxygen vacancies is described as

$$ O_O \rightarrow V_{O}^{\cdot\cdot} + \frac{1}{2} O_2 + 2e', \quad K = [V_{O}^{\cdot\cdot}]n^2P(O_2)$$

FIGURE 1 Schematic view of MLCC. For color version of this figure, the reader is referred to the online version of this book.
where $K$ is the equilibrium constant, and the electron concentration is

$$n = \frac{K^\frac{1}{2}}{|V_{O}^{*}|^2P(O_2)^\frac{3}{2}}$$

The substitution of Mg ions to Ti ions can be described as

$$\text{MgO} \rightarrow \text{O}_2 + \text{Mg}^{2+} + V_{O}^{*}$$

From the above relationships, Mg ion substitution to Ti sites increases the oxygen vacancy concentration $[V_{O}^{*}]$, resulting in the decrease of electron concentration $n$. Because of the decrease of charge carrier, the resistivity of BaTiO$_3$ ceramics does not degrade as shown in Figure 3, enabling us to use inexpensive nickel internal electrodes in MLCCs.

In modern electronic circuits, the multilayer ceramic capacitor (MLCC) is an indispensable passive component for manufacturing miniaturized electronic circuits. The production of MLCC increases every year, especially high-capacitance MLCCs are gradually replacing the market of tantalum capacitors. Most of high-capacitance MLCCs use Ni internal electrodes and satisfy the EIA X7R specifications, where dielectric layers have the ‘core—shell’ structure (Figure 4), in which one grain of BT-based ceramics consists of inner core of almost pure BT and outer shell of BT with
additives [4–7]. However, several problems such as nonlinear dielectricity, capacitance aging, size effect, and reliability have arisen to prevent further increase of the capacitance density.

2. NONLINEAR DIELECTRICITY OF MLCCs

Nonlinear dielectricity means the electric field dependence of permittivity. It becomes important in high-capacitance MLCCs because the electric field applied on one dielectric layer increases with decreasing layer thickness. Nonlinear dielectricity causes capacitance degradation under high DC-bias voltage and temperature stability degradation under high AC voltage. Nonlinear dielectric responses of high-capacitance MLCCs are attributable to the ferroelectricity of BaTiO$_3$ [8].

AC-field dependence of permittivity of (Ba$_{0.79}$Zr$_{0.21}$)TiO$_3$ (BZT), dielectrics in MLCCs with EIA Y5V specification, is shown in Figure 5. The permittivity exhibits a maximum at a certain electric field. This behavior can be explained by the model of relaxors; the increase of permittivity is due to the enhancement of dipole polarization of thermally fluctuating dipoles in polar nano-regions (PNRs) by the application of high fields. This enhancement saturates at a certain field, giving rise to the maximum of permittivity [9, 10]. The temperature variation of permittivity of BZT measured with different AC fields is shown in Figure 6. As the AC field increases, the temperature at permittivity peak ($T_m$) decreases and the peak becomes asymmetric. In relaxors, it is known that no phase transition occurs at $T_m$. The decrease of permittivity at low temperatures is caused by the freezing of thermally fluctuating dipoles. Those freezing dipoles were forced to move by high AC fields, contributing dipole polarization; therefore, permittivity measured by high AC field does not decrease even at low temperatures. Thus, the shift of $T_m$ and the peak broadening to low temperature side with increasing AC field can be explained.

To elucidate the dielectric properties of the core and the shell parts in X7R-MLCCs, we have prepared three types of MLCCs with only core grains, only shell grains, and core–shell grains. The composition of core grains was BaTiO$_3$ while that of shell grains was (Ba$_{0.95}$Ho$_{0.05}$)(Ti$_{0.975}$Mg$_{0.015}$Mn$_{0.01}$)O$_3$. Figure 7 shows dielectric properties of the three kinds of MLCCs measured with various AC fields [7]. Comparison between Figure 7(a) and (c) revealed that the dielectric properties of X7R-MLCCs were dominated by the core part. The formation of core–shell structure depressed the AC-field dependence of permittivity at low temperatures. Relatively flat temperature dependence was observed for MLCCs with only core grains (Figure 7(a)), although the core grains consisted of pure BaTiO$_3$ which shows a sharp permittivity peak at 130 °C. This is because the ferroelectricity of BaTiO$_3$ was depressed in core grains of 200–300 nm by the size effect. On the other hand, the dielectric property of shell parts (Figure 7(b)) is similar to that of BZT (Figure 6), indicating that the dielectric response of the shell grains is explained by the mechanism of relaxors. Consequently, the dielectric layers of X7R-MLCCs consist of BaTiO$_3$ core parts.
showing depressed ferroelectricity by the size effect and the shell parts of BaTiO$_3$-based solid solutions showing relaxor-like behaviors. Grains in the dielectric layers vary in the dopant distribution at the shell parts, giving rise to the distribution of $T_m$ in Figure 7(b). The role of shell parts is to increase the permittivity at low temperatures; the shell parts contribute the temperature stability of MLCCs. However, in the case of ultra-thin dielectric layers, the electric field applied to the dielectric layers increases and the temperature stability of permittivity will be achieved without the shell parts. It should be also noted that the permittivity of the shell parts is lower than that of the core parts (Figures 7(a) and (b)), which means that the shell parts degrade the capacitance of MLCCs.

3. CAPACITANCE AGING IN MLCCs

Capacitance aging is a phenomenon wherein the capacitance of MLCCs somehow decreases with time under DC-bias voltages at relatively high temperatures [6,11]. The capacitance-aging problem of BaTiO$_3$ has been studied for a long time [12-14] but these studies mostly dealt with aging without DC fields. MLCCs with high capacitances are mainly used as bypass capacitors in power lines of electric circuits to stabilize the DC voltage of the lines, which means that the capacitors are used under DC fields for long durations. One of the most important applications of high-capacitance MLCCs is bypass capacitors connecting to power lines of CPU. These MLCCs are used under DC-bias voltages at relatively high temperatures. The decrease of capacitance sometimes makes it difficult to supply enough current to CPU, which is not predicted by electronic circuit engineers.

Figure 8 shows a typical capacitance-aging curve of X7R-MLCCs measured under a DC electric field. $\delta C$ (%) is defined as $\delta C = (C - C_0)/C_0 \times 100$, where $C_0$ is the capacitance measured without a DC field at the temperature shown in the figure. It is possible to see that the reduction of the capacitance proceeds in two stages as pointed out by Nomura et al. [11]. The decrease of the capacitance in the first stage takes place immediately after the application of a DC field. The first stage is therefore attributable to the nonlinear dielectric permittivity of dielectric materials and it should not be included in the capacitance-aging phenomenon. The second stage starts after the first stage and continues up to $10^5$ s and more. The $\delta C$ in the second stage depends on temperature, indicating that a thermal activation process is involved in the aging phenomenon. Figure 9 shows the change in the aging behavior with MnO content ($x$) into the dielectric with the composition of $100\text{BaTiO}_3-0.5\text{Ho}_2\text{O}_3-0.5\text{MgO}-x\text{MnO}-1.5\text{BaSiO}_3$. The relaxation time of the aging obviously decreased with increasing MnO content and the $\delta C$ of MLCCs with different Mn contents tended to saturate to a certain value after a long aging time. The $\delta C$ after a long aging time is independent of the Mn content, which indicates that the internal field or space charges are less related with the aging phenomenon.

The aging behavior measured at different temperatures is shown in Figure 10. The relaxation time was reduced with temperature, indicating that a thermally activated process is involved in the aging phenomenon. It must be
noted that the capacitance aging was markedly depressed at 150 °C, which is above the Curie temperature of BaTiO₃, because ferroelectric domains deeply concern with the aging phenomenon. In Figure 11, the aging behavior of MLCCs fired in reducing or air atmosphere was compared. The aging in the second stage is markedly depressed in the air-fired specimen. It is also notable that the δC after the aging is almost consistent for the two specimens. In other words, the aging in the second stage could be moved to the first stage by firing in air atmosphere. The Mn ions in

FIGURE 7 Temperature dependence of permittivity of MLCCs with (a) core, (b) shell, and (c) core–shell grains.

FIGURE 8 Typical examples of capacitance aging in X7R-MLCCs measured at room temperature and at 110 °C. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 9 Effect of MnO addition on capacitance aging at 110 °C. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 10 Capacitance aging measured at different temperatures.

the core of the reduced specimen have a trivalent (Mn³⁺) or divalent (Mn²⁺) state, which clamp the domain-wall motion. It is well known that Mn is an effective additive for making the ‘hard’ PZT ceramics, in which the domain-wall motion is clamped by the defect dipoles associated with

FIGURE 11
Mn$^{3+}$ and/or Mn$^{2+}$ substituted with B-site ions in the perovskite structure [15,16]. In the MLCCs fired in air atmosphere, Mn$^{3+}$ and/or Mn$^{2+}$ are possibly changed to Mn$^{4+}$, which is not effective in clamping the domain-wall motion. This change results in a relatively free movement of domain walls by the first application of a DC field; therefore, the $\delta C$ in the second stage moved to the first stage in the air-fired specimen. The $\delta C$ after the aging was consistent for the two specimens because the firing conditions did not affect considerably the nonlinear permittivity and domain structure of MLCCs.

The capacitance aging in the second stage is caused by the 90° domain switching in BT in the core. The mechanism of the capacitance aging in X7R-MLCCs is shown in Figure 12. Most of the domain walls in the core are clamped by the defect dipoles formed by negatively charged Mn sites and positively charged oxygen vacancies. The movement of unclamped domain walls may be involved in the $\delta C$ in the first stage, although the $\delta C$ in the first stage is due mainly to the nonlinear dielectric permittivity of the relaxor-like shell. In the second stage, the continuous application of DC fields moves the electrons on Mn sites to eliminate the defect dipoles with time. This enhances the domain-wall motion to align the c-axis of BT to the electric field direction, resulting in the dielectric permittivity due to the permittivity difference between $\varepsilon_{33}$ and $\varepsilon_{11}$. The movement of electrons is a thermally activated hopping process. Because the increase in Mn sites in the core enhances the hopping of electrons, relaxation time is reduced with increasing MnO content ($x$) as shown in Figure 9. The difference between $\varepsilon_{33}$ and $\varepsilon_{11}$ is almost independent of Mn content; so, the $\delta C$ after the aging tends to saturate to a certain value as shown in Figure 9.

Figure 13 shows the change in the aging behavior with the DC field. It should be noted that the first stage was
enhanced but the second stage was depressed to disappear when a high DC field was applied. This is because a high DC field moves the domain walls quickly in the first stage in spite of the clamping of domain walls by defect dipoles. The same phenomenon is normally observed in ferroelectrics when an applied DC field exceeds the coercive field. The difference in $\delta C$ after the aging is purely due to the effect of nonlinear permittivity because the domain switching is completed after the second stage. Figure 14 shows capacitance ($C$) vs voltage ($V$) and electric displacement ($D$) vs electric field ($E$) curves measured for the specimen in Figure 13. The clear hysteresis observed in the $D-E$ curve indicates that domain switching definitely takes place in the dielectrics. By assuming that the difference in $\delta C$ after the aging in Figure 13 was caused only by the nonlinear permittivity, it was possible to separate the effect of the nonlinear permittivity and that of the domain switching from the experimental results. The two additional curves in Figure 14 (b) show both effects. The $D-E$ curve observed could be explained by the sum of the nonlinear permittivity and domain switching. The effects of the nonlinear permittivity and domain switching were also separated and are shown in Figure 13. It is important that the domain switching effect on the $\delta C$ is consistent for all curves in the figure and the $\delta C$ due to the nonlinear permittivity increases with DC field.

4. SIZE EFFECT OF BaTiO$_3$ CERAMICS

It is now recognized that the permittivity shows a maximum at about 1 $\mu$m, i.e., the permittivity of BT ceramics increases with decreasing grain size in the micron level while that decreases in the submicron level. Arlt et al. [17] proposed that the 90° domain walls were the origin of the enhancement of permittivity. To study the intrinsic size effect, it is important to prepare fine-grained BaTiO$_3$ ceramics without internal stresses or impurities because the crystal structure and the dielectric property of BT ceramics depend on internal stress and impurities. The two-step sintering method [18–20] is effective to prepare fine-grained BT ceramics without internal stress or impurity. Figure 15 shows the grain size dependence of permittivity for the BT ceramics prepared using conventional and
two-step sintering methods [21]. The permittivity of pure BaTiO₃ ceramics increases with decreasing grain size when the grain size is above 1.1 μm, whereas it decreases when the grain size is below 1 μm. The maximum permittivity of the BT ceramics is 8000 for a grain size of 1.1 μm.

Figure 16 shows scanning electron microscope (SEM) images of the BT sample surface after chemical etching with various grain sizes [21]. The observations reveal that most of the grains had simple lamellar structures of 90° domains. The 90° domain width decreases with decreasing grain size. To understand the domain contribution to the permittivity, the measurement of ultrawide range dielectric spectra of the BaTiO₃ ceramics is very effective [22, 23]. Figure 17 shows the dielectric spectra of the BaTiO₃ ceramics with various grain sizes. In this figure, the Debye-type dielectric relaxations observed below 10⁹ Hz are attributable to the dipole polarization due to the domain-wall vibration induced by the electric fields. On the other hand, the dielectric dispersions observed over 10¹⁰ Hz are attributable to ionic polarization. Figure 18 shows the grain size dependence of the permittivity due to dipole and ionic polarizations. The increase in permittivity with decreasing grain size to 1 μm is due to the 90° domain contribution. In contrast, for grain sizes below 1 μm, the dipole polarizability and the ionic polarizability decrease with decreasing grain size. It suggests that the domain contribution to the permittivity, the measurement of ultrawide range dielectric spectra of the BaTiO₃ ceramics is very effective [22, 23]. Figure 17 shows the dielectric spectra of the BaTiO₃ ceramics with various grain sizes. In this figure, the Debye-type dielectric relaxations observed below 10⁹ Hz are attributable to the dipole polarization due to the domain-wall vibration induced by the electric fields. On the other hand, the dielectric dispersions observed over 10¹⁰ Hz are attributable to ionic polarization. Figure 18 shows the grain size dependence of the permittivity due to dipole and ionic polarizations. For grain sizes over 1 μm, the dipole polarizability due to domain-wall vibration increases with decreasing grain size. It indicates that a high 90° domain-wall density enhanced the dipole polarizability due to the domain-wall vibration. Moreover, the ionic polarizability also increases with decreasing grain size. The 90° domain structure induces the lattice strain into the BaTiO₃ lattices around the domain wall. It is believed that the lattice strain region around the 90° domain wall has high ionic polarizability [24, 25]. Therefore, high domain-wall density enhances the ionic polarizability as well as the dipole polarizability. The increase in permittivity with decreasing grain size to 1 μm is due to the 90° domain contribution. In contrast, for grain sizes below 1 μm, the dipole polarizability and the ionic polarizability decrease with decreasing grain size. It suggests that the domain contribution
decreases when grains became smaller than 1 \( \mu m \). However, it is still difficult to explain the mechanism of permittivity decrease below grain sizes in the submicron range. For further studies, it is necessary to clarify theoretically the weakening mechanism of dipole and ionic polarizability for the fine-grained BaTiO3 ceramics with sizes below 1 \( \mu m \).

Grain size dependence of the permittivity of dielectric layers in X7R-MLCCs (Taiyo Yuden, Co., Japan) is shown in Figure 19 [26]. The permittivity simply decreases with decreasing grain size. To understand the size effect of BaTiO3-based dielectric layers in X7R-MLCCs, it should be recalled that dielectric layers of the MLCCs consist of the “core—shell” grains (Figure 4). The core—shell structure is a complex structure in one grain, where the shell part is heavy-doped BaTiO3 showing relaxor-like behavior, while the core part is almost pure BaTiO3 showing depressed ferroelectric behavior due to the size effect of BaTiO3. Figure 20 shows the permittivity of dielectric layers with different grain sizes obtained by the SPLINE analysis of experimental data taken for MLCCs with different grain sizes [25]. It should be noted that the temperature stability of the permittivity was markedly degraded when the grain size became smaller than 120 nm. In the MLCCs with the core—shell structure, the shell dominated the permittivity at low temperatures, while the core dominated the permittivity at high temperatures within the X7R specification. The degradation of temperature stability observed in small grains indicated that the portion of shell part increased in small grains. The shell parts are formed by controlling the diffusion of additive ions from the grain boundary, which means that the formation of thin shell parts becomes difficult as the grain size becomes small, giving rise to the increase of the shell part in small grains. In the size effect of dielectric layers in MLCCs, the core—shell structure plays an important role.

5. RELIABILITY OF MLCCS — LIFETIME IN HALT

The increase of capacitance density of MLCCs has been realized by the reduction of dielectric layer thickness but enough reliability (a long lifetime in HALT) has been another criterion to be achieved in the development of MLCCs. It is worth to describe the difference in breakdown, capacitance aging, and lifetime in the HALT. A typical relation between DC voltage and leakage current of X7R-MLCCs is shown in Figure 21(a) [27]. The application of a high voltage shown by X-point brings about insulation break of MLCCs, which is called breakdown and the voltage of X-point is called breakdown voltage. By applying a DC voltage far below the breakdown voltage, the capacitance of MLCCs decreases with time down to a few tens of percents, which is called capacitance aging caused by the domain-wall motion in the core parts as mentioned in the section of capacitance aging. Figure 21(b) shows the leakage current variation with time in the HALT of MLCCs where a DC voltage below the breakdown voltage is applied at relatively high temperatures of 120–200 °C. The leakage current gradually increases with time, and, at a certain time, the current increases steeply to the breakdown of MLCCs. This time is called a lifetime in the HALT. The breakdown in the HALT is caused by the accumulation of oxygen vacancies at the cathode.

The model of oxygen vacancy distribution before and after the HALT is shown in Figure 22 [28]. Plenty of oxygen vacancies are formed in the core parts of almost pure BaTiO3 after sintering in reducing atmosphere. However, the number of oxygen vacancies in the shell parts should be smaller than those of the core parts because some additives in the shell parts work as anti-reducing agents. Positively charged oxygen vacancies (\( V'_O \)) migrating to the cathode under a DC field are blocked at the shell parts and
grain boundaries. In this sense, grain boundaries and the shell parts contribute to the reliability of X7R-MLCCs. An interesting question is the role of rare earth elements such as Dy, Ho, and Y to improve the lifetime in the HALT. The relation between Ho content and the lifetime in HALT is shown in Figure 23. The lifetime rapidly increases from the Ho content of 0.2 atom% and tends to saturate with increasing Ho content. Analysis of local Ho distribution indicated that Ho segregated at the shell parts and grain boundaries; similar results were obtained for Dy and Y. Thus, rare earth elements such as Dy, Ho, and Y change the properties of shell parts and grain boundaries to depress the migration of oxygen vacancies. These rare earth elements are incorporated into the A-sites of perovskite structure as a donor when the content is low. The addition of donor additives decrease the number of oxygen vacancies and act as a block of oxygen vacancy migration. The similar

FIGURE 21 (a) A typical relation between DC voltage and leakage current of X7R-MLCCs and (b) the leakage current variation with time in the HALT of MLCCs [27]. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 22 Distribution of oxygen vacancies in the core–shell structure before and after HALT [28]. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 23 Relation between lifetime in HALT and Ho content in the dielectrics of X7R-MLCCs [27].
situation is valid for grain boundaries but the change of electric properties at grain boundary is more effective for the improvement of lifetime in HALT.

Figure 24 shows the relation between DC voltage and leakage current of X7R-MLCCs added with different amounts of Ho. Solid circles in the figure indicate that the leakage current was not stable but increased to breakdown during measurements. MLCCs added with 1.5 mol% of Ho stably passed a relatively large current under high field, which is possibly due to a tunneling current. The resistance of core parts, shell parts, grain boundaries, and electrode interfaces was determined by the equivalent circuit analysis of MLCCs as shown in Figure 25. The electrode interface has the highest resistance and the grain boundaries have the second highest resistance. As the resistance of electrode interfaces should be independent of Ho contents, the resistance change, giving rise to the change of leakage current shown in Figure 23, is due to the grain boundaries, which means that the addition of Ho decreases the resistance of grain boundaries to stably pass the current under DC voltage.

The resistance of grain boundaries is higher than that of core parts and shell parts. The DC voltage applied to MLCCs is divided to core parts and shell parts according to their resistance. If the resistance of grain boundaries is extremely high, DC voltage is mainly applied to grain boundaries, giving rise to the enhancement of oxygen vacancy migrations across grain boundaries followed by the breakdown of grain boundaries by heat generation. The enhancement of oxygen vacancy migration across the grain boundary leads to the breakdown of the whole device by the accumulation of oxygen vacancies at the electrode interfaces. On the contrary, if grain boundaries stably pass a relatively high current, the oxygen migration across grain boundaries is depressed because the DC voltage does not concentrate to grain boundaries, giving rise to the improvement of the lifetime in HALT.

The contribution of Dy, Ho, and Y to the improvement of the lifetime is interpreted as follows: 1) these rare earth elements are possibly incorporated into both A-sites at the shell parts and decrease the oxygen vacancy concentration to depress the migration of oxygen vacancies, 2) rare earth elements incorporated into grain boundaries decrease the resistance of grain boundaries to avoid the voltage concentration leading to the breakdown at grain boundaries, and 3) the decrease of DC voltage applied on the grain boundaries depresses the oxygen vacancy migration across the grain boundaries and their accumulation at electrode interfaces.

6. COMPUTER SIMULATION AND FURTHER PROSPECTIVE OF MLCC TECHNOLOGY

The reduction of dielectric layer thickness demands the simultaneous reduction of BaTiO₃ grain size to keep the several grain boundaries in one dielectric layer. However, the dielectric permittivity reduced with BaTiO₃ grain size because of the size effect. Consequently, there must be a limit of capacitance density in the current technologies of MLCC. Computer simulation software (MLCC-simulator)
was developed to optimize the MLCC structure and to predict the limit of capacitance density of MLCCs [25]. The highest capacitance density predicted by the simulation was 47 μF in 1608 mm chip size where the electrode thickness was 0.4 mm and the BaTiO₃ grain size was 93 nm. Figure 26 shows the dielectric properties of this MLCC. The temperature stability was markedly degraded when grains smaller than 120 nm were used. The third simulation seemed to be close to the technical limit of the current technology. The results of simulations gave the following important information: 1) there is no advantage to use BaTiO₃ grains smaller than 80 nm, 2) the reduction of Ni internal electrode thickness is effective to obtain high capacitance, 3) enough reliability should be obtained with the small number of grains in one dielectric layer, 4) the reduction of dielectric layer thickness is effective to obtain 22–47 μF in 1608 chip size, and 5) the limit of capacitance in 1608 chip size is about 47 μF as long as we use the current technology. Finally, we have to stress that the technical limit was predicted by assuming there is no progress of the current technology.

The fact that several grain boundaries are necessary to achieve a long lifetime in HALT indicates that the grain size decreases with decreasing dielectric layer thickness, giving rise to the degradation of permittivity of dielectric layers. The simulation indicated that the highest capacitance achieved by the current technology is 47 μF in 1608 chip size, which is close to the density of latest products. This limit is mainly determined by the number of grains needed in one dielectric layer and the thickness of Ni internal electrodes. We will discuss on the issues concerning dielectrics.

Microstructure control of dielectrics is the key issue to increase the capacitance density. The ideal structure is that the shell part is as thin as possible to obtain the high permittivity, and the distribution of additives at grain boundaries is precisely controlled to achieve the high reliability. To realize them, one of the most important issues is the selection of raw BaTiO₃ powders. Hoshina et al. [29] analyzed the internal structure of BaTiO₃ nanoparticles and revealed that the BT nanoparticles consisted of three regions: surface cubic layers, internal tetragonal layers, and gradient lattice strain layers (GLSLs) between them. The c/a ratio determined by the XRD analysis decreases with increasing thickness of cubic layers, which means that the c/a ratio is the index of defect concentration in BaTiO₃ powders. In the formation of core–shell structures, additives diffuse from the surface to the inner part of BaTiO₃ particles. The diffusion should be as slow as possible to form the core–shell structure in fine BaTiO₃ grains but the defects in crystals usually increase the diffusion constant. It is therefore difficult to form the core–shell structure using BT powders with low c/a ratio. The collapse of core–shell structure degrades the permittivity and its temperature stability.

The size effect is an intrinsic effect of BT and it may be difficult to completely overcome it as long as BT-based ceramics are used in dielectric layers. We cannot expect to double the capacitance if we decrease the dielectric layer thickness to half any more. The increase of capacitance density should be preceded by the microstructure control which implies the formation of core–shell structure with thin shells and proper distribution of additives at grain boundaries. The formation of ideal core–shell structure is achieved by the selection of BT powders with fewer defects as well as the optimization of temperature profiles in the sintering.

REFERENCES

[1] Photograph was provided by Taiyo Yuden Co. Ltd., Japan.
Lead-Free Piezoelectric Ceramics

Tadashi Takenaka
Faculty of Science and Technology, Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278-8510, Japan.

Chapter Outline
1. Introduction 429
2. Perovskite-structured Piezoelectric Ceramics 430
3. Summary 443
References 444

1. INTRODUCTION

Piezoelectric materials play an important role for electronic devices such as actuators, sensors, accelerators, ultrasonic motors, transducers, filters and resonators, and micro electromechanical systems (MEMS). Important ferroelectric oxides for piezoelectric ceramics are restricted to perovskite-type, tungsten bronze-type, and bismuth layer-structured compounds. The most widely used piezoelectric materials are perovskite-type PbTiO₃-PbZrO₃ (PZT)-based multi-component systems (PZT systems) [1–3] because of their excellent piezoelectric properties. However, it has been recently desired to use lead-free materials for environmental protection during the waste disposal of products. For example, legislation has been enforced in the EU as the draft Directives on Waste from Electrical and Electronic Equipment (WEEE) since January 1st, 2004, Restriction of Hazardous Substances (RoHS) since July 1st, 2006, and End-of-Life Vehicles (ELV) since July 1st, 2003. Therefore, lead-free piezoelectric materials have been attracting attention worldwide [4–6] as new materials in place of PZT-based piezoelectric ceramics.

Lead-free piezoelectric materials, such as piezoelectric single crystals, e.g., langasite [7], and ferroelectric ceramics with a perovskite structure [8–59], a tungsten bronze structure [60,61], and bismuth layer-structured ferroelectrics (BLSF) [62–68] have been extensively reported for mainly in the past 10 years. Recently, various perovskite-structured ferroelectrics such as BaTiO₃ [BT], (Bi₁/₂Na₁/₂)TiO₃ [BNT], (Bi₁/₂K₁/2)TiO₃ [BKT], KNbO₃ [KN], (K,Na)NbO₃ [KNN], and their solid solutions have been actively studied [16–59] as candidates for lead-free piezoelectric ceramics. These ceramics have been widely studied and are suitable for actuator and high-power applications because of their relatively large piezoelectric constants, d, among lead-free piezoelectrics. However, there are some problems such as low Curie temperatures, Tᵣ, or low depolarization temperatures, T₅, difficulties in poling treatments, and/or low relative densities. The ferroelectric strengths of perovskite-structured ABO₃-type oxides are representative by PbTiO₃ [PT] in 2–4 type, KNbO₃ [KN] in 1–5 type, and BiFeO₃ [BF] in 3–5 type, respectively. However, all of them, except PT, do not show their strong ferroelectrics as high as those of PZT. Therefore, no lead-free materials displays better piezoelectric properties than those of PZT-based systems. To replace PZT-based systems, it is necessary that the required piezoelectric properties for various applications are classified and developed for each application. For example, the perovskite-type ceramics seem to be suitable for actuator and high-power applications. On the other hand, BLSF ceramics seem to be candidate materials for ceramic resonator applications.

In this section, a brief survey of non-lead-based piezoelectric ceramics is given and dielectric, ferroelectric, and piezoelectric properties of typical lead-free perovskite ferroelectric ceramics such as BaTiO₃ [BT], Bi₁/₂Na₁/₂TiO₃ [BNT], Bi₁/₂K₁/2TiO₃ [BKT], and KNbO₃ [KN]-based systems for piezoelectric applications are described as superior candidates for environmentally friendly lead-free piezoelectric ceramics to reduce some damage to the earth.
2. PEROVSKITE-STRUCTURED PIEZOELECTRIC CERAMICS

2.1. BaTiO3 [BT]-based Ceramics

Barium titanate, BaTiO3 [BT] [8,9], is the first-discovered ferroelectric oxide with perovskite structure. This ceramic has a relatively high electromechanical coupling factor, $k_{33}$ (~0.50), and a piezoelectric strain constant, $d_{33}$ (~190 pC/N), and has been partially used for piezoelectric applications such as sonors. However, the working temperature range of the BT is narrow for actual piezoelectric applications because the BT has a low Curie temperature, $T_c$ (~120–135 °C) [8,9,54,55]. On the other hand, bismuth potassium titanate, (Bi$_{12}$K$_{1/2}$TiO$_3$ [BKT]), is a typical lead-free ferroelectric with a perovskite structure of tetragonal symmetry at room temperature and a relatively higher $T_c$ (~380 °C) [13]. Hiruma et al. [42] reported the electrical properties of BKT ceramics prepared by the hot-pressing (HP) method. To expand the working temperature range, that is, to elevate the $T_c$ of BT ceramic, dielectric and piezoelectric properties of $(1-x)$BaTiO$_3$-xBKT solid solution (BT-BKT10x) system were also investigated [39,94]. The $T_c$ increased linearly with increasing the BKT content (x) and the $T_c$ of BT-BKT20 (x = 0.2) reaches higher than 200 °C.

Recently, large piezoelectric strain constants, $d_{33}$ (=350 pC/N) and $d_{33}$ (=450 pC/N), of BT ceramics were reported, respectively, in the microwave sintered BT by using hydrothermally prepared fine particles [54] and in the BT prepared by the two-step sintering method [55]. These large $d_{33}$ values may be caused by the elevated very high relative free permittivity ($\varepsilon_f$/$\varepsilon_0$) higher than 5000.

2.2. KNbO$_3$ [KN]–NaNbO$_3$ [NN]–LiNbO$_3$ [LN] System

Potassium niobate, KNbO$_3$ [KN], ceramics have attracted much attention as a candidate material for lead-free piezoelectric applications, because the single-crystal KN has a large piezoelectricity and a high Curie point [10,30,32]. KN has an orthorhombic symmetry at room temperature, and has phase transitions at ~−10, 225 and 425 °C corresponding to rhombohedral → orthorhombic → tetragonal → cubic, respectively. The electromechanical coupling factor of the thickness-extensional mode, $k_{0}$, in a KN crystal reaches as high as 0.69 for the 49.5° rotated X-cut about the Y-axis, which is the highest among current lead-free piezoelectrics [49]. KN single crystals are known to have high piezoelectric activities. However, it is difficult to obtain a dense ceramic body of KN by the ordinary firing process. To obtain the dense KN-based ceramic, the hot-press (HP) method or liquid-phase sintering by additive dopants was investigated [31,33]. On the other hand, electrical properties of potassium–sodium niobates, KNbO$_3$–NaNbO$_3$ [KNN] system, were reported by Egerton et al. [11,14]. Their works on ceramics in the system indicated that relatively low dielectric constants and high electromechanical coupling factors could be obtained over a wide range of compositions. However, it is difficult to realize the desired structure in a ceramic form because the sintering of these materials in an air requires long soaking periods to achieve sufficient densification. There have been many reports on KN solid-solution systems such as KNbO$_3$–NaNbO$_3$ [KNN]. Good piezoelectric properties, such as a large planar coupling factor, $k_p$ = 0.56, and a large remanent polarization, $P_r$ up to 30 μC/cm$^2$, have been observed for KNN ceramics [11]. An excellent piezoelectric property of 416 pC/N in textured (K$_{0.5}$Na$_{0.5}$)NbO$_3$-based ceramics has recently been reported by Saito et al. [35]. Figure 1 shows (a) the piezoelectric strain constant, $d_{33}$, as a function of the content of Ta and Li ions and (b) the $d_{33}$, as a function of the Curie temperature [35]. Table 1 summarizes piezoelectric properties of the textured LF4T compared with those of PZT4 [35]. Recently, Guo et al. [36] also reported that $(1-x)(K,Na)NbO_3$–xLiNbO$_3$ solid solution $(1-x)$ KNN-xLN shows the MPB composition between the orthorhombic and the tetragonal symmetries near 0.05 < x < 0.07 with the higher Curie temperature of 450 °C and the higher $d_{33}$ of 240 pC/N in Figures 2 and 3, respectively.

2.3. KNbO$_3$ [KN]-based Ceramics

In terms of KNbO$_3$ [KN] ceramic, there are few papers and limited number of reports on electrical properties because of a deliquescence behavior and a poor sinterability of KN ceramic by the conventional fabrication process in air. Recently, the dense KN-based ceramics could be obtained by using the modified conventional ceramic firing technique with good dielectric, ferroelectric, and piezoelectric properties [44–46]. A developed preparation procedure has been proposed to improve the deliquescence and sinterability behavior of KN-based ceramics such as KNbO$_3$ [KN] and KNbO$_3$ + MnCO$_3$ x wt% [KN-Mn x]. To fix the problems of deliquescence and poor sinterability behavior, two conditions were optimized in whole sample preparation process. The first one is a calcination process for improvement of the deliquescence behavior, and the second one is a milling process for achieving fine KN powders and finally dense KN-based ceramic bodies.

Figure 4 shows the temperature dependences of the dielectric constant $\varepsilon_s$ and the loss tangent tan δ of the KN ceramic calcined at (a) 1000 and (b) 600 °C. The Curie temperature $T_c$ of the KN ceramic calcined at 1000 °C was 424 °C and the phase transition temperature $T_2$ from an orthorhombic symmetry to a tetragonal symmetry was approximately 230 °C. There were no significant changes in $T_c$ or $T_2$ for both (a) and (b) ceramics.
The KN samples for piezoelectric properties were poled at 5 kV/mm for 2–5 min at 150 °C. Note that the sample in the (33) mode was approximately in the full-poling state as judged from the maximum phase, $\theta_{\text{max}} = 87.4^\circ$, of impedance in the inductance region between the resonance and anti-resonance frequencies with the electromechanical coupling factor, $k_{33} = 0.492$, the piezoelectric constant, $d_{33} = 91.7 \text{ pC/N}$, and the mechanical quality factor, $Q_m = 325$, respectively. Figure 5 shows the temperature dependences of $d_{33}$ and $k_{33}$ for the KN ceramic as determined by the resonance–anti-resonance method. The piezoelectric response disappeared above 430 °C, which corresponded to $T_c$. The $k_{33}$ values in the orthorhombic region from RT to 225 °C were almost stable at approximately 0.5. Above the phase transition temperature, $T_2$, the $k_{33}$ slightly decreased to approximately 0.35. The $d_{33}$ was constant at about 100 pC/N below 200 °C and showed a maximum larger than 200 pC/N at $T_2$ of 225 °C.

The sintering temperatures of Mn-doped KN ceramics shift to lower temperatures and the sintering temperature range gradually becomes wide with increasing MnCO$_3$ content. The milled powders had fine particles as well as
submicrons. The preparation of a fine particle powder is one of the key points to obtaining dense KN-based ceramics as reported by Birol et al. [40] The crystal structures of KN and KN$_{1-x}$Mn$_x$ (x = 0.05 – 1.6) were determined by the X-ray diffraction patterns. All of sintered ceramics had a single phase of perovskite structure with orthorhombic symmetry.

Figure 6 shows the P–E hysteresis loops of KN–Mn0.05 ceramics at RT and 200 °C with well-saturated loops. The remanent polarization, $P_r$, was 23 μC/cm$^2$ and the coercive field, $E_c$, was 13 kV/cm at RT. This result indicates that Mn-doping for KN is effective in obtaining high-resistance specimens under high electric fields and high temperatures.

The coupling factor, $k_{33}$, of KNbO$_3$ + MnCO$_3$ x wt% [KN–Mn x] decreased with increasing the amount of doped Mn ions. This tendency is similar to the relative density behavior as a function of Mn content. Figure 7 shows the frequency dependence of impedance on the (33), (31), and (p)-modes for the KN–Mn0.1. This sample was poled under 5 kV/mm at 150 °C. The phase, $\theta$, in the impedance for the (33) mode reversed very well from capacitive to inductive region, and the maximum phase, $\theta_{\text{max}}$, reached 88.1° which is very close to 90°. Electromechanical coupling factor, $k_{33}$, indicated 0.51. The $k_{33}$ values of KN single crystal are 0.61 and 0.56, reported by S. Wada et al [32] and M. Zgonik et al [47], respectively. The $k_{33}$ value of the KN–Mn0.1 ceramic is about 84–91% for that of the single crystal and is reasonable value compared with that of the full-poling state in randomly oriented ceramics with the orthorhombic symmetry. In the case of the (31) mode, the $\theta_{\text{max}}$ was 77.0°. It is not high enough to say that the specimen was poled fully. It is probably due to insufficiency of applied electric field, $E_a$, in the poling treatment. At present, we cannot apply higher $E_a$ without a surface discharge and breakdown of the specimen. The $k_{31}$ showed 0.19, which is lower than that expected from the single crystal. It is also probably due to the difficulty in poling.

On the other hand, the piezoelectric properties for the thickness-shear mode have not been investigated enough in the ceramic form, even though the KN single crystal shows large shear mode piezoelectric properties such as $k_{34}$ and $d_{34}$ [32]. This is due to a problem about spurious peaks in the resonance and anti-resonance characteristics. Therefore, it has been difficult to estimate piezoelectric
properties of (15) mode for the KN-based ceramic and which has not been reported sufficiently so far. Figure 8 shows frequency dependence of impedance, $Z$, in the (15) mode for KN-Mn0.1 with a wide frequency range. The fundamental frequency, $f_{(1)}$, and third and fifth harmonics, $f_{(3)}$ and $f_{(5)}$, can be clearly observed. The $k_{15}$ values [48] from $f_{(3)}/f_{(1)}$ and $f_{(5)}/f_{(3)}$ were 0.48 and 0.50, respectively, which were almost consistent with those calculated by the resonance–anti-resonance method.

To minimize the influence of spurious peaks, we prepared the sample with larger aspect ratio by decreasing the sample thickness to 0.2 mm. Figure 9 shows the frequency dependence of impedance, $Z$, in the (15) mode for the KN-Mn0.1 ceramic of 0.2 × 8 mm$^2$. It is easy for the $k_{15} = 0.55$ in Figure 9 to be calculated using the resonance and the anti-resonance frequencies. This $k_{15}$ of the ceramic KN seems to be a reasonable value as compared with those of the $k_{15} (0.44)$ and the $k_{24} (0.88)$ from the KN single crystal [32,49].

Table 2 summarizes piezoelectric properties of the KN–Mn0.1 ceramic. The relative free permittivity, $\varepsilon_{11}/\varepsilon_0$, was 611. Piezoelectric strain constant, $d_{15}$, reached 207 pC/N which was large because of large $k_{15} (=0.55)$, $\varepsilon_{11}/\varepsilon_0 (=611)$, and $\varepsilon_{33}^E (=25.7 \text{ pm}^2/N)$. It indicated the possibility that larger $d_{15}$ could be obtained as the poling state was further enhanced. The shear mode vibration of KN–Mn0.1 ceramic seems to be a considerably promising candidate for actuator and high-power applications.

### 2.4. (Bi$_{1/2}$Na$_{1/2}$)TiO$_3$ [BNT]-based Ceramics

Bismuth sodium titanate, (Bi$_{1/2}$Na$_{1/2}$)TiO$_3$ [BNT] [12,13,16,43,69–71], is a perovskite-structured ferroelectric with rhombohedral symmetry (R3C) at room temperature (RT) and their phase transitions are complicated. The phase transition temperatures, $T_{R\rightarrow T}$, from rhombohedral to tetragonal (the temperature $T_m$ of the maximum dielectric constant), and $T_{T\rightarrow C}$, from tetragonal to cubic (Curie temperature, $T_C$), are approximately 340 °C and 540 °C on heating, respectively, for BNT single crystals [70,71,79].

The BNT ceramic shows the strong ferroelectric properties of a large remanent polarization, $P_r = 38 \mu\text{C/cm}^2$, and relatively high piezoelectric properties compared with other lead-free piezoelectric ceramics. Therefore, the BNT is considered to be an excellent candidate as a key material of lead-free piezoelectric ceramics. However, data on piezoelectric properties of the BNT ceramic are scarce because it is difficult to pole this ceramic due to a large coercive field, $E_c (=73 \text{ kV/cm})$, except in specialized work [43]. In the last two decades, BNT-based solid solutions [15–29] and A-site-substituted BNT [72–74] that can be poled easily were studied. The electromechanical coupling factor, $k_{33}$, of the BNT ceramic as the end member of solid solutions varies from 0.25 to 0.40 because of the difference in the sintering conditions. The BNT ceramic needs a high sintering temperature of more than 1200 °C to obtain a dense body. It is thought that the vaporization of Bi ions occurred during the sintering process at temperatures

![Figure 6](image6.png)

**FIGURE 6** $P$–$E$ hysteresis loops of KN–Mn0.05 at RT and 200 °C.

![Figure 7](image7.png)

**FIGURE 7** Frequency dependences of impedance, $Z$, in (33), (31), and (p) modes for KN–Mn0.1.
higher than 1200 °C, resulting in the poor poling treatments because of the low resistivities. From the thermograph (TG, weight loss) measurement, the weight loss caused by the Bi vaporization was carried out at over 1130 °C. Various processes and methods are thought to prevent the Bi vaporization and to obtain the stoichiometric BNT ceramic. Thus, the BNT ceramic should be sintered at 1100 °C and lower temperatures.

Two approaches were attempted to sinter the BNT ceramic with higher density and higher resistivity as follows [43]: (1) the addition of Bi$_2$O$_3$ to the BNT ceramic (BNT + Bi$_2$O$_3$ x wt% [BNT-x]) sintered at 1225 °C for 5 min to 10 h to compensate for the insufficiency of Bi ions and (2) the low sintering temperature process with a long soaking time of 30–100 h in the ordinary firing method and the hot-pressing (HP) method to suppress the Bi vaporization.

Figure 10 shows the resistivity, $\rho$, of BNT-x as a function of excess Bi$_2$O$_3$ content (x). The $\rho$ shows the largest value of about $10^{14}$ Ω-cm at $x = 0.3$. It is clear that the $\rho$ is improved by adding the excess Bi$_2$O$_3$ and the optimum charge neutrality was observed for the BNT-0.3. Therefore, the BNT-0.3 seems to be a stoichiometric neutral state of the BNT.

Two approaches under low sintering temperature were tried to increase the density. The first one is to keep the longer soaking time at low temperature during the ordinary

### Table 2: Piezoelectric Properties in Various Vibration Mode of KN–Mn0.1 Ceramic

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative free permittivity</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon_{33}/\varepsilon_{0}$</td>
<td>291</td>
</tr>
<tr>
<td>$\varepsilon_{11}/\varepsilon_{0}$</td>
<td>611</td>
</tr>
<tr>
<td>Electromechanical coupling factor</td>
<td></td>
</tr>
<tr>
<td>$k_{33}$</td>
<td>0.51</td>
</tr>
<tr>
<td>$k_{31}$</td>
<td>0.19</td>
</tr>
<tr>
<td>$k_{15}$</td>
<td>0.55</td>
</tr>
<tr>
<td>$k_{0}$</td>
<td>0.32</td>
</tr>
<tr>
<td>$k_i$</td>
<td>0.37</td>
</tr>
<tr>
<td>Frequency constant</td>
<td></td>
</tr>
<tr>
<td>$N_{33}$</td>
<td>2822</td>
</tr>
<tr>
<td>$N_{11}$</td>
<td>2730</td>
</tr>
<tr>
<td>$N_{15}$ (Hz m)</td>
<td>1798</td>
</tr>
<tr>
<td>$N_p$</td>
<td>3535</td>
</tr>
<tr>
<td>$N_t$</td>
<td>3093</td>
</tr>
<tr>
<td>Piezoelectric strain constant</td>
<td></td>
</tr>
<tr>
<td>$d_{33}$</td>
<td>80.8</td>
</tr>
<tr>
<td>$d_{31}$ ($\mu$C/N)</td>
<td>-29.7</td>
</tr>
<tr>
<td>$d_{15}$</td>
<td>207</td>
</tr>
<tr>
<td>Piezoelectric voltage constant</td>
<td></td>
</tr>
<tr>
<td>$g_{33}$</td>
<td>31.4</td>
</tr>
<tr>
<td>$g_{31}$ ($\times 10^3$ V m/N)</td>
<td>-11.5</td>
</tr>
<tr>
<td>$g_{15}$</td>
<td>34.2</td>
</tr>
<tr>
<td>Elastic compliance constant</td>
<td></td>
</tr>
<tr>
<td>$s_{33}$</td>
<td>9.74</td>
</tr>
<tr>
<td>$s_{11}$ ($\mu$m²/N)</td>
<td>7.54</td>
</tr>
<tr>
<td>$s_{55}$</td>
<td>25.8</td>
</tr>
<tr>
<td>Mechanical quality factor</td>
<td></td>
</tr>
<tr>
<td>$Q_m$ ($33$)</td>
<td>391</td>
</tr>
<tr>
<td>$Q_m$ ($15$)</td>
<td>132</td>
</tr>
</tbody>
</table>
sintering process, and the other one is to utilize the HP method to prepare the BNT ceramic with high density. The relative density ratio of the BNT-0 with no excess Bi$_2$O$_3$ increases with increasing the soaking time in the ordinary firing process at 1100 °C. The highest density ratio of 96% was obtained at 100 h. On the other hand, densities of the hot-pressed BNT-0 (HP-BNT-0) ceramics under some conditions were higher than those of the non-hot-pressed one. The HP-BNT-0 ceramic pressed at 200 kg/cm$^2$ under 1100 °C for 30 min showed a highest density ratio of more than 98%, and might be expected to have excellent piezoelectric properties.

Figure 11 shows the frequency characteristics of the impedance, $Z$, for BNT-0 and HP-BNT-0 ceramics. The $k_{33}$ of the BNT-0 sintered at 1100 °C for 100 h and the HP-BNT-0 by the HP method at 1100 °C for 30 min were 0.44 and 0.48, respectively.

Table 3 summarizes the dielectric and piezoelectric properties of BNT-0.3 and HP-BNT-0 ceramics. The BNT ceramic is superior as a key material for some lead-free piezoelectric applications because of the relatively high electromechanical coupling factor, $k_{33}$, and the relatively high piezoelectric strain constant, $d_{33}$. The $k_{33}$ and the $d_{33}$ of the BNT-0.3 sintered at 1225 °C for 30 min and the HP-BNT-0 by hot-pressed at 1100 °C for 30 min were 0.47 and 92 pC/N, and 0.48 and 93 pC/N, respectively.

2.5. (Bi$_{1/2}$Na$_{1/2}$)TiO$_3$ [BNT]—(Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BKT]—BaTiO$_3$ [BT] System

The important feature is that the BNT forms a morphotropic phase boundary (MPB) with the other perovskites having a tetragonal symmetry, such as PbTiO$_3$ [PT], BaTiO$_3$ [BT], and (Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BKT], and these MPB compositions show excellent piezoelectric properties [20,29,34,41,51–53,56,75]. Therefore, BNT-based solid solutions have attracted considerable attention as lead-free piezoelectric materials because of the MPBs. Also, the easy preparation of the dense BNT-based ceramics is an advantage for manufacture.

Two solid-solution systems, that is, (1–$x$)BNT–$x$BT (BNBT100$^x$) and (1–$y$)BNT–$y$BKT (BNKT100$^y$), have already been reported by Takenaka et al. [20] and Sasaki et al. [29], respectively. It is reported that the MPB existed at $x = 0.06$–0.07 for BNBT100$^x$ and $y = 0.16$–0.20 for BNKT100$^y$, respectively. Then, the dielectric and piezoelectric properties of the three-component system, $a$(Bi$_{1/2}$Na$_{1/2}$)TiO$_3$–$b$(Bi$_{1/2}$K$_{1/2}$)TiO$_3$–$c$BaTiO$_3$ (BNBK), were investigated [34,41,51,52], focusing on the MPB compositions. Figure 12 shows the phase relation of the BNK system around the MPB area. The MPBs of both BNBT6 and BNKT16 exist on the rhombohedral side, and the MPBs of both BNBT7 and BNKT20 exist on the tetragonal side around the MPB region, respectively. It was found by X-ray diffraction that the MPB between the rhombohedral and tetragonal phases exists between the two systems of BNBK1 [a(BNBT6)–(1–a)(BNKT16)] and BNBK2 [a(BNBT7)–(1–a)(BNKT20)] ($a = 0, 0.2, 0.4, 0.6, 0.8$, and 1, for each system).

Figure 13 shows the compositional dependence of the piezoelectric constant, $d_{33}$, for BNBK2:1(x), where $x$BNT–$y$BKT–$z$BT, [x + y + z = 1, $y$:$z = 2$:1], as
a function of the content (x) of BNT. High $d_{33}$ values were obtained near the MPB composition, being highest just on the tetragonal side. A maximum $d_{33}$ value of 181 pC/N was obtained in BNBK2:1(0.88). On the tetragonal side, the $d_{33}$ decreased because both the $k_{33}$ and the free permittivity, $\epsilon_r$, decreased. The $d_{33}$ values of BNBK2:1(0.98) and BNBK2:1(0.30) were found to be 80.3 and 79.9 pC/N, respectively.

It is important for BNT-based solid solutions to investigate the actual working temperature for use in practical applications. For that reason, the depolarization temperature, $T_d$, was accurately determined from the temperature dependence of piezoelectric properties [51–53]. The temperature dependences of $k_{33}$ are shown in Figure 14 for (a) the rhombohedral side of $x = 0.92–0.98$, (b) the MPB composition of $x = 0.88–0.90$, and (c) the tetragonal side of $x = 0.30–0.84$ in BNBK2:1(x), respectively. It is recognized that the rhombohedral and tetragonal sides exhibit more stable $k_{33}$ values with temperature than those for compositions around the MPB. The $T_d$ and the temperature, $T_m$, with the maximum dielectric constant in the temperature dependence of BNBK2:1(x) are summarized in Figure 15. The $T_m$ was determined by the temperature of the maximum $\epsilon_r$ in the temperature dependence of dielectric constant, $\epsilon_r$. The $T_m$ of BNBK2:1(x) was approximately the same at 300 °C; however, the $T_d$ was largely decreasing near the MPB composition and increased with decreasing BNT content (x). These results indicate that the $T_d$ is dependent on lattice anisotropy (tetragonality), $c/a$. Figure 16 demonstrates the relationship between the $T_d$ and the $d_{33}$ of the tetragonal and rhombohedral sides. It is necessary for actual applications to obtain the value in the right upper corner (high $d_{33}$ and high $T_d$) of this figure. However, the $d_{33}$ and the $T_d$ indicated a trade-off relationship. The $d_{33}$ as a function of the $T_d$ on the tetragonal side is higher than that on the rhombohedral side.

Figure 17 [6] shows the strain, $S$, of BNBK2:1(0.88), BNBK2:1(0.78), and BNBK2:1(0.98) as a function of the applied electric field, $E_a$, measured at 0.1 Hz. The magnitudes of their strains, $S$, are $S$ (MPB) > $S$ (tetragonal) > $S$ (rhombohedral) for their compositions. Table 4 [6] summarizes the piezoelectric properties of BNBK2:1(x) ($x = 0.78, 0.88$ and $0.98$) including the $d_{33}^*$ and the $S$ obtained from the results of Figure 17. The $d_{33}^*$ defined by the equation,

$$d_{33}^* \text{[pm/V]} = \frac{\text{Strain}, \frac{S}{E_a \text{[kV/mm]}}}{10^6}$$  \hspace{1cm} (1)$$

for example, is 188 pm/V with relatively high $T_d$ (206 °C) in the tetragonal composition ($x = 0.78$). Finally, Table 5 summarizes the depolarization temperatures, $T_d$, and piezoelectric properties of rhombohedral, MPB, and tetragonal compositions of the BNBK2:1(x) ternary system.
2.6. (Bi$_{1/2}$Na$_{1/2}$)TiO$_3$ [BNT]—(Bi$_{1/2}$Li$_{1/2}$)TiO$_3$ [BLT]—(Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BKT] system

The BNT presents the low depolarization temperature $T_d$ of 185 °C, and the MPBs of BNT-based solid solutions show a particularly low $T_d$ of approximately 100 °C [51, 53, 56, 76, 77], even though the MPBs have excellent piezoelectric properties. The piezoelectric working temperature range is limited to below the $T_d$ because piezoelectric properties disappear at the $T_d$ and higher temperatures. Therefore, it is important to increase the $T_d$ of the MPB composition to enable its practical use in piezoelectric applications. From the effects [53] of Li- and K-substitution on the $T_d$ in the A-site of BNT, it is seen that the $T_d$ of BNT increases when the small amounts of Li and K are substituted. Phase transition temperatures, $T_d$, $T_{R-T}$, and $T_m$, of a solid solution, (1$-x$)(Bi$_{1/2}$Na$_{1/2}$)TiO$_3$-$x$(Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BNKT100x] are summarized in Figure 18. Field-induced strains of BNKT100x under unipolar driving at 0.1 Hz are shown in Figure 19, and the piezoelectric constant, $d_{33}$, and the normalized strain, $d_{33}^* ( = S_{max}/E_{max} )$, at 80 kV/cm are summarized in Figure 20. All $d_{33}^*$ values were higher than the $d_{33}$, because the $d_{33}^*$ includes the domain contribution due to the high applied voltage and the low measuring frequency [80]. The highest value was obtained near the MPB composition for BNKT100x. Moreover, the ratios of the $d_{33}^*$ to the $d_{33}$ are larger for the tetragonal compositions than for the rhombohedral compositions due to the difference in the domain structures.

Double substitution of Li and K is thought to be effective in increasing the $T_d$ of BNT-based solid solutions. Figure 21 shows the phase relation of $x$(Bi$_{1/2}$Na$_{1/2}$)
Li-substitution is effective in increasing $T_d$ of BNT-based solid solution. In addition, $T_d$ of BNLKT4-100z and BNLKT8-100z showed the maximum at $z = 0.08$ at the rhombohedral side as the same as BNLKT0-100z. Although very few data on increase $T_d$ have been reported previously at the rhombohedral composition of BNT-based solid solutions [59], the $T_d$ could be increased up to 221 °C for BNLKT4-8.

Recently, piezoelectric ceramics have attracted much attention for high-power devices, such as ultrasonic motors and transducers [82–85]. These devices were mainly composed of Pb(Zr,Ti)O$_3$ (PZT)-based piezoelectric ceramics with a high mechanical quality factor $Q_m$, which is called hard PZT. However, the resonant vibration of hard PZT becomes unstable at a vibration velocity of approximately 1.0 m/s, resulting in $Q_m$ markedly decreasing and vibration velocity not increasing. Moreover, PZT contains a large amount of PbO; therefore, recently, there has been much interest in lead-free piezoelectric ceramics as a material for replacing PZT-based ceramics. As materials for lead-free high-power applications, SrBi$_2$Nb$_5$O$_{15}$ [86,87] and (Sr,Ca)$_2$NaNb$_5$O$_{15}$ [61] have been reported.

The relationship between the $T_d$ and the $d_{33}$ and also the piezoelectric properties of BNLKT100y-100z system indicate that it has the high possibility of use for high-power applications including wt% MnCO$_3$-doped BNLKT100y-100z (abbreviated to BNLKT100y-100zMnw). Figure 23 shows the piezoelectric strain constant, $d_{33}$, and the mechanical quality factor, $Q_m(33)$, in the 33-mode, of BNLKT4-100z as a function of the content ($z$) of BKT. The $d_{33}$ reaches about 180 pC/N at the MPB composition ($z = 0.20$). On the other hand, the $Q_m(33)$ was the highest, of approximately 200 at $z \leq 0.08$, and decreased to below 90 at $z \geq 0.18$. Figure 24(a) shows the temperature dependence of the $k_{33}$ for BNLKT4-8Mnw ($w = 0$ and 0.6). It is found that the $k_{33}$ is almost constant up to the $T_d$. Moreover, the $T_d$ gradually decreases with increasing the Mn concentration ($w$), as shown in Figure 24(b), with $T_d = 204$ °C for BNLKT4-8Mn0.6.

The decrease in the $T_d$ with increasing the Mn content ($w$) indicates that the Mn ion substituted into either the A-site or B-site of BNLKT4-8. According to previous reports, both $\varepsilon_{\text{max}}$ and $T_m$ decrease with an increasing

<table>
<thead>
<tr>
<th>$X$ in BNBK2:1(x)</th>
<th>$T_d$ [°C]</th>
<th>$K_{33}$</th>
<th>$d_{33}$ [pC/N]</th>
<th>$d_{33}'$ [pm/V] at $E_a = 80$ kV/cm</th>
<th>Strain [%] at 80 kV/cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.98</td>
<td>Rhombohedral</td>
<td>200</td>
<td>0.42</td>
<td>80</td>
<td>121</td>
</tr>
<tr>
<td>0.88</td>
<td>MPB</td>
<td>113</td>
<td>0.56</td>
<td>181</td>
<td>240</td>
</tr>
<tr>
<td>0.78</td>
<td>Tetragonal</td>
<td>206</td>
<td>0.45</td>
<td>126</td>
<td>188</td>
</tr>
</tbody>
</table>

| TABLE 4 Piezoelectric Properties of BNBK2:1(x) ($x = 0.78$, 0.88 and 0.98) |
amount of Mn in BNT and BNT-BT [88,89]. In BNLKT4-8Mn, although the temperature dependence of dielectric constant showed that the $\varepsilon_{\text{max}}$ decreased with increasing $w$, the $T_m$ was almost constant at 272 °C [81].

Considering the valence state of Mn ions in BNT, the $\text{Mn}^{4+}$ substitution into the B-site probably decreases $T_m$ because the ionic radius of Mn$^{4+}$ is 0.530 Å, which is similar to that of Al$^{3+}$. In contrast, the substitution of Mn$^{2+}$ and Mn$^{3+}$ into the A-site or B-site of BNT should increase the $T_m$ of BNT. Therefore, the decay of $T_d$ and the lack of variation in $T_m$ indicate that Mn ions exist in BNLKT4-8, in the mixed state of Mn$^{2+}$ or Mn$^{3+}$, and Mn$^{4+}$. The resistivity was maximum at $w = 0.2$, and, then, decreased with increasing Mn concentration $w$. The vaporization of A-site ions such as Bi, Na, and K occurs during sintering. Therefore, the very small amount of Mn probably compensates for the A-site vacancies as a donor, resulting in an increase in resistivity [88,89]. In contrast, Mn$^{2+}$ or Mn$^{3+}$ in the B-site works as an acceptor; therefore, resistivity decreases at $w > 0.2$. Generally, acceptor ions associate with oxygen vacancies, and cause
domain pinning, thereby increasing in $Q_m$ [82,90–92]. Therefore, the $Q_m$ of BNLKT4-8Mn increases with increasing $w$.

The high-power characteristics of BNLKT4-8, BNLKT4-8Mn0.6, and PZT-H were evaluated by the high-power characteristic measurement [85]. The small-amplitude $Q_m(31)$ values of BNLKT4-8, BNLKT4-8Mn0.6, and PZT-H are 440, 740, and 1770, respectively. Figure 25 shows the variation in $Q_m(31)$ as a function of the vibration velocity $v_{0-p}$. It is found that the $Q_m(31)$ values of BNLKT4-8 and BNLKT4-8Mn0.6 decrease slower than those of PZT-H. Although the small-amplitude $Q_m(31)$ of PZT-H was twice higher than that of BNLKT4-8Mn0.6, that of BNLKT4-8Mn0.6 was larger than that of PZT-H at $v_{0-p} > 0.6$ m/s. Moreover, the $Q_m(31)$ of BNLKT4-8Mn0.6 was higher than 400, even at $v_{0-p} = 1.5$ m/s.

It is considered that the small decay of $Q_m(31)$ under a high-amplitude vibration for BNLKT4-8 and BNLKT4-8Mn0.6 is attributed to the high coercive field $E_c$. It is known that the $E_c$ at the rhombohedral composition is larger than that at the MPB and on the tetragonal side for BNT-based solid solutions [51]. The BNLKT4-8 has high $E_c$ of 60 kV/cm, which is 4 times larger than that of PZT-H. Moreover, domain-wall motion is suppressed by oxygen vacancies associated with the doping of acceptor ions [82]. Therefore, the acceptor-ion-doped rhombohedral composition of BNT-based solid solutions is stable even at a vibration velocity higher than PZT-H.

2.7. (Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BKT]-based Ceramics

Bismuth potassium titanate, (Bi$_{1/2}$K$_{1/2}$)TiO$_3$ [BKT], is a typical lead-free ferroelectric with a perovskite structure of tetragonal symmetry at room temperature and has a relatively high Curie temperature, $T_c$, of 380 °C [13]. This indicates that BKT has a certain promise as a candidate for lead-free piezoelectrics in a wide working temperature range. However, there are few reports about this material owing to its poor sinterability, except for our reports [42].
This problem has restricted extensive activities of researchers in investigations for the BKT-based solid-solution systems.

Recently, Hiruma et al. [42, 93] reported the electrical properties of BKT ceramics prepared by the hot-pressing (HP) method. The optimum sintering temperature seems to be $1060 \text{ } ^\circ\text{C} \sim 1080 \text{ } ^\circ\text{C}$. Figure 26 shows the temperature dependences of dielectric constant, $\varepsilon_s$, and dielectric loss tangent, $\tan \delta$, for the hot-pressed BKT ceramics sintered at (a) $1060 \text{ } ^\circ\text{C}$ and (b) $1080 \text{ } ^\circ\text{C}$.

The $\tan \delta$ curves in Figure 26 show two peaks. High-temperature peaks show frequency dispersions; therefore, it is thought that these peaks are related to the $T_c$. On the other hand, low-temperature peaks are almost independent of frequencies; therefore, it is considered to indicate the second phase transition, $T_2$, between tetragonal and pseudo-cubic symmetries, whose temperatures of BKT-HP1060 $^\circ\text{C}$ and BKT-HP1080 $^\circ\text{C}$ are about 340 and 315 $^\circ\text{C}$, respectively. The $T_c$ of Bi$_2$O$_3$, La$_2$O$_3$, or MnCO$_3$-doped BKT ceramics showed a tendency to decrease with an increase in the content of dopants.

Figure 27 shows $D-E$ hysteresis loops of BKT-HP1060 $^\circ\text{C}$ and BKT-HP1080 $^\circ\text{C}$. Well-saturated $D-E$ hysteresis loops with low leakage current were obtained for all specimens at RT. The $P_t$ and the coercive fields $E_c$ were 22.2 $\mu$C/cm$^2$ and 52.5 kV/cm for BKT-HP1080 $^\circ\text{C}$, and 14.2 $\mu$C/cm$^2$ and 47.3 kV/cm for BKT-HP1060 $^\circ\text{C}$, respectively. The different tendencies seem to be the difference in grain size 0.2 $\mu$m for BKT-HP1060 $^\circ\text{C}$ and 0.4 $\mu$m for BKT-HP1080 $^\circ\text{C}$.

A solid-solution system, $(1-x)$ (Bi$_{1/2}$K$_{1/2}$)TiO$_3$-xBaTiO$_3$ [BKT–BT100x] [6, 95], was investigated for evaluating their characterizations and electrical properties by using randomly oriented and grain-oriented samples. Especially, the compositions near the BKT ($x = 0 \sim 0.4$) were focused as the lead-free piezoelectric ceramics with the wide working temperature. X-Ray diffraction patterns of BKT–BT100x ceramics with $0 \leq x \leq 1$ show a single phase of perovskite structure with tetragonal symmetry at room temperature. The sintered ceramics of BKT–BT indicated higher relative densities than 95%, even in the compositions of the BKT side as shown in Table 6.

Figure 28 shows the temperature dependence of electromechanical coupling factor, $k_{33}$, and the phase, $\theta$, in the impedance–frequency characteristics of the (33)-mode for BKT [42] ceramic. This figure indicates that the BKT and BKT–BT ceramics seem to be attractive for
higher-temperature applications compared with the BT ceramic.

Figure 29 shows lattice constants, $a$ and $c$, and lattice anisotropy, $c/a$, as a function of the amount ($x$) of BT in BKT–BT100x ceramics. Both BKT ($x = 0$) and BT ($x = 1$) have the same crystal structure of tetragonal symmetry; however, the $c/a$ indicated nonlinear tendency. The $c/a$ shows the maximum (~1.025) at the composition around $x = 0.2$, which is a very large value among the lead-free piezoelectric materials. This result corresponds almost to that shown in Buhrer’s report [13].

From temperature dependences of dielectric constant, $\varepsilon_r$ and loss tangent, $\tan \delta$, for the BKT–BT100x ceramics, the Curie temperature, $T_c$, linearly shifted to lower temperatures with an increase in the amount of BT content ($x$), as shown in Figure 30. The $T_c$ of BKT–BT80 ($x = 0.8$) shows still a higher temperature than 200°C. However, both the $\varepsilon_r$ at RT and at the $T_c$ decrease with increasing $x$. The $T_2$ in Figure 30 shows the second phase transition temperature from tetragonal to pseudo-cubic phases, existing near 270°C in BKT. The $T_2$ of BKT ceramic was reported by Ivanova [96].

Furthermore, grain orientation effects for piezoelectric properties were investigated in BKT–BT100x using a reactive template grain growth (RTGG) method [97]. Piezoelectric strain constant, $d_{33}$, for randomly oriented BKT–BT10, 20, and 30 are 73.4, 69.1, and 67.6 pC/N, respectively. These values are relatively small for practical use to actuators. So, we tried to prepare the textured samples by the RTGG method to enhance their piezoelectric properties. Textured specimens were prepared using the RTGG method with matrix and templates of plate-like Bi$_4$Ti$_3$O$_{12}$ (BIT) particles for BKT–BT. Calcination and sintering temperatures were 900–1000°C and

<table>
<thead>
<tr>
<th>$\rho_0/\rho_x$ (%)</th>
<th>OF</th>
<th>RTGG</th>
<th>OF</th>
<th>RTGG</th>
<th>OF</th>
<th>RTGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$ (%)</td>
<td>98.5</td>
<td>97.8</td>
<td>99.3</td>
<td>95.4</td>
<td>98.0</td>
<td>94.3</td>
</tr>
<tr>
<td>$k_{33}$</td>
<td>0.35</td>
<td>0.37</td>
<td>0.36</td>
<td>0.33</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>$\varepsilon_{33}/\varepsilon_0$</td>
<td>602</td>
<td>560</td>
<td>532</td>
<td>501</td>
<td>461</td>
<td>426</td>
</tr>
<tr>
<td>$s_{33}^e$ (pm$^2$/N)</td>
<td>8.2</td>
<td>10.5</td>
<td>8.2</td>
<td>10.1</td>
<td>7.8</td>
<td>13.0</td>
</tr>
<tr>
<td>$d_{33}$ (pC/N)</td>
<td>73.4</td>
<td>84.5</td>
<td>69.3</td>
<td>70.7</td>
<td>67.6</td>
<td>83.3</td>
</tr>
<tr>
<td>$d_{33}'$ (pm/V)</td>
<td>103</td>
<td>168</td>
<td>116</td>
<td>143</td>
<td>103</td>
<td>134</td>
</tr>
</tbody>
</table>

$\rho_0$: relative density ratio, $F$: orientation factor, $k_{33}$: electromechanical coupling factor, $\varepsilon_{33}$: free permittivity, $\varepsilon_0$: elastic constant, $d_{33}$: piezoelectric strain constant, $d_{33}'$: calculated $d_{33}$ by the Eqn 1.
The dense KN-based ceramics, KNbO$_3$+MnCO$_3$ x wt% [KN-Mn x], can be obtained by using the modified conventional ceramic firing technique with good dielectric, ferroelectric, and piezoelectric properties. The electromechanical coupling factor, $k_{33} = 0.55$, and the piezoelectric strain constant, $d_{33} = 207$ pC/N, are obtained at Mn-doped KN–Mn0.1. The Bi-excess bismuth sodium titanate, (Bi$_{1/2}$Na$_{1/2}$)TiO$_3$ (BNT) + Bi$_2$O$_3$ x wt% [BNT-x], and hot-pressing BNT [HP-BNT] ceramics were prepared and their piezoelectric properties were investigated. All of BNT-x ceramics sintered at 1225 °C showed a high density ratio, more than 97% to the theoretical density. The BNT-0.3 seems to be a stoichiometry because of the measurement results from the resistivity, $\rho$, Curie temperature, $T_c$, and microstructure. BNT-0.3 ceramic showed a relatively large electromechanical coupling factor, $k_{33} (=0.47)$, and piezoelectric constant, $d_{33} (=93$ pC/N). The large piezoelectricity, $k_{33} (=0.48)$ and $d_{33} (=98$ pC/N), with the high density ratio (98%), could be obtained for the first time on the HP-BNT ceramic sintered at 1100 °C for 30 min and pressed at 200 kg/cm$^2$ as lead-free piezoelectric materials.

In the case of the ternary system $x$(Bi$_{1/2}$Na$_{1/2}$)TiO$_3$-$y$(Bi$_{1/2}$K$_{1/2}$)TiO$_3$-$z$BaTiO$_3$ ($x + y + z = 1$), [BNBKy$z$ ($x$)], enhanced piezoelectric properties were obtained near the MPB composition, and the highest electromechanical coupling factor, $k_{33}$, and $d_{33}$ were 0.58 for BNBK2:1(0.89) and 181 pC/N for BNBK2:1(0.88). Nevertheless, the $T_d$ shift to lower temperatures (about 100 °C) around the MPB compositional shift to a value higher than 200 °C for $x < 0.78$, with $k_{33} = 0.452$ and $d_{33} = 126$ pC/N for BNBK2:1(0.78). This ternary system shows high potential (with properties such as large $d_{33} > 250$ pC/N and high $T_d > 250$ °C) for actuator applications.

A small amount of Li-substitution was very effective for increasing the $T_d$ because the $T_d$ is related to the variation of lattice distortion. The $T_d$ of $x$(Bi$_{1/2}$Na$_{1/2}$)TiO$_3$-$y$(Bi$_{1/2}$Li$_{1/2}$)TiO$_3$-$z$(Bi$_{1/2}$K$_{1/2}$)TiO$_3$ ($x + y + z = 1$) (BNLKT100y-100z) three-component system increased from 185 °C to 221 °C in BNLKT4-100z at the rhombohedral composition. Although excess Li-substitution enhanced the piezoelectric properties, $T_d$ drastically decreased with increasing amount of Li-substitution. Considering both high $T_d$ and high $d_{33}$, the tetragonal composition of BNLKT4-100z is optimal for piezoelectric actuator applications. The rhombohedral composition of BNLKT4-8 seems to be suitable for high-power applications. In addition, a small amount of Mn-doping is very effective for improving $Q_m$. The high-power characteristics of BNLKT4-8Mn0.6 are superior to those of PZT-H at approximately $\nu_{p-p} > 0.6$ m/s. Therefore, a Mn-doped BNT-based solid solution with rhombohedral symmetry is a promising candidate for lead-free high-power applications.
The \((1-x)(\text{Bi}_{1/2}\text{K}_{1/2})\text{TiO}_3-x\text{BaTiO}_3\) [BKT–BT100] system was selected to elevate the \(T_C\) of BT to higher temperatures than 200 °C. The \(T_C\) of BKT–BT80 was about 230 °C. From these results, it is clear that the BKT–BT system is a superior candidate for lead-free piezoelectric materials for use in high-temperature applications. The piezoelectric properties of the textured and nontextured BKT–BT10, 20, and 30 are studied. The \(d_{33}\) in textured specimen of BKT–BT10 was improved as compared with that of nontextured specimen from 73.4 to 84.5 pC/N.

To replace PZT-based systems, it is necessary that special features of each lead-free material correspond to the required piezoelectric properties for each application such as high-temperature sensors, high-frequency applications, and temperature-stable ceramic resonators. Future trends in the research and development of environmentally gentle lead-free piezoelectric ceramics seem to be focused on textured grain orientations realized using the templated grain growth (TGG) and seeded polycrystal conversion (SPC) methods. In addition, domain controls including engineered domain of lead-free ferroelectric ceramics will be very important to enhance the piezoelectric activities. The trend toward thick and thin films with lead-free piezoelectric ceramics seems to be a necessity for FBAW and MEMS applications.

REFERENCES


[27] Nagata H, Takenaka T. Lead-free piezoelectric ceramics of \((\text{Bi}_{1/2}\text{Na}_{1/2})\text{TiO}_3\)-\(\text{K}^+\text{NbO}_3\)-1/2\((\text{Bi}_2\text{O}_3\cdot \text{Sc}_2\text{O}_3)\) system. Jpn J Appl Phys 1998;37:5311–4.


[29] Takenaka T, Nagata H. Ferroelectric and piezoelectric properties of lead-free \((\text{Bi}_{1/2}\text{Na}_{1/2})\text{TiO}_3\)-\(\text{K}^+\text{NbO}_3\)-1/2\((\text{Bi}_2\text{O}_3\cdot \text{Sc}_2\text{O}_3)\) ceramics. Proc. the 11th IEEE International Symposium on the Applications of Ferroelectrics (ISAF XI ’98) (IEEE Catalog No.98CH36245); 1999, pp. 559–562.


Heat Capacity Study of Functional Ceramics

Hitoshi Kawaji
Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Chapter Outline
1. Introduction 447
2. Thermodynamic Properties of Formation and Growth of Ferroelectric Nanoregion in Relaxors 448
3. Thermodynamic Studies of Giant Particle-Size Effect on the Phase Transition in Dielectric Crystals 451
4. Conclusion 453
References 453

1. INTRODUCTION

The stable state of materials is determined by the enthalpy of the materials at zero kelvin, which means that the most energetically favorable state is realized at that temperature. However, the stable state at a finite temperature cannot be determined by the enthalpy alone, but the Gibbs energy, which includes both enthalpy and the entropy term, should be considered to discuss the stability of the materials. This means that some disorder, which increases the entropy of the materials, will be excited with increasing the temperature. On the other hand, the large amount of disorder may be obtained abruptly at a temperature, i.e., a phase transition. The phase transition changes the physical properties drastically and thus influences the functionality of the material. For example, the superconductivity of a copper oxide superconductor disappears above the critical temperature and the finite resistivity appears above the temperature. The situation that the functionality of materials can be obtained below a phase transition is also seen in the ferromagnetism, ferroelectricity, and so on. There are some cases that the functionality of the materials appears above the phase transition temperature as the superionic conductivity of silver iodide. These indicate the importance of the investigation of phase transitions for understanding of the functionality of materials. The phase transition is governed by the Gibbs energy at constant pressure and thus the study of Gibbs energy is essential for the understanding the phase transition properties. The heat capacity measurements inform us the temperature dependence of the enthalpy and the entropy, thus the Gibbs energy, of materials and are important for the investigation of phase transitions. In addition, the entropy change due to the phase transition can be obtained only by the heat capacity measurements.

To obtain the data of enthalpy and entropy for the determination of stability of materials, the absolute value of the heat capacity should be measured. The precise date can be obtained only by the adiabatic calorimetry, which had been used to confirm the third law of thermodynamics. The principle of the measurements is very simple, in which the temperature increase by the Joule heating is precisely measured under the adiabatic condition. First, the sample is held under the adiabatic condition, which can be obtained by vacuum insulation and by an adiabatic shield which is kept at the same temperature with the sample; thus, the temperature of the sample is maintained at constant. Next, the known energy is added to the sample, keeping the adiabatic condition. After the addition of the energy, the sample reaches at a temperature. The heat capacity of the sample is obtained by dividing the Joule energy by the temperature increment. In general, the increment of the each heating is about 1–2 K below room temperature. Actually, the heat capacity of the empty sample vessel with the thermometer and the heater is measured before the sample measurement and subtracted from the total heat capacity including sample to obtain the heat capacity of the sample. The precision and the accuracy of the measurements are determined by the accuracy of the adiabatic
control and of the temperature scale of the thermometer. The precision is estimated about 0.1% of the value at room temperature, if about 1 g of the sample is used for the measurements. The precision and accuracy of the measurements decrease with decreasing the amount of the sample. Thus, in the case that the only small amount of the sample is available, a relaxation method is applicable. The measurements with the precision of about 1% are possible using the sample of about several tens of milligrams [1]. However, a special care is needed to determine the phase transition enthalpy and the entropy of a first-order phase transitions using the relaxation method. Above room temperature, DSC is also available to determine the heat capacity with the precision of about 1% using about several tens of milligrams of sample [2].

2. THERMODYNAMIC PROPERTIES OF FORMATION AND GROWTH OF FERROELECTRIC NANOREGION IN RELAXORS

In this section, the results of the heat-capacity study of the mechanism of relaxor behavior in perovskite oxides are discussed. The perovskite oxides are well known as the dielectric functional materials, especially lead complex oxides show superior electromechanical properties and are studied from the application points of view. In the group of the lead complex oxides Pb(B′,B′′)O₃, there are some compounds which show a large and frequency-dependent peak in the dielectric constant. These compounds are called relaxor due to the relaxation properties in the dielectric constant [3]. Some compounds of relaxor show no macroscopic change in the crystal structure, meaning that the cubic symmetry above room temperature is maintained till zero kelvin.

The lead complex oxides Pb(B′,B′′)O₃ are classified into several groups. In the group of compounds with the formula Pb(B⁴⁺₃,B⁵⁺₂)O₃, the typical relaxors, Pb(Mg₁/₃Nb₂/₃)O₃ (PMN) and Pb(Mg₁/₃Ta₂/₃)O₃ (PMT), are known, in which the transition metal ions are distributed almost randomly in B site [4]. In the case of Pb(B⁴⁺₁/₂,B⁵⁺₁/₂)O₃, the long-range ordering of B site can occur and the dielectric properties are influenced by the degree of the ordering [5]. The sample with highly ordered B site shows normal ferroelectric or antiferroelectrics property and the sample with less ordered shows relaxor behavior. The relaxor property has been investigated vigorously and some microscopic models have been proposed [6–12]. However, the precise mechanism is still an open question and the thermodynamic investigation has been required to clarify the mechanism of the relaxor properties.

PMN (Pb(Mg₁/₃Nb₂/₃)O₃) is the relaxor found very early and studied most extensively. The crystal structural analysis by X-ray diffraction and neutron diffraction indicates that lead ions occupy the off-centered position by 0.03 nm from the original perovskite A site, and oxide ions also occupy in the displaced position as shown in Figure 1 [13]. The averaged crystal structure belongs to the cubic space group Pm₃m in the whole temperature range and no structural phase transition is observed [14,15]. On the other hand, very small polarized region, which is called as polar nanoregion (PNR), is reported in the cubic matrix of the compounds [16,17]. The local symmetry of the PNR is of rhombohedral R3m and the polar axis of each PNR orientated randomly to eight <111> directions of the cubic matrix. Thus, the polarizations are canceled and thus no averaged polarization appears. The volume fraction of PNR increases with decreasing the temperature, and reaches about 20% at about 200 K. The growth of PNR freezes at about 200 K and the size of PNR is limited to about 10 nm [13,17]. The formation of PNR is reported to start below about 600 K in PMN by Burns et al. [18]. In addition to the formation of PNR, the short-range ordering of B-site cations is reported in PMN by TEM observation [19–25]. The superlattice reflection indicates the presence of locally 1:1 ordered region in the compound. The region with about 2 nm size is called chemically ordered domain (COD) and is distributed randomly in the disordered matrix keeping the cubic symmetry on the average [25,26]. The heat capacity study is carried out for the typical relaxor to detect the formation and growth of PNR and COD in relaxors and to understand the mechanism of the relaxor properties from the thermodynamic point of view [28]. Figure 2 shows the dielectric constant of PMN and PMT measured at various frequencies. The broad dielectric peaks are observed...
The peak temperature in the real part decreases with decreasing the measurement frequency and the temperature of the corresponding peak in the imaginary parts also decreases, indicating the relaxation properties. The relaxor properties in PMT are similar to PMN except is the ones observed below about 100 K. Figure 3 shows the heat capacity measured using an adiabatic calorimeter. No remarkable heat capacity anomaly is observed in the figure around the temperature where the dielectric peaks are observed. However, the heat capacity curves of PMN and PMT are crossing at 250 K: the heat capacity of PMN is smaller than that of PMT below 250 K and vice versa above 250 K. It is very curious and indicates the existence of excess heat capacity in PMN above 250 K. Figure 4 shows the Debye characteristic temperature calculated from the heat capacity data using the equation

\[ C_v = 3nR \left( \frac{T}{\Theta_D} \right)^3 \int_0^{\Theta_D/T} \frac{x^3e^x}{(e^x - 1)^2} \, dx. \]

The material which completely follows the Debye lattice vibration model shows constant Debye characteristic temperature in the whole temperature range. However, the Debye characteristic temperature depends on the temperature in the real materials, and the typical example is that of BMT (Ba(Mg\(_{1/3}\)Ta\(_{2/3}\))O\(_3\)) shown in Figure 4. At very low temperatures, the Debye temperature is determined by acoustic phonons and scarcely depends on the temperature. With increasing the temperature, the Debye characteristic temperature shows a minimum around at \(\Theta_D/20\) due to the effect of the zone boundary phonons and then approaches a certain value. However, the each curve of Debye characteristic temperature of PMN and PMT shows another minimum at 250 K and 320 K, respectively, as in Figure 4, which corresponds to the excess heat capacity in the materials. The excess heat capacity separated from the total heat capacity is shown in Figure 5. The excess heat capacity spreads over a wide temperature range from about 100 K to above 500 K in PMN. The onset temperature corresponds...
to the Burns’ temperature [18] below which PNR is considered to begin to form in PNM. The peaks are observed around the temperatures where the dielectric peaks appear in both materials. These results indicate that the excess heat capacity corresponds to the formation of PNR in the relaxors. The entropy calculated from the excess heat capacity is 3.3 J K\(^{-1}\) mol\(^{-1}\) for PMN and 2.9 J K\(^{-1}\) mol\(^{-1}\) for PMT. The large values show that the formation of PNR relates to order–disorder mechanism in the materials as expected from the disordered crystal structure. Assuming the order–disorder mechanism, the expected entropy is calculated to \(\Delta S = R\ln 8 = 17.3\) J K\(^{-1}\) mol\(^{-1}\) for the perfect ordering, which is 5 times larger than that of the measured value. The small entropy value obtained form the excess heat capacity corresponds the fact that the growth of PNR does not fully progress in the crystal.

PST (Pb(S\(_{1/2}\)Ta\(_{1/2}\))O\(_3\)) is another type of relaxor, in which the ordering of B site is possible and influences the dielectric properties. The degree of ordering can be controlled by quenching and annealing at elevated temperature [28–32]. The highly ordered sample shows ferroelectric property and the less ordered shows relaxor behavior. The heat capacity and the dielectric constant of the highly ordered sample, which is grown by a flux method around 1200 K, are shown in Figure 6. The heat capacity shows a sharp peak due to the ferroelectric phase transition at 310 K and the corresponding sharp dielectric peak is observed at the same temperature independently of the measuring frequency. Figure 7 shows the heat capacity and the dielectric constant of the less ordered sample which is quenched from 1830 K. The heat capacity peak moved to lower temperature and is smaller than those of the highly ordered sample. The
dielectric peak depends on the measuring frequency: the peak temperature decreases with decreasing the frequency, which is a characteristic feature of relaxor. Figure 8 shows the Debye characteristic temperature of both samples of PST. In the less-ordered sample, a broad anomaly like PMN is superimposed to the small dip due to the remaining phase transition, indicating that the broad heat capacity anomaly is essential for the relaxor behavior. It is very interesting that the broad heat capacity anomaly is observed above the phase transition temperature. The entropy calculated from the excess heat capacity is 4.1 J K$^{-1}$ mol$^{-1}$ and 3.3 J K$^{-1}$ mol$^{-1}$. The values are still very small compared to the theoretical value and indicate the incompleteness of ordering in the both samples [27,33]. The detailed analysis of the mechanism is still needed.

3. THERMODYNAMIC STUDIES OF GIANT PARTICLE-SIZE EFFECT ON THE PHASE TRANSITION IN DIELECTRIC CRYSTALS

It has been known that the crystal structure and the physical properties of materials are affected from its particles size. The typical example of ceramics is of ferroelectric properties of BaTiO$_3$. The large crystal of BaTiO$_3$ shows the ferroelectric phase transition around 400 K. However, the phase transition temperature decreases with decreasing the particle size and the phase transition disappears at around 10 nm. The other phase transitions in BaTiO$_3$ are also considered to behave similarly. Figure 9 shows the heat capacity of the BaTiO$_3$ powders heated at various temperatures [34]. The fine powders are used as starting materials and the averaged particle size increases with increasing the heated temperature. The observed heat capacity peaks correspond to orthorhombic to tetragonal phase transition in BaTiO$_3$ and grow with the heated temperature, thus the particle size. In general, these particle-size effects can be analyzed in terms of the influence of the surface of the fine particles. The relative population of the atom near surface increases drastically with decreasing the particle size especially in nanometer scale, and the surface effect becomes considerable. On the other hand, a peculiar particle-size effect on the phase transition is found in some dielectric substances, in which the boundary of phase transition behavior locates around 1 mm not at nanometers or micrometers.

BZG (BaZnGeO$_4$) is a ferroelectrics and shows an incommensurate phase transition. Figure 10 shows the heat capacity of BZG crystals [35]. Two heat capacity peaks are clearly seen in the crystal with large size: the higher-temperature peak corresponds to the known phase transition from phase V–IV and the lower phase transition is newly found (the lowest temperature phase is phase VI). However, the sample with less than 0.1 mm shows only one heat capacity peak and the phase VI does not appear. The crystal structure change due to the newly found phase transition is confirmed by the X-ray diffraction [36,37]. The surface effect should be ignored in millimeter size and another mechanism should be considered.

More recently, CZP (CsZnPO$_4$) is found to be another material showing similar giant particle size effect on
phase transition [38–40]. Figure 11 shows the heat capacity of CZP with different particle size. The sample with large size above 0.1 mm shows a heat capacity anomaly due to the phase transition around 220 K, but no heat capacity anomaly is observed in the samples below 0.1 mm. This behavior is confirmed by X-ray diffraction as shown in Figure 12. The single crystal of CZP shows a jump in the lattice parameter at 220 K in the heating direction and at 200 K in the cooling direction. The jump of the lattice parameter and the thermal hysteresis clearly indicates the first-order transition nature. On the other hand, the powdered sample shows no jump in the lattice parameter and the temperature dependence indicates that the high temperature phase is supercooled until the lowest temperature. In addition to the giant particle-size effect on the phase transition, the phase transition shows a relaxation behavior. Figure 13 shows the heat capacity anomaly of the samples annealed below the phase transition temperature for various time intervals. The heat capacity anomaly grows with increasing the annealing time. The kinetics of the growth of the heat capacity anomaly can be analyzed using the Avrami model and the results are shown in Figure 14, indicating a heterogeneous nucleation and growth. The mechanism of this curious giant particle-size effect and the relaxation properties of the phase transition have not yet been understood clearly. However, the high-temperature phase is ferroelastic and the ferroelastic domain may influence the phase transition property. The detailed studies are required.
4. CONCLUSION

Heat-capacity studies of functional materials have the importance that not only the temperature dependence of thermodynamic functions, which govern the satiability of the materials, can be determined but also the entropy which is the measure of disorder in the materials and is directly determined only by the precise heat capacity measurements. The heat-capacity studies of the formation and growth of ferroelectric nanoregion in relaxors indicate that the entropy change due to the formation of the nanoregion is limited only one-fifth of the perfect ordering, and the disorder remains and freezes at low temperature. Such information can be obtained only by the low-temperature heat-capacity measurements, and the usefulness of the thermodynamic is clearly shown in the first topic. In addition, all the possible types of thermal agitation contribute to the heat capacity of materials; thus, the heat-capacity measurements have an advantage to find novel phenomena occurring with the temperature change as shown in second topic. The thermodynamic study of materials is classic but still effective in the investigations of the functional materials.

REFERENCES


[29] Stenger CGF, Scholten FL, Burggraaf AJ. Ordering and diffuse phase transitions in Pb(Sr0.5Ti0.5)O3 ceramics. Solid State Commun 1979;32:989.


Chapter 6.4
New Frontiers Opened Up Through Function Cultivation in Transparent Oxides

Hideo Hosono
Frontier Collaborative Research Center and Materials and Structures Laboratory, Tokyo Institute of Technology, Mail Box R3-1, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Chapter Outline
1. Introduction 455
2. Research Concept and Strategy 456
3. Light metal TCO: 12CaO·7Al2O3 with Built-in Nano-Porous Structure 457
4. Transparent Amorphous Oxide Semiconductors 468
5. Iron-Pnictide Superconductors 477
6. Future Challenge: Ubiquitous Element Strategy 484
Acknowledgments 485
References 485

1. INTRODUCTION

Top 10 of the Clerk number showing the order in natural abundance of elements in the earth crust are listed as oxygen, silicon, aluminum, iron, calcium, sodium, potassium, magnesium, hydrogen, and titanium. This means that our territorial environment consists of oxides of these abundant metals and water. Human beings have developed civilization utilizing these oxides to date, i.e., Al2O3, SiO2, CaO, and MgO for the Stone Age, Fe for the Iron Age, and the present Information Age is supported by Si semiconductor materials and SiO2 glass fibers. It is no doubt that the coming age will be created through utilizing these abundant materials intelligently.

Oxide ceramics is probably among the oldest of man-made materials owing to the abundance and easy availability of the ingredients. Although most oxides are optically transparent, which is important for optical applications, it has been believed that active functions based on “excited electrons”, such as in crystalline semiconductor materials, are not possible. For example, alumina and glasses, which are representative oxides, are optically transparent but electrically insulating.

However, during the past two decades techniques to purify oxides have drastically advanced. It is now possible to obtain highly pure materials, resulting from concentrated studied on fine ceramics in the 1980s. For instance, metal oxide chemicals containing sub-ppm impurities are now commercially available. Furthermore, the thin film deposition technique of oxides was much advanced through intensive study on high Tc superconductors for electronic applications in the past decade. Taking advantage of this situation, our group has explored transparent conducting oxides (TCOs) while following a working hypothesis established based on a consideration of chemical bonding and point defects, resulting in more than ten new TCOs. Important among them are p-type TCO, CuAlO2, reported in Nature (1997) [1] and a series of amorphous TCOs in which the Fermi level is controllable by intentional doping (1996) [2]. The former is particularly important because most active functions in semiconductors originate from PN-junctions [3]. The absence of practical application of transparent oxide semiconductors is due primarily to the missing of P-type TCO. It is therefore now possible to examine the possibility of invisible circuits based on transparent oxides as shown in Fig. 1. As for optical materials, Hosono et al. reported novel photosensitive or photorefractive glasses [4], SiO2 glasses [5], and a CaF2 crystal [6] for excimer laser applications. These materials were found through accumulated fundamental research on defects in glass and dielectric crystals. This study started from October 1, 1999 under such a situation.
2. RESEARCH CONCEPT AND STRATEGY

A research concept “Transparent Electro-Active Materials (TEAM)” was born in a dream “we wish to convert alumina to transparent semiconductor”. Transparent oxides are most abundant and stable on earth, and environmentally friendly. Although they have been used as ingredients of traditional materials such as porcelain, cement, and glass, few active functions have been found in them. In fact, these materials are described as typical insulators in college textbooks. A widely accepted view “a transparent oxide cannot be a platform for electro-active materials” comes from phenomenological observation. I think it possible to realize a variety of active functionalities for transparent oxides by appropriate approach based on an original view for these materials.

What is unique for transparent oxides? I think there are 2 characteristic features: a wide variety of crystal structures and that constituting elements are characteristic of oxides. No such vast variety is seen in typical semiconductor or compound semiconductors. The crystal structure of the latter is limited to the diamond-type. This striking difference arises from that in the chemical bonding nature between oxides and compound semiconductors, i.e., although the ionic nature controlling a long range ordering is primary in oxides, the covalent nature determining the local coordination structure contributes to the bonding significantly, depending on the type of a metal cation. On the other hand, bonds in compound semiconductors are almost covalent and, as a consequence, local coordination configuration is tetrahedral. This feature results in the striking difference in nature between hole and electron pathways. The conduction band minimum (CBM) that works as an electron pathway is primarily composed of metal vacant orbitals, whereas 2p-orbitals of electro-negative oxygen ions dominantly contribute to the valence band maximum (VBM) which functions as a hole pathway. This electronic nature is the basis for materials designing in transparent oxide semiconductors. Figure 2 summarizes the chemical bonding in oxides in comparison with a pure covalent compound and a pure ionic compound. Here we focus on transparent oxides with low-dimensional crystal structure, i.e., nano-porous structure and lowered structure. If an electron could be injected to the nano-sized cage, the resulting cages may be regarded as coupled quantum dot system for which one may expect a novel electronic structure. For layered structures composed of a narrow semi-conducting layer sandwiched by wide gap oxides, we may anticipate a unique electronic state of a 2D-electronic nature.

Another is a variety of excited states. Although transparent oxides composed of abundant elements appear less interesting in the ground state, there are so many interesting excited states. Representative examples are a self-trapped exciton that works as energy localization center, leading to persistent defect formation and a polaron composed of a charge carrier and its associated polarization and host lattice deformation field. This feature provides valuable opportunities to modify transparent dielectric crystalline and amorphous oxides toward novel functions and devices by electronic excitation. Several representative articles on this subject are described in the literature [7–11]

FIGURE 1  Impact of discovery of a p-type transparent conductive oxide. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 2  Characteristics of transparent oxides in comparison with covalent semiconductor and ionic crystal. For color version of this figure, the reader is referred to the online version of this book.
3. LIGHT METAL TCO: 12CaO·7Al₂O₃ WITH BUILT-IN NANO-POROUS STRUCTURE

A wide variety of materials and crystal structure is a unique feature of oxides. Electro-conductive oxides are, however, restricted to transition metal or heavy metal cation-bearing materials. Representative light metal oxides that are most abundant in our territorial crust are classified as typical transparent insulators as described in textbooks. They are the main constituents of traditional ceramics such as porcelain, cement, and glass, and no one has succeeded in converting them to persistent electronic conductors until a report [12] in 2002 utilizing the nano-structures built-in crystal structure of 12CaO·7Al₂O₃ (C12A7). This section describes the electronic structure of C12A7 and various approaches to electron doping along with unique properties of electron-doped C12A7.

3.1. Crystal Structure of 12CaO·7Al₂O₃

This compound can be easily prepared by a conventional solid-state reaction in air, but is not formed in O₂ or moisture-free atmosphere [13]. The crystal lattice of 12CaO·7Al₂O₃ (C12A7) with a microporous structure has a cubic lattice constant of 1.199 nm, and the unit cell includes two molecules [14]. It is characterized by a positively charged lattice framework [Ca₂₄Al₂₈O₆₄]⁴⁺ having 12 crystallographic cages with a per unit cell as shown in Figure 3. Each cage with an inner diameter of ~0.44 nm and a formal charge of +1/3 is coordinated by 8 cages to form a 3-dimensional structure by sharing a monomolecular open mouse. This structure is similar to a zeolite but the excess charge of the cage wall is reverse. Free oxygen ion O₂⁻ accommodates in 1/6 out of 12 cages in the unit cell to reserve electroneutrality in the stoichiometric state. The remaining two oxide ions (O₂⁻) are entrapped as counter anions in the cages in the stoichiometric composition and called free oxygen ions because the cage diameter is larger by ~50% than the diameter of O²⁻. This fact is associated with the fast oxygen ion conduction in this material [15]. Monovalent anions, such as hydroxyl ions (OH⁻) or halogen anions, X⁻(X⁻ = F⁻ and Cl⁻), are known to substitute for a part or all of the free oxygen ions to form derivatives [Ca₂₄Al₂₈O₆₄]⁺⁺⁻·4(X⁻). The replacement of a free O₂⁻ ion by two monovalent anions likely stabilizes the structure through charge delocalization [16]. C12A7 crystals can be easily synthesized by conventional solid-state reactions using CaCO₃ and Al(OH)₃ or Al₂O₃. Single crystals can be grown from the melt by using a floating-zone (FZ) method [17] or the Czochralski (CZ) technique [18].

3.2. Electronic Structure and Our Approach

Figure 4 shows the band structure of stoichiometric C12A7 calculated by density functional method with GGA approximation using VASP code. Two major features are evident from the figure [19]. (1) The O₂p levels of the free oxygen ions are located ~1 eV above the valence bands. (2) There are two types of conduction bands, i.e., one is a conventional conduction band that is primarily composed of Ca 4s levels, and is called the framework conduction band (FCB). Another is a conduction band composed of 3-dimensionally connected sub-nanometer-sized empty cages, which is not comprised of orbitals of specific ions constituting the cage wall but is constituted of a coupled “particle-in-a box”, and is called the ‘cage conduction band (CCB)” [19–21]. CCB, which is unique to C12A7, is located ~2 eV below from the bottom of FCB and has moderate energy dispersion [22]. These features led us to an idea that if the free oxygen ions are replaced by electrons, the resulting state may exhibit high electronic conduction; the energy level of the FCB bottom is too high for electron doping, whereas the location of CCB is marginal for successful electron doping compared in the doping limit [23,24] for other n-type semiconductors. Various methods to inject electrons to the CCB are described in the next section.

3.3. Electron Doping

3.3.1. Light-Induced Doping for C12A7:H⁻

Hayashi et al. [12] demonstrated a process by which the transparent insulating oxide C12A7 can be converted into an electronic conductor. H⁻ ions are incorporated into the sub-nanometer-sized cages of the oxides by a thermal treatment in a hydrogen atmosphere at 1300 °C; subsequent illumination of the resulting C12A7:H⁻ with ultraviolet light (λ < 350 nm) results in a conductive state that persists after irradiation ceases. The photo-activated material exhibits moderate electrical conductivity (~0.5 S cm⁻¹) at room temperature, with visible light absorption losses of...
only one percent for 200-nm thick films, as shown in Figure 5. The incorporation of \( \text{H}^– \) anions in the cage may be understood through the following chemical reactions:

\[
\text{O}_2^–(\text{cage}) + \text{H}_2(\text{atmosphere}) \rightarrow \text{OH}^–(\text{cage}) + \text{H}^–(\text{cage}),
\]

(1)

or

\[
\text{O}_2^–(\text{cage}) + \text{H}_2(\text{atmosphere}) \rightarrow 1/2\text{O}_2(\text{atmosphere}) + 2\text{H}^–(\text{cage}).
\]

(2)

Specimens C12A7: \( \text{H}^– \) look white as polycrystalline powder or colorless and transparent for single crystals and is insulating (conductivity \( \lesssim 10^{-10} \) S cm\(^{-1} \) at RT). However, upon illumination with ultraviolet light (wavelength \( < 330 \) nm), two absorption bands centered at 2.8 eV and 0.4 eV are induced, and simultaneously, a drastic increase in the conductivity was observed along with the coloration. The conductivity increased from \( \lesssim 10^{-9} \) S cm\(^{-1} \) to 0.5 S cm\(^{-1} \). The Seebeck coefficient of the conductive sample is negative \((-360 \) mV K\(^{-1} \)), indicating n-type conduction. Figure 6 shows the temperature dependence of the conductivity \( (\sigma) \). Log \( \sigma \) is proportional not to \( T^{-1} \) but to \( T^{-1/4} \) over a wide temperature range 300–50 K. This characteristic indicates that the conduction is not controlled by a simple thermal activation process but more probably by a variable-range hopping of carriers.

A possible mechanism for the photo-induced conversion of C12A7: \( \text{H}^– \) from an insulator to a conductor is proposed, as summarized in Figure 7. First, UV-illumination induces an electron emission from the encaged \( \text{H}^– \) ions (\( \text{H}^– \rightarrow \text{H}^0 + \text{e}^– \)). Then, an empty cage electro-statistically...
captures the photo-induced electron to form an F$^+$-like center, i.e., an electron trapped at the site of an oxygen vacancy. As the electron is weakly bound in the cage, the wave function would spread spatially. Consequently, electrons may migrate throughout the crystal by variable-range hopping. The resultant H$^+$ atoms rapidly diffuse throughout the cages and combine with the remaining free O$^{2-}$ ion to form OH$^-$ releasing another electron. That is to say, a H$^+$ ion works as a double photo-electron donor in C12A7. The validity of this model was justified by observation of H$^0$–F$^+$ pair generation in the sample illuminated at 4 K [25] and theoretical calculations [22]. The electron doping from H$^-$ is possible as well by an electron-beam irradiation in place of UV-radiation [26]. The quantum efficiency of conduction electron generation evaluated for 25 keV electron-beam is 0.02. Since electron-beam can be focused to a level beyond the limit by UV-light, direct writing of electro-conductive patterning down to several ten nanometers is possible.

### 3.3.2. Doping by Chemical Reduction

Several chemical processes were developed for electron doping via removal of the free O$^{2-}$ ion. Thermal annealing of C12A7 single crystals in Ca metal vapor at 700 °C led to the electrical conductivity of ~80 S cm$^{-1}$ at 300 K [27]. However, Ca treatment needs a long duration (e.g. 10 days for a 0.4-mm thick sample) to extract the free oxygen ions because the upper annealing temperature is limited to ~700 °C due to chemical reactions of Ca vapor with SiO$_2$ glass tubes. Further, the replacement may be incomplete due to the formation of a surface CaO layer that acts as a diffusion barrier of the free O$^{2-}$. These drawbacks were overcome by employing metal Ti [28] in place of metal Ca because metal Ti forms stable non-stoichiometric oxides with high electronic conductivity over a wide chemical composition range ($1 < x < 2$ in TiO$_{2-x}$) and the out-diffusion of the free oxygen ions from C12A7 is not blocked even if a thick titanium oxide layer completely covers the sample surface. In addition, higher annealing temperatures up to ~1,300 °C can be employed utilizing an observation that Ti does not react with C12A7 up to this temperature. Since an increase in the temperature from 700 to 1100 °C enhances the diffusion of the free oxygen ions in C12A7 by several orders of magnitude, the treatment time can be drastically shortened.

Figure 8 shows the temperature dependence of the electrical conductivities ($\sigma$) for the Ti-treated single crystals with different electron concentrations ($N_e$). The $N_e$ values were evaluated from the optical reflection spectra of the single crystals. The Arrhenius plot has a negative slope when the electron concentration is low, indicating that a thermally activated mechanism such as hopping controls the migration of the electrons. On the other hand, the temperature dependence vanishes at $N_e = 1 \times 10^{21}$ cm$^{-3}$ ($N_{th}$), demonstrating that electron conduction is degenerated, and finally the slope becomes positive around room temperature at larger $N_e$ values. Conductivity extrapolated to 0 K ($\sigma_{0K}$) as a function of $N_e$ (Figure 9) shows this systematic change more clearly: $\sigma_{0K}$ starts to increase from zero to a finite value at $N_e = N_{th}$, confirming an insulator–metal (IM) transition and $\sigma_{0K}$ super-linearly increases with $N_e$, showing a sharp increase in the drift mobility (from ~0.1 to 4 cm$^2$ V$^{-1}$ s$^{-1}$) for $N_e > N_{th}$ (inset of Figure 9).

These observations demonstrate that insulating C12A7 is converted to metallic state by replacing more than half of the free oxygen ions by electrons. This IM transition accompanies a sharp increase in the drift mobility ($\mu_e$) from ~0.1 to 4 cm$^2$ V$^{-1}$ s$^{-1}$ as $N_e$ increases. It should be noted that the doping and conduction mechanisms in C12A7 are much
different from those in conventional semiconductors due to the unique crystal and electronic structures of C12A7.

In typical semiconductors such as phosphorus-doped Si [29], IM transition occurs due to the percolative overlap of the donor wave functions as explained by Mott in terms of the critical donor concentration $N_c$, which follows the empirical relation $N_c^{1/3}R_c \sim 0.25$ ($R_c$ denotes the radius of the electron orbital) [30]. When the $N_{th}$ value of $\sim 1 \times 10^{21}$ cm$^{-3}$ is put into this relation, the obtained $R_c$ value is of $\sim 0.26$ nm, which is close to the radius of the crystallographic cage embedded in C12A7 ($\sim 0.25$ nm). Thus, it is feasible that the IM transition in C12A7 occurs via the percolation of the ‘donor’ orbitals. However, this view cannot be directly applicable to C12A7 because even an ‘undoped’ stoichiometric C12A7 crystal has a ‘donor wave function’ in each empty cage as a component of the CCB. Although the ‘donor wave function’ intrinsically exists, the electrons doped to the CCB do not contribute to the band conduction at $N_e < N_{th}$ at low temperatures because the electron is localized in a cage forming an $F^+$-like center [12,25] due to the lattice relaxation, as described above. However, a high electron doping may average out the deformations of the cage geometry throughout the crystal and it converts the localized $F^+$-like state to a delocalized state, invoking an IM transition.

This model for the IM transition is validated by MEM/Rietveld analysis and $ab$-initio density functional theory (DFT) calculations [28]. Figure 10 displays the electron density maps in the (001) plane of stoichiometric C12A7.

**FIGURE 9** Conductivity extrapolated to 0 K and mobility as a function of carrier electron concentration. When each $O^{2-}$ ion in the cages is replaced by 2 electrons, $N_e$ corresponds to $2.3 \times 10^{21}$ cm$^{-3}$. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 10** Electron density map of cages in insulating and metallic conducting C12A7 determined by MEM-Rietvelt analysis of X-ray diffraction pattern. This density is on the (001) plane. Note that two Ca$^{2+}$ ions at the axial position are split into 2 sites corresponding to an $O^{2-}$ entrapped cage and untrapped cage. For color version of this figure, the reader is referred to the online version of this book.
and metallic C12A7 obtained by the MEM/Rietveld analysis [31]. The electron densities of the Ca\(^{2+}\) ions on the cage wall of stoichiometric C12A7 have long tails like a dumbbell shape, indicating that the map is a superposition of the two deformed O\(^{2−}\)-encaging cages and the ten undeformed empty cages. The distances between the Ca\(^{2+}\) ions (D\(_{\text{Ca−Ca}}\)) in the deformed and empty cages obtained by the Rietveld analysis are 4.22 Å and 5.65 Å, respectively. Thus, the map in Figure 10 provides solid evidence that the O\(^{2−}\)-encaging cage shrinks largely from the empty cage. On the other hand, the map in Figure 12c indicates that the cage structure of [Ca\(_{24}\)Al\(_{28}\)O\(_{64}\)]\(^{4+}(4e^-)\) clearly differs from that of stoichiometric C12A7. Each atom has a spherical electron density, indicating that the local structural deformation is removed in [Ca\(_{24}\)Al\(_{28}\)O\(_{64}\)]\(^{4+}(4e^-)\). That is, all the twelve cages are homogenized and D\(_{\text{Ca−Ca}}\) (5.64 Å) becomes close to that of the empty cage, suggesting the electrons are distributed uniformly over the twelve cages within the accuracy of the MEM/Rietveld analysis. Figure 11 shows relaxed lattice structures calculated by DFT with the isosurfaces of the wave function (|\(\Psi|\(^2\)) of 2p level of the free oxygen ions and those of the CCB bottom (isosurface state density) for (i) stoichiometric C12A7, (ii) [Ca\(_{24}\)Al\(_{28}\)O\(_{64}\)]\(^{4+}(O^{2−})(2e^-)\) and (iii) metallic C12A7. The DFT results are summarized in

**FIGURE 11** Isosurface state density by DFT calculation on relaxed structure and schematic band structure of C12A7. For color version. For color version of this figure, the reader is referred to the online version of this book.
Figure 13 as a schematic energy level diagram. Figure 11a(i) demonstrates that two free oxygen ions respectively occupy the cages at the center and the corners of the unit cell and the other cages are empty in the stoichiometric C12A7. The calculated $D_{Ca-Ca}$ of the $O_2^-/C_0$-encaging cage is 4.27 Å, in good agreement with the observed value of 4.22 Å. Further, each empty cage has an s-like state, and the connection of these states forms the CCB even in the stoichiometric C12A7. On the other hand, the $O_2^-/C_0$-encaging cage has no such CCB state because of the entrapping of $O_2^-/C_0$, whose 2p energy level is located slightly above the valence band maximum of the cage framework (Figure 11c(i)). Thus, stoichiometric C12A7 has no electron in the CCB and is a band insulator [22].

The DFT result for $[Ca_{24}Al_{28}O_{64}]^{4+}(O_2^-)/(2e^-)$, where a free oxygen ion is extracted from a cage and two electrons occupy two of the other cages in the unit cell, demonstrates that $D_{Ca-Ca}$ (5.57 Å) of the electron-encaging cage becomes larger compared to that of the $O_2^-/C_0$-encaging cage (4.27 Å), but it is still smaller than that of the empty cage (5.77 Å). The wave function at the CCB minimum (Figure 11b(ii)) implies that the electron is still localized in a specific cage, forming an $F^+$-like center (Figure 11c(ii)).

Figure 11b(iii) is the relaxed lattice structure with the isosurface state density at the CCB minimum of $[Ca_{24}Al_{28}O_{64}]^{4+}(4e^-)$, revealing that the $D_{Ca-Ca}$ (5.63 Å) of the electron-encaging cage further expands from that of $[Ca_{24}Al_{28}O_{64}]^{4+}(O_2^-)/(2e^-)$ and it becomes very close to...
that of the empty cage. It also indicates the isosurface state density in each cage is equivalent, indicating that the electrons are delocalized over the cages and consequently the cage conduction band has a rather large dispersion of ~2 eV.

Thus, the DFT calculation, consistent with the MEM/Rietveld analysis, qualitatively explains why C12A7 shows IM transition upon electron doping. That is, the successive reduction of deformation of the electron-encaging cage with increasing the electron concentration induces the IM transition, which accompanies the sharp increase in the electron mobility. It is noteworthy that simple ionic oxides such as MgO have not exhibited such a high conductive state although the F/F⁺-like centers can be incorporated at high concentrations similar to the present case [32]. Thus, the three-dimensionally connected sub-nanometer-sized cage structure in C12A7 plays an essential role in the appearance of the metallic state.

3.3.3. Doping via Melt Processes

The development of a simple synthesis process, which is compatible for mass production of C12A7:e⁻ is strongly required for practical applications. Possible candidate to satisfy the requirement is a direct synthesis of C12A7:e⁻ from the melt or glass state. Extensive studies regarding the relationship between the melting condition and corresponding precipitated phases in the CaO-Al₂O₃ system [33] verify that the presence of template anions to compensate the positive charge of the C12A7 framework is inevitable for the precipitation of the C12A7 phase. More specifically, the C12A7 phase is reproducibly prepared from the stoichiometric C12A7 melt in wet air [13], involving O²⁻ and/or OH⁻ ions that act as the template. On the contrary, the thermal treatment in inert gas atmospheres, where template anions are seemingly absent, hinders the formation of the C12A7 phase and only a mixture phase of C₃A and CA (CaO·Al₂O₃) is formed. If we could precipitate the crystalline C12A7 under an appropriate reducing atmosphere, there is a possibility of direct synthesis of C12A7:e⁻ from the melt or glass. The critical issue for the direct synthesis is to find a novel template ion that can exist in the melt or glassy phase even when they are exposed to reducing atmospheres. With efforts in searching for such templates, successful results were obtained for the process via the reduced melt [34] or glass state [34], where the C₂²⁻ ion most likely acts as a template for the C12A7 phase formation [35].

Figure 12(a) shows schematically a direct solidification process to prepare C12A7:e⁻ from the C12A7 melt. C12A7 powders of the stoichiometric mixture of CaO and Al₂O₃ are melted in a semi-airtight carbon crucible with a carbon cap at 1600 °C for 1 hr in air. Heating the carbon crucible at 1600 °C produces a strongly reducing atmosphere (PO₂ = ~10⁻¹⁶ atm) inside the crucible. The crucible is slowly cooled down at a rate of ~400 °C/hr. X-ray diffraction measurements clarify that solidified products are mixtures of C₃A and CA. Then, they are re-melted and solidified again according to the same temperature profile as that of the first run. The second run results in the formation of the C12A7 crystalline phase with green color and electrical conduction.

Raman spectra of the products after the first run and thermal desorption measurement in the cooling process in the second run suggest that C₂²⁻ ions are dissolved into the melt from the carbon crucible to compensate for the oxygen deficiency caused by the strongly reducing atmosphere. Furthermore, the ionic radius of C₂²⁻ (1.2 Å) is similar to O²⁻ (1.4 Å), which supports that the C₂²⁻ anions serve as the template in a reducing atmosphere, instead of the free oxygen ions in an oxidizing atmosphere. Moreover, the C₂²⁻ Raman band is not observed in the C12A7:e⁻ and CO gas is evolved in the cooling process, which implies that the encaged C₂²⁻ ions are only stable in the nucleation and/or initial stage of the crystallization and they are released from the cages, leaving electrons in the lattice during the cooling process. Thus, the electron formation processes in the cages may be expressed as follows:

C₂²⁻(melt) → C₂²⁻(cage) → 2C(solid) + 2e⁻(cage), (3)
and/or

C₂²⁻(cage) + 2O²⁻(cage) → 2CO(atmosphere) + 6e⁻(cage) (4)

Figure 12b shows schematically the glass-ceramics process for the preparation of C12A7:e⁻. Oxygen-deficient glass, which is obtained by a rapid quench of the C12A7 melt in a carbon crucible placed in air, is crystallized in an evacuated silica tube by heating to 1000 °C, which is higher than the crystallization temperature (>900 °C). The resultant C12A7 samples were green in color and exhibited rather high electrical conductivity, indicating that electrons are incorporated into the cages instead of free oxygen ions. Similar to the direct solidification process, C₂²⁻ ions, which are likely incorporated in the glass, are considered to act as the template. The resulting C12A7:e⁻ having 10¹⁹–10²⁰ cm⁻³ exhibits 0.1–1S/cm at RT.

3.4. Thin Film Fabrication of Electron-Doped C12A7

Thin film fabrication is needed for device application of C12A7:e⁻. Bulk formation processes described above are not applicable for thin film fabrication because the thickness of reaction layers formed by metal treatment is comparable to the conventional film thickness. Thus, different processes
are required to be invented. Two processes to be described below have been reported for this purpose.

### 3.4.1. Ion Implantation

Ion implantation, a nonequilibrium physical process, has such advantages over conventional chemical doping processes that high concentration of ions can be introduced in a designated local area with an excellent controllability [29,30], and thus it is widely used for the modification of thin films.

Miyakawa et al. [36,37] conducted Ar⁺ ion implantation into polycrystalline C12A7 thin films and found that the electron anions are incorporated into the cages as a result of kicking out the free oxygen ions from the cages through the nuclear collisions. The polycrystalline C12A7 thin films (~800 nm in thickness) were prepared on MgO(100) single-crystal substrates by a pulsed laser deposition technique at RT and a subsequent annealing at 1100 °C for 1 hr in air to crystallize the deposited amorphous films. Ion implantation was carried out at temperatures from RT to 600 °C in a vacuum less than ~2 × 10⁻⁸ torr. The Ar⁺ fluence (accumulated dose) set to vary from 1 × 10¹⁶ to 1 × 10¹⁸ cm⁻². Moreover, the C12A7 framework is stable for hot implantation, although a crystalline film is converted to amorphous phase if the implantation is performed at RT.

**Figure 13** shows the electrical conductivities of 300 keV Ar⁺-implanted films subjected to the implantation at high temperature of 600 °C as a function of Ar⁺-ion fluence. The electrical conductivity of the film increases with implanted fluence: films with fluences less than 1 × 10¹⁷ cm⁻² are insulators, while those with 5 × 10¹⁷ to 1 × 10¹⁸ cm⁻² exhibit electrical conductivity of ~0.5 and 1 S cm⁻¹, respectively. The enhancement of the electrical conductivity is accompanied by the appearance of optical absorption bands at 0.4 and 2.8 eV, indicating the F⁺-like centers are simultaneously formed by the hot implantation process. These results suggest that the free oxygen ions are kicked out from the cages by the collision between the free oxygen ions and Ar⁺ ions, leaving electrons in the cages. As a result, the electrons are directly produced by Ar⁺-implantation via the extrusion of free oxygen ions through the following reaction:

\[
\text{O}^2⁻(\text{cage}) \rightarrow \frac{1}{2}\text{O}_2(\text{atmosphere}) + 2e⁻(\text{cage})
\]  

(5)

It is noted in samples implanted with less than 1 × 10¹⁷ cm⁻² Ar⁺ ions that the conductivity and the optical absorption intensities increase after UV-light irradiation. Since these phenomena are commonly observed in H⁻ incorporated C12A7, it suggests that H⁻ ions are created into the cages by the hot implantation of Ar⁺ ions. It was considered that the formation of the H⁻ ions by Ar⁺-implantation is caused by preexisting OH⁻ groups that are introduced during the post-annealing process for the crystallization of amorphous films in air. That is, electrons are generated by the extrusion of free oxygen ions through reaction (5) and simultaneously protons (H⁺) are created by the decomposition of preexisting OH⁻ groups in the as-prepared film through reaction (6):

\[
\text{OH}⁻(\text{cage}) \rightarrow \text{O}^2⁻(\text{cage}) + \text{H}⁺(\text{attached to cage wall})
\]  

(6)

Then, H⁺ ions immediately react with the electrons in the cages, leading to the formation of H⁻ ions as follows:

\[
2e⁻(\text{cage}) + \text{H}⁺(\text{attached to cage wall}) \rightarrow \text{H}⁻(\text{cage})
\]  

(7)

Thus, H⁻ ions are formed in lower Ar⁺-implanted samples, leading to light-induced electronic conduction and coloration.

**Figure 14** shows the schematic illustration of the electron and H⁻ formations by Ar⁺-implantation as mentioned above. Ar⁺-implantation produces the electrons and protons in the cages by nuclear collisions and electronic excitation effects through the reactions (5) and (6). When the number of the electrons produced by reaction (5) is less than that of protons, all produced electrons are captured by H⁺ ions through reaction (7) and H⁻ ions are formed in the cages. However, when the generated electron concentration...
exceeds the proton concentration, the excess electrons themselves are directly introduced into the cages.

Since higher concentration of the electrons is incorporated three-dimensionally in a designated narrow space, the synthesis of the C12A7:e⁻ thin films by ion implantation paves a way to realize novel applications such as patterned electric wiring, electron injection electrode, and cold electron emitter.

3.4.2. Chemical Reduction

Miyakawa et al. [38] found a method for fabrication of C12A7:e⁻ thin films by using a deposition of oxygen-deficient amorphous C12A7 layer on crystalline C12A7 thin films. Figure 15 shows the processes. First, amorphous thin films of ~400 nm thick were deposited at RT by the pulsed laser deposition (PLD) method using ceramic C12A7 targets. The partial O₂ pressure was ~1 × 10⁻³ Pa during the depositions, and the substrates employed were MgO(100). Polycrystalline(p-) C12A7 thin films were obtained by heating the resulting amorphous (a-) C12A7 thin films to 1100 °C in an ambient atmosphere. Next, a-C12A7 (~30 nm thick) layers were deposited on the p-C12A7 thin films at 700 °C in an O₂ pressure of ~1 × 10⁻³ Pa by PLD. Last, the over-layered a-C12A7 layer was removed by chemical—mechanical polishing (CMP). By tuning the duration and temperature of post-annealing temperature, the electron concentration in C12A7 thin film can be increased up to ~2 × 10²¹ cm⁻³ (the theoretical maximum).

12SrO·7Al₂O₃ (S12A7) with the same crystal structure as C12A7 is known as a metastable phase [39]. Miyakawa et al. [40] applied this method to S12A7 and successfully obtained S12A7:e⁻ thin films by tuning the temperature and atmosphere for the post-annealing of a-S12A7. Figure 16 shows the DC conductivities of resulting C12A7:e⁻ and S12A7:e⁻ thin films as a function of temperature. It is evident that both samples exhibit metallic conduction. The Hall mobility and carrier concentrations in the C12A7:e⁻ film were evaluated to be 2.5 cm²(Vs⁻¹) and 1.3 × 10²¹ cm⁻³, respectively. Similar values (1.0 cm²(Vs⁻¹) and 1.4 × 10²¹ cm⁻³) were obtained for the S12A7:e⁻ thin films. No carrier doping was observed in the same conditions without...
deposition of a-C12A7 on pc-C12A7 or S12A7. Deposition of Au layers or Ti layers was examined for carrier generation. The former did not work, whereas the latter did. These observations imply that the role of a-C12A7 deposition on p-C12A7(S12A7) is to exact free O\textsuperscript{2−} accommodated in the cages and inject electrons into the cages.

### 3.5. Unique Properties of Electron-Doped C12A7 and Applications

C12A7:e\textsuperscript{−} is an exceptional material that exhibits metal–insulator transition in light metal oxides. Their maximum conductivity at RT is 1500 S cm\textsuperscript{−1}, which is higher by a factor of 2 than that of graphite and is comparable to metal manganese. The optical transmittance of thin films (100 nm thick) is >60% in the visible range. The conductivity and visible transmittance of metallic C12A7:e\textsuperscript{−} is obviously inferior to that of ITO compared to TCO, but C12A7:e\textsuperscript{−} has totally unique physical properties, i.e., very low work function and superconductivity.

#### 3.5.1. Low Work Function

Toda et al. [21] evaluated work function (ϕWF) of C12A7:e\textsuperscript{−} by photoelectron yield spectroscopy (PYS) using bulk C12A7:e\textsuperscript{−} single crystals. A fresh surface of the sample was prepared by filing the surface with a diamond file in a preparation chamber attached to the measurement chamber under a vacuum of ~5 × 10\textsuperscript{−7} Pa before each photo-electron measurement. Figure 17a shows photoemission current (I\textsubscript{PE}) as a function of V\textsubscript{BIAS} at different excitation photon energies (h\nu). I\textsubscript{PE} rapidly increases with V\textsubscript{BIAS} up to 10 V and levels off. I\textsubscript{PE} is related to h\nu according to Eqn (8):

\[ \ln(I\textsubscript{PE}T^{-2}) = \ln(F(h\nu - \phiWF)\frac{k_B T}{h\nu}) + Z, \quad (8) \]

where

\[ F(x) = \int_{0}^{\infty} \frac{y}{1 + \exp(y - x)} dy \]

Here \( k_B \) is the Boltzmann constant, \( T \) the temperature, and \( Z \) a constant. Figure 17b shows \( (I\textsubscript{PE})^{1/2} - h\nu \) plots (the approximated Fowler plot) for samples A and B with different \( N_e \) at V\textsubscript{BIAS} values of 10 V, 20 V, and 30 V. The ϕWF value was determined to be 2.4 eV by obtaining the best-fit result to Equation (8). It is seen that this ϕWF value gives fairly good agreement with the theory for all the V\textsubscript{BIAS} values especially at h\nu around the ϕWF value. These results indicate that the vacuum level (\( E\textsubscript{VAC} \)) locates at ~2.4 eV above \( E\textsubscript{F} \), ~0.4 eV above FCBM, and ~7.9 eV above VBM (see Figure 18).

Here we consider the reason why C12A7:e\textsuperscript{−} has a lower ϕWF than those of alkali and alkaline earth metals (Na = ~2.8 eV, K = ~2.3 eV, Rb = ~2.2 eV, Cs = ~2.1 eV, and Ca = ~2.9 eV) and why C12A7:e\textsuperscript{−} is chemically stable regardless of the low ϕWF value. Figure 18 compares the energy structures of the F\textsuperscript{−} center (an oxygen vacancy trapping an electron) in CaO and CCB in C12A7:e\textsuperscript{−}. CCB in C12A7:e\textsuperscript{−} is composed of 3-dimensionally connected sub-nanometer-sized cages, in some of which electrons are entrapped. An electron in the cage is octahedrally coordinated with 6 Ca\textsuperscript{2+} ions constituting the cage wall. Thus, the electron-trapped cage is an analog to the F\textsuperscript{−} center in CaO except for two points: (1) the separation between the electron and the Ca\textsuperscript{2+} ion is longer by ~50 % than that in the F\textsuperscript{−} center in CaO, and (2) an oxygen ion directly bonding with the Ca\textsuperscript{2+} ion in the lattice framework of C12A7:e\textsuperscript{−} is connected one or two tetrahedrally coordinated Al\textsuperscript{3+} ions. As a consequence, the net negative charge affecting at the cage center is smaller than that at the F\textsuperscript{−} center in CaO. These differences cause a smaller Madelung potential for C12A7:e\textsuperscript{−} than for CaO,

### Figure 17

Evaluation of work function (WF) in C12A7:e\textsuperscript{−} single crystals by photoelectron yield spectroscopy. Two samples with different carrier electron concentrations (1 × 10\textsuperscript{19} or 2 × 10\textsuperscript{21} cm\textsuperscript{−3}) were used for the measurements. Photocurrent vs. bias voltage for different excited photon energy. (b) \( I\textsubscript{PE} - h\nu \) plots. The intercept yields ϕWF of 2.4 eV. For color version of this figure, the reader is referred to the online version of this book.
resulting in the ~3.6 eV shallower electronic levels in the cages of C12A7:e− compared with the alkali and alkaline earth metal oxides is also understood from its crystal structure. When C12A7:e− is exposed to an ambient atmosphere, the rigid structure of the lattice framework made of Ca−O and Al−O bonds prohibits in-diffusion of H2O and O2 molecules into a cage to react with the low work function electrons at around room temperature. This geometrical restriction is the primary cause for chemical and thermal stability of C12A7:e−, providing an opportunity for practical application.

Recently, several applications of C12A7:e− have been reported, i.e., cold [41] and thermoionic [42] electron emission sources, cathode materials in organic light-emitting diode (OLED) [43], and efficient electrode material in fluorescent lamps [44]. Better performance is obtained by using C12A7:e− for each application compared with the currently used materials. These are examples of straightforward applications based on unique properties of C12A7:e−. An intriguing application has been found for a chemical reducing reagent available for various organic reactions in aqueous media [45].

3.5.2. Superconducting Transition [46]

In order to investigate whether the superconducting transition occurs in the metallic C12A7:e− or not, two samples exhibiting the highest conductivity in each crystal growth method (FZ and CZ) were used for measurements of electrical properties at low temperatures. Here, the C12A7 electrode denotes that C12A7 in which most of free O2− ions were replaced by electrons, i.e. an electron serves as an anion in this crystal. The electrical conductivities at 300 K of the FZ and CZ single-crystal electrides were ~810 S cm−1 and ~1500 S cm−1, respectively.

Figure 19 shows the temperature (T) dependence of the electrical resistivity (ρ) of two samples. A sharp drop in resistivity is observed in the FZ single-crystal electride with an onset temperature of ~0.2 K. However, electrical resistivity of the metallic C12A7 electride prepared from the CZ single crystal never becomes zero down to 90 mK, as shown in the inset of the figure. Figure 19 (inset) shows that the onset temperature in the FZ sample decreases with the application of magnetic field, which strongly suggests that the FZ single-crystal C12A7 electride is a superconductor below a transition temperature of ~0.2 K. The superconductivity disappears by the magnetic field above a magnetic field of 30 mT. Bulk superconductivity was confirmed by the magnetic measurement, as shown in the right panel of Figure 19. It is confirmed that the FZ single-crystal electride with a conductivity of ~770 S cm−1 (sample B) also becomes a superconductor with the onset temperature of ~0.19 K. The superconducting transition was further observed with the onset temperature of ~0.16 K and ~0.4 K in epitaxial thin film electrides with Ne of 1.6 × 1021 cm−3 and 2.0 × 1021 cm−3, respectively. The variation in Tc in both single-crystal and thin film electrides suggests that the increase in Ne leads to an increase in Tc in the C12A7 electride. The critical magnetic field Hc(0) is estimated to be ~19 mT for the bulk and ~33 mT for the thin film, using
a relation of $H_c(T) = H_c(0)[1 - (T/T_c)^2]$. The low transition temperature and the small critical magnetic field imply that C12A7-e$^-$ is a type-I superconductor. Perfect diamagnetism ($\chi \sim 1$) below $0.20$ K in the $\chi$-$T$ curve was also observed, confirming the bulk superconductivity in the C12A7 electrode and 100% volume fraction of the superconducting phase.

Here, the mechanism [47] for superconducting transition in the C12A7 electride is described. In the Bardeen–Cooper–Schrieffer theory, $T_c$ is expressed by McMillan’s formula

$$T_c = \frac{Q_D}{1.45} \exp\left[-1.04 \cdot \frac{1 + \lambda}{\left(\lambda - \mu^* \cdot (1 + 0.62 \cdot \lambda)\right)}\right],$$

where $Q_D$ is the Debye temperature, $\lambda$ is the phonon-electron coupling constant, and $\mu^*$ is the Coulomb pseudopotential. $\lambda$ is provided by $\lambda = N(0)/V$, where $N(0)$ is the electron concentration at the Fermi energy, and $V$ is the phonon-mediated coupling constant for a single electron. Using the experimentally obtained value of $Q_D = 604$ K [48] from heat capacity and the assumption of $\mu^*$ as 0.2 in the low electron density system, we obtain $\lambda = -0.45$, which is much larger than those of alkali metals, as mentioned below. The observed increase in $T_c$ with $N_e$ is consistent with McMillan’s formula, provided that $N_e$ is correlated with $N(0)$.

Although the feature that the conduction electron in the electride is that the s-electron is similar to that of the alkali metal, superconducting transition is observed in C12A7 electrode but not in alkali metals at ambient pressure except $T_c = 0.4$ mK of Li [49]. The difference is attributable to the magnitude of the electron–phonon interaction. The anionic electron may interact strongly with the rigid lattice framework that is formed due to the strong covalent and ionic bonds among $\text{Ca}^{2+}$, $\text{Al}^{3+}$, and oxygen ions. These facts possibly lead to the high values of $\lambda$ (-0.45) and $Q_D$ (604 K), which are much higher than those of alkali metals such as Li ($\lambda = 0.38$, $Q_D = 420$ K), Na ($\lambda = 0.20$, $Q_D = 150$ K), and K ($\lambda = 0.14$, $Q_D = 100$ K). That is, the strong electron–phonon interactions likely resulted from the covalent and ionic bond nature of the oxide framework is responsible for the emergence of superconductivity in the C12A7 electrode. Stable electrides with a rigid framework structure and with electrons populated in free space is a new category of superconductors.

4. TRANSPARENT AMORPHOUS OXIDE SEMICONDUCTORS

4.1. History of Oxide Semiconductors and Oxide TFTs

Figure 20 shows the history of oxide semiconductors and TFTs based on them. The research on oxide TFTs started in the 1960s but almost disappeared in the open-accessible literature after that. The research reappeared in the 1990s and became rather active in the 2000s. The primary engine for accelerating the research on the oxide semiconductors and oxide TFTs is the need in rapidly expanding flat panel displays. High performance and easily processable TFTs are needed to meet strong demand for backplane to drive organic LED panels as well as next-generation LCD panels workable at higher frame rates.

An oxide semiconductor is N-type except for several materials. This is a natural consequence of chemical bonding nature in oxides. Figure 21 shows the schematic band structure of oxide and silicon. The valence band maximum (VBM) in oxides is primarily composed of 2p-orbitals of oxygen ions with strong electro-negative nature. So, the location of VBM is so deep that positive holes are difficult to be doped. On the other hand, the situation is reverse for the conduction band minimum (CBM), which is composed of metal cation’s orbitals and is located at
a rather lower level. This is the reason why P-type doping is extremely difficult but not for N-type doping in oxides [50].

Research on oxide TFTs was restricted on polycrystalline samples until late 2004. Polycrystalline N-type oxide semiconductors represented by ZnO and SnO$_2$ are sensitive to the environmental atmosphere because adsorption of oxygen-related species on the surface/grain boundaries traps a carrier electron and desorption induces the reverse process. The instability of polycrystalline oxide TFTs, which is the major obstacle, comes from the intrinsic nature of surfaces and grain boundaries in oxide semiconductors. This obstacle was practically resolved by using amorphous oxide semiconductors (AOSs) in place of polycrystalline forms in 2004 [51]. Amorphous semiconductors have outstanding advantages in processing over crystalline semiconductors. They have no grain boundaries and their thin films can be deposited at low temperatures. These features make them very favorable for large-sized thin film application such as TFTs.

It was widely believed that the electron transport properties of crystalline semiconductors are largely degraded when they become amorphous on the basis of several examples such as silicon, chalcogenides, and transition metal oxides. However, this was not true for ionic oxide semiconductors, as will be described. Ionic amorphous oxide semiconductors, which are called 'transparent amorphous oxide semiconductors (TAOS)' named after their band gaps, have large electron mobilities comparable to those in the corresponding crystalline analogs. TAOS-TFTs have mobilities of $10^{-5}$ cm$^2$(Vs)$^{-1}$, which are larger by 1–2 orders of magnitude than that in a-Si:H TFTs and can be fabricated by the conventional sputtering process.

![FIGURE 21](image) Schematic energy bands (left) Ionic nontransition metal oxides. (right) Covalent compounds. ex Si. For color version of this figure, the reader is referred to the online version of this book.
4.2. Material Design Concept of TAOS with High Mobility

In ionic oxides, the nature of CBM which works as an electron pathway totally differs from that of VBM which act as a hole pathway. CBM in ionic oxides is primarily composed of vacant s-orbitals of a cation, and the contribution of oxygen 2p-orbitals, which are dominant at VBM, is rather small. The spatial spread of this vacant s-orbital is so large that direct overlap between the s-orbitals of the neighboring cations is possible in post-transition metal oxides, and therefore an effective mass of electron is small in these oxides. In most amorphous materials, structural randomness appears prominently as the bond angle distribution. When the bond angle has a large distribution, how is the effective mass (in other words, the transfer rate between neighboring cation s-orbitals) modified for carrier electrons? We considered two cases for this question: covalent semiconductors and ionic semiconductors. In the former case, the magnitude of the overlap between the vacant orbitals of the neighboring atoms is very sensitive to the variation in the bond angle. As a consequence, rather deep localized states would be created at somewhat high concentrations and thereby the drift mobility would be largely degraded due to carrier scattering with these defects. On the other hand, the magnitude of the overlap in the latter case is critically varied by the choice of metal cations: when the spatial spread of the s-orbital is larger than the inter-cation distance, the magnitude should be insensitive to the bond angle distribution because the s-orbitals are isotropic in shape. As a consequence, we may anticipate that these ionic amorphous materials have large mobility comparable to that in the corresponding crystalline form.

In the case that the spatial spread of the metal s-orbital is small, such a favorable situation cannot be expected. The spatial spread of the s-orbital of a metal cation is primarily determined by the principal quantum number (n) and is modified by the charge state of the cation, as shown in Figure 22 [2,52]. Thus candidates for a transparent semiconductor having large electron mobilities comparable to those of the corresponding crystals are transparent oxides constituted by post-transition metal cations with an electronic configuration (n-1)d^{10}ns^0, where n ≥ 5.

Figure 23 illustrates the comparison in orbital drawings of Si and a post-transition metal (PTM) oxide between crystalline and amorphous states. The drastic reduction of the electron mobility in the amorphous state from c-Si may be understood intuitively from the figure, whereas medium mobility in c-PTM oxides is reserved even in the amorphous state. In a sense, the situation of CBM in PTM oxides is similar to that in amorphous metal alloys in the aspect that metal orbitals dominantly constitute the electron pathways. The conductivity of amorphous/liquid metal alloys remains at a slightly lower level compared with the corresponding crystalline phases. The structure of amorphous metal is modeled by dense random packing of metal spheres and occupation of a metalloid at the interstitial position. We may consider that in a-PTM oxides the vacant ns-orbitals of the metal cations work as metal atoms in a metal.

4.3. Electron Transport in Amorphous Oxide Semiconductors

Ionic amorphous oxide semiconductors have several common properties that are not seen in conventional amorphous semiconductors. First is their large electron mobility such as 10−40 cm²(Vs)^−1, which is higher by 1−2 orders of magnitude than that in a-Si:H or by several orders of magnitude than that in glassy semiconductors or amorphous chalcogenides. Second is that a degenerate state can be realized by doping. This is very different from the other amorphous semiconductors. For instance, c-Si is easily changed to the degenerate state by carrier doping (~several 100 ppm), but no such a state has been attained in a-Si:H to date. That is, conduction takes place by hopping or percolation between tail states in conventional amorphous semiconductors. This is the reason why drift mobility in the amorphous state is so small compared with that in the crystalline state. On the other hand, in a post transition oxides Fermi-level can exceed the mobility gap easily by carrier doping, leading to band conduction. It is considered that this striking difference originates from that in chemical bonding nature between the materials, i.e., strong ionic bonding with spherical potential is much favorable to form a shallow tail state having small density of state.
The last unique feature is that electron carriers are primarily injected by chemical doping. Substitutional doping, which is effective in crystalline semiconductors and a-Si:H, is very inefficient. Each ion is primarily connected by ionic bonding and has a large flexibility in a coordination structure compared with the tetrahedral unit in a-Si:H. Thus, doping (or substitution) of aliovalent ions does not yield charge carriers if neutrality of the total formal charge of the constituent ions is maintained. Therefore, an effective doping way is to alter the stoichiometry of oxygen ion by controlling the oxygen gas pressure during deposition processes or ion implantation of cations with a low electron affinity.

Among them amorphous In$_2$O$_3$-Ga$_2$O$_3$-ZnO (IGZO) has been extensively studied as the semi-conducting channel layer of transparent TFTs since the first report [51] in late 2004. Figure 24 summarizes electrical properties (Hall mobility $\mu_{\text{Hall}}$ and carrier concentration $N_e$) for the films in the In$_2$O$_3$-Ga$_2$O$_3$-ZnO system [53]. All the films were deposited at the same condition: i.e., on SiO$_2$ glass.
substrate at RT and oxygen partial pressure of 1.0 Pa. Although pure In$_2$O$_3$ and ZnO films exhibited large Hall mobilities of ~34 and ~19 cm$^2$ V$^{-1}$ s$^{-1}$, respectively, they were crystalline even if the films were deposited at RT. Moreover, it is not easy to control carrier concentration down to <10$^{17}$ cm$^{-3}$ in these films without compensation doping. Pure Ga$_2$O$_3$ forms amorphous films but carrier doping, i.e., formation of a shallow oxygen vacancy, was very difficult irrespective of the deposition condition examined. Thus, these end-member materials in this ternary system are not appropriate because of local nonuniformity due to grain boundaries, no stable amorphous phase, and/or the difficulty in carrier generation. As known in glass science, incorporation of aliovalent and different size cations is effective to enhance amorphization, and it is very favorable to introduce network forming cations. Indeed, stable amorphous phases are formed in the binary systems of In$_2$O$_3$–Ga$_2$O$_3$ (a-IGO) and ZnO–Ga$_2$O$_3$ (a-GZO) and in the ternary system of In$_2$O$_3$–Ga$_2$O$_3$–ZnO (a-IGZO). Both Hall mobility and carrier concentration rapidly decrease with increasing the Ga$^{3+}$ ion content. Hall mobilities in the a-IGO films decreased from ~25 cm$^2$ V$^{-1}$ s$^{-1}$ at $N_d$ ~ 10$^{20}$ cm$^{-3}$ to ~1 cm$^2$ V$^{-1}$ s$^{-1}$ at ~10$^{18}$ cm$^{-3}$ as the Ga$^{3+}$ ion content increased from 30% to 50%. However, we should note that the Hall mobility values here are not the maximum potential of these materials because Hall mobility largely depends on carrier concentration in TAOS due to the presence of potential barriers arising from structural randomness, as seen in Figure 25: Carrier mobility strongly depends on carrier concentration, and large mobilities are obtained at carrier concentrations larger than a threshold value (e.g., ~10$^{18}$ cm$^{-3}$ for a-IGZO) [13]. However, introduction of high-density carriers (e.g., >10$^{20}$ cm$^{-3}$) became very difficult in the larger Ga content films. This result indicates that large mobility is not easily obtained in the a-IGZO films with large Ga contents if one tries to dope carriers by impurity doping or introducing oxygen vacancies. This feature is unfavorable for TCO application, but it would not be disadvantage for semiconductor device applications because the difficulty in carrier doping by oxygen vacancy suggests better controllability and stability of carrier concentration, especially at

![Figure 25](image1.png)

**FIGURE 25** Schematic illustration of near the conduction band bottom. Arrows denote conduction paths of electron for the cases of $E_F > E_{cB}$ or $E_F < E_{cB}$. For color version of this figure, the reader is referred to the online version of this book.

![Figure 26](image2.png)

**FIGURE 26** Tail state density in a-IGZO estimated from TFT characteristics. Estimation is valid energy range below ~1 eV from the CBM. For comparison, data on a-Si:H are shown. For color version of this figure, the reader is referred to the online version of this book.
low concentrations. Even if high-density doping is difficult by choosing deposition condition, it is still possible to induce high-density carriers by an external electric field if TFT structures are employed, which may make it possible to utilize the potential large mobilities that may be available at large carrier concentrations. Figure 26 compares tail state density in a-IGZO with a-Si:H [54]. The tail state density around the CBM controlling N-channel mobility in a-IGZO is lower by 3-orders of magnitude than that of a-Si:H. Such a low tail state density makes it possible to upshift the Fermi-level to the CBM, which induces band conduction giving a high mobility comparable to that in the crystal. So far, no amorphous semiconductor in which band conduction occurs has been found except for TAOS.

Hall mobilities larger than $30 \, \text{cm}^2\, \text{V}^{-1}\, \text{s}^{-1}$ are obtained in the $\text{In}_2\text{O}_3-\text{ZnO}$ (a-IZO) systems [55], which are used as amorphous TCOs. However, controllability of carrier concentration and stability of low carrier state are not satisfactory in a-IZO films.

### 4.4. Unique Features of TAOS-TFTs

TAOS-TFTs have three unique characteristics compared with other TFTs. First is high field mobility $>10 \, \text{cm}^2\, (\text{Vs})^{-1}$. Second is easy fabrication at low temperature using conventional DC sputtering. The last is a large process allowance. The TFTs fabricated at unoptimized conditions exhibits poor performance, but the TFT performance can be much improved to that prepared under the optimized condition just by annealing at an appropriate temperature far below the crystallization temperature of TAOS. Figure 27 is an example of a-IZO-TFTs showing effectiveness of post-annealing to improve TFT performance. The annealing temperature is $250–300 \, ^\circ\text{C}$ which is much lower than the crystallization temperature ($>500 \, ^\circ\text{C}$). No distinct structural change around each metal cation was noted before and after annealing. Pronounced annealing effects are observed commonly for TAOS-TFTs [56,57]. It is worth noting that TFT performance is practically

---

**FIGURE 27** Effects of post-annealing on the performance of a-IGZO-TFTs. Top: transfer curve of TFTs fabricated under optimized and unoptimized conditions. Annealed in ambient atmosphere. Bottom: effect of various atmosphere during annealing. For color version of this figure, the reader is referred to the online version of this book.
determined by this post-annealing treatment. This is similar to the case of amorphous metal and provides a large process window for deposition. Humidity in annealing atmosphere enhances the improvement of the TFT devices [58].

Figure 28 shows the performance histograms of a-IGZO-TFTs that were fabricated on a glass substrate by conventional sputtering and subsequently annealed [59]. About 100 TFTs were fabricated from a 1 cm x 1 cm area of a-IGZO thin film. The TFT exhibits excellent uniformity and high average performance. The value of $\mu_{\text{sat}}$ resides within the range of 0.5 cm$^2$(Vs)$^{-1}$ and $\mu = 0.11$ cm$^2$(Vs)$^{-1}$ (0.76% of the average value), demonstrating the excellent uniformity of the a-IGZO-TFTs. It strongly suggests that the a-IGZO-TFTs essentially have a good short-range uniformity and are advantageous in integrated circuits and large area.

Short channel TFTs are needed for integrated circuits. It was reported by the Samsung group in 2008 that a-IGZO-TFTs keep the almost same mobilities, ON/OFF ratios, threshold voltages, and subthreshold slopes when the channel length is reduced down to 50 nm [60]. These results indicate that a-IGZO-TFT is a candidate for selection transistor in 3D-cross-point stacking memory.

Table 1 shows comparison of TAOS-TFTs with a-Si:H and poly-Si-TFTs summarized by Jeong et al. in 2008 [61]. It is obvious that a-IGZO-TFTs meet the high mobility, homogeneity, and low cost requirements. The major technical issue was instability under biasing. TAOS-TFTs are sensitive to the light corresponding to the sub-band gap and two performance changes are induced, positive shift of the threshold voltage ($\Delta V_{\text{th}}$) and increase in off-current (mobility and subthreshold slope remain unchanged). For example, such a change is induced by illumination of light shorter than 460 nm [62]. The magnitude of $\Delta V_{\text{th}}$ depends on intensity and wavelength of light. The extent is rather smaller than that of a-Si:H but is larger than that in poly-Si-TFTs.

4.5. Progress in TAOS-TFT as the FPD Backplane

TAOS-TFTs are several unique features as follows: (1) high electron mobility ranging 10–30 cm$^2$(Vs)$^{-1}$ depending on the material systems, (2) being capable by conventional sputtering at low temperatures, (3) controllable performance by post-annealing, (4) optically transparent, (5) only N-channel operation, no inversion occurs due to high tail and defect state density above the VBM, (6) a large material selection rage for gate insulator due to (4), and (7) no short channel effect down to 50 nm. Utilizing these advantages, extensive studies are being continued on the application of TAOS-TFTs s to the backplane for FPDs as well as novel display structures.

4.5.1. Novel Display Structure

An innovative e-paper display structure called the “front drive” type was recently proposed by Ito et al. of Toppan Printing [63]. Alignment of the TFT array to a color filter array is a troublesome process in the display assembling.
because a-Si:H is nontransparent and due to variation in substrate dimension with aging. Their idea to escape this difficulty was to directly deposit the TAOS-TFT arrays on the color filter arrays utilizing the low temperature process and optical transparency simultaneously. This is a first demonstration of a device structure benefited from optical transparency of TAOS. Figure 29 shows the front-drive structure applied to an electronic paper based on E-ink imaging film. Taking advantage of transparency of TAOS, the TFT and the color filter arrays can be positioned at the viewing side of displays. This display structure is applicable to other displays, facilitating the alignment. In particular, this structure appears to make it possible to apply substrate materials with large expansion coefficients such as soda-lime glasses and plastics.

4.5.2. Driving Backplane of OLED and LCD Panels

Conventional active-matrix (AM) flat panel displays are based on amorphous or polycrystalline silicon thin-film transistor (TFT) technology. Limitations of the amorphous silicon (a-Si:H) include visible light sensitivity and a low field-effect mobility, which reduce the pixel aperture ratio and driving ability for some applications. A typical example is an OLED that is driven by an electric current. Since the luminous intensity of OLEDs is proportional to the flow current, higher mobility TFTs are required. Although polycrystalline silicon TFTs have a larger field-effect mobility, its uniformity over large area is not acceptable for high yield manufacturing. Over last several years, there has been a great interest in thin-film transistors made of transparent oxide semiconductors. This is mainly due to metal oxide semiconductor thin-film transistors’ unique advantages such as visible light transparency, a large-area uniform deposition at low temperature, and high carrier mobility. However, a conventional metal oxide semiconductor such as zinc oxide (ZnO) is

![Table 1: Comparison of TFT Technology](#)

<table>
<thead>
<tr>
<th></th>
<th>a-Si:H TFT</th>
<th>Poly-Si TFT</th>
<th>Amorphous oxide TFT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generation</strong></td>
<td>8–10G</td>
<td>4G (LTPS/8G)</td>
<td>8G/8G (LTPS)</td>
</tr>
<tr>
<td><strong>Channel</strong></td>
<td>a-Si:H</td>
<td>LTPS/SPC</td>
<td>a-InGaZnO₄</td>
</tr>
<tr>
<td><strong>TFT Mask Steps</strong></td>
<td>4–5 masks</td>
<td>5–9 (11) masks</td>
<td>4–5 masks</td>
</tr>
<tr>
<td><strong>Mobility (cm²/Vs)</strong></td>
<td>&lt;1</td>
<td>50–100 (or larger)</td>
<td>1–30</td>
</tr>
<tr>
<td><strong>TFT uniformity</strong></td>
<td>Good</td>
<td>Poor</td>
<td>Good</td>
</tr>
<tr>
<td><strong>Pixel TFT</strong></td>
<td>NMOS</td>
<td>PMOS, CMOS</td>
<td>NMOS (CMOS?)</td>
</tr>
<tr>
<td><strong>Pixel circuit (OLED)</strong></td>
<td>Simple/Complex (1T + 1C/4T + 2C)</td>
<td>Complex (ex. 5T + 2C)</td>
<td>Simple/Complex (2T + 1C)</td>
</tr>
<tr>
<td><strong>Cost/Yield</strong></td>
<td>Low/High</td>
<td>High/Low</td>
<td>Low/High</td>
</tr>
<tr>
<td><strong>TFT reliability</strong></td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td><strong>Vₜ shift (Iₜₕₑₜₜ = 3 μA, 30khr)</strong></td>
<td>&gt;30 V</td>
<td>&lt;0.5 V</td>
<td>&lt;1 V</td>
</tr>
<tr>
<td><strong>Circuit integration</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Process T</strong></td>
<td>150–350 °C</td>
<td>250–550 °C</td>
<td>RT–400 °C</td>
</tr>
<tr>
<td><strong>Display Mode</strong></td>
<td>LCD, OLED</td>
<td>LCD, OLED (small size)</td>
<td>LCD, OLED, and E-paper</td>
</tr>
<tr>
<td><strong>Substrate</strong></td>
<td>Glass, metal, (plastic)</td>
<td>Glass, metal, (plastic)</td>
<td>Glass, metal, and plastic</td>
</tr>
</tbody>
</table>

Taken from J.K. Jeong, Samsung SDI/Inha Univ, M.C. Sung, LG Electr. 2nd Oxide TFT Workshop (2007), IMID2009.

![Figure 29: New display structure called ‘Front-Drive Structure’ utilizing optical transparency and RT processing.](#)
polycrystalline in nature even at room temperature. The
grain boundaries of such polycrystalline metal oxides could
affect device properties, uniformity, and stability over
large area.

Recently, OLED and LED panels driven by a-IGZO-
TFT backplane have been presented in SID and related
meetings. Figure 30 shows photographs of some FPDs with
a-IGZO-TFT backplane exhibited at the SID and associated
exhibitions. In SID ’07, LG electronics presented a 4-inch
size AM full color OLED using this backplane. This was
a first demonstration of OLED based on an oxide semi-
conductor TFTs. Subsequently, they presented a flexible
OLED at IMID ’07 by depositing a-IGZO on a thin stain-
less plate. SID ’07 was a memorial year for TAOS-TFTs
because a technical session on oxide TFT was launched for
the first time at this conference, which is the largest and
most important in the display area. In SID ’08, oxide TFTs
were highlighted as the backplane of LCD as well as
OLED. Samsung SDI demonstrated a 12.1” WXGA
AMOLED prototype display fabricated by an a-IGZO-TFT
backplane. They successfully fabricated a-IGZO THF array
by the same lithographic process as a-Si:H TFTs. TAOS-
TFTs are now attractive as a backplane of next-generation
LCDs. Larger size and high scanning frequency operation
are requirements for next-generation LCDs. When the
display size is increased to >60 inch, a greater number of
TFTs is required to keep the same pixel density. In such
a case, the ‘on’-resistance of TFT arrays cannot be
negligibly small. Furthermore, Frame frequency doubling
(60 to 120 Hz) is now spreading. Thus, the performance of
a-Si:H TFTs is insufficient to meet these requirements.
This is a reason why a-IGZO-TFTs are seriously examined
for LCD application. Samsung Electronics and SAIT
presented a 15-inch AM-LCD panel using a-IGZO back-
plane. An excellent review article [23] on OLED-panels
using a TAOS-TFT backplane was published by a Samsung
group. High activity for application of a-IGZO-TFTs to the
backplane of OLED and LCD panels is continuing to date,
through SID-2009 and SID 2010. The application is
directed to a 2k × 4k LCD panel, which operates at 240 Hz
or OLED panel. Besides Samsung groups, Sony and AUO
exhibited a 11.7-inch OLED panel and a 32-inch LED
panel, respectively.

4.6. Perspective

Fundamental research on TAOS-TFTs has been rapidly
progresses in the last 5 years. This is particularly true for
a-IGZO-TFTs, which are highly anticipated as the
universal backplane. The main advantages of a-IGZO-
TFTs are scalability, uniformity, availability of low
temperature process, and a wide range of material choices
for gate insulators. As a result, the major instability issue
appears to be resolved for practical applications. The
interest in industry is moving to manufacturing technology
and production instrument for sputter-deposition of
a-IGZO-TFTs on G8 glass substrates out into market. Figure 31 shows the photograph of sputtering targets (crystalline IGZO) for a G8-sized glass substrate. The mass production of FPD panels which drive by this backplane started from 2012.

The next technical challenge in TAOS-TFTs will be the development of a nonvacuum process [26]. Although research on this subject is rapidly increasing, examination of device reliability has not been performed yet. The ultimate goal of this research is to fabricate high performance TFTs on plastics or polymer-coated papers leading to flexible or paper displays. Invention of innovative material processing beyond the traditional sol–gel or solution process would be needed.

Another technical challenge in oxide TFTs is to realize C-MOS. Transparent oxide semiconductors (TOS) were born from transparent conductive oxides (TCO). The requirements for TOS rather differ from those for the latter. Control of carrier concentration and carrier type is essentially important for the former. The current status of TOS is far from this ideal situation, in particular for carrier polarity control. Although many papers have reported on p-type TOSs including p-ZnO, no p-channel TFT with a field-effect mobility of $>0.1 \text{cm}^2(\text{V} \cdot \text{s})^{-1}$ has been realized to 2007. It is considered that instability and/or high gap state density is the primary reason. For example, Cu$_2$O is a well-known p-type semiconductor and has attracted as the active layer since the first TFT parent proposed by Heil in 1935. Matsuzaki et al. fabricated epitaxial thin films and obtained a Hall mobility of $\approx 100 \text{cm}^2(\text{V} \cdot \text{s})^{-1}$ at the hole concentration of $\approx 10^{13} \text{cm}^{-3}$, which is comparable to that in single crystalline Cu$_2$O. However, Cu$_2$O-based TFT did not operate enough and the estimated field-effect mobility remained $\approx 0.1$ [64]. This striking difference comes from large tail state densities. Such a situation appears to be similar to other p-type oxide semiconductors. In 2008, Ogo et al. [65] reported a p-channel TFT with the mobility of 1.4 cm$^2(\text{V} \cdot \text{s})^{-1}$ employing SnO (not SnO$_2$) as the active layer. This is a first demonstration of a p-channel oxide TFT with a mobility $>1 \text{cm}^2(\text{V} \cdot \text{s})^{-1}$, which was a long-standing target in this area. The next goal is fabrication of complementary (C) MOS by combining p-channel and n-channel oxide TFTs. Although a monopolar channel is enough for TFTs for the backplane of displays, C-MOS is applicable for logic circuits. Exploiting bipolar semiconductive oxides with low tail state densities which can be fabricated at low temperatures is leading as the most essential work for this goal.

Oxide semiconductors are easy to fabricate by conventional sputtering and are robust to oxygen and radiation, in general. If oxide-based C-MOS structure can be fabricated on various types of substrates including plastics, flexible electronic circuits would be promising. Of course, the formation of a hetero-junction between a TOS and an organic semiconductor is a practical and promising way to active device application [66] such as a photosensor, a C-MOS, and a solar cell.

### 5. IRON-PNICTIDE SUPERCONDUCTORS

#### 5.1. Discovery of Iron-Based Superconductors

##### 5.1.1. From p-Type Transparent Semiconductor to Magnetic Semiconductors

A layered compound composed of alternatively stacked active layers is an attractive platform for exploration of wide gap semiconductors and/or functions due to the following reasons: (1) carrier doping layer can be separated spatially from the carrier transport layer, which may enhance carrier mobility through suppression of impurity scattering; (2) electron and/or holes are confined in the active layers which serve as wells, stabilizing exciton at room temperature; (3) magnetic interaction tends to be weakened because of intrinsic magnetic instability in 2 dimensions; and (4) the band gap is increased due to suppression of 3-dimensional band dispersion. We chose
a Cu\(^{+}\)-based mixed oxyanion layered material as the candidate for p-type transparent oxide semiconductors (p-TOS). In general, a hole at the valence band top tends to localize at oxygen 2p-orbitals primarily constituting the VBM, due to its strong ionic nature. For p-type TCOs, how to delocalize the positive hole at the VBM of oxides works as a strategy. We focused a Cu\(^{+}\) with [Ar](3d\(^{10}\)) configuration which has 3d electron levels comparable to those of O 2p, leading to the discovery of p-TOS, CuAlO\(_2\) with a layered structure (delafossite). Further improved performance was obtained in a mixed oxychalcogenide LaCuOCh (Ch=chalcogenide) [67]. We found the unique optoelectronic properties for epitaxial thin films of LaCuOCh, i.e., large hole mobility, RT-stable exciton [68], a large 3rd order optical nonlinearity [69], 2D-electronic structure [70], and blue light-emitting diode utilized PN junction [71]and efficient excitonic emission.

Intriguing properties of transition metals primarily come from the freedom combined with spin and charge of 3d-electrons. Although Cu\(^{+}\) with a closed 3d-shell structure is appropriate for keeping optical transparency (a large gap), no such properties can be expected for its compounds. We thus determined to start the exploitation of new magnetic semiconductors in the LaTMPnO system (where T\(_M\) is a 3d transition cation with +2 charge state) with the same crystal structure as LaCuChO as shown in Figure 32.

### 5.1.2. Electromagnetic Properties of LaTMPnO

The binary compounds of transition metal pnictides exhibit a diversity of electromagnetic properties by changing the combination of T\(_M\) and Pn elements. However, few studies on electromagnetic properties of layered compounds LaTMPnO had been performed notwithstanding that the existence of many compounds were already reported. We tried to synthesize the bulk compounds LaTMPnO (TM:3d transition metal, Pn:P, As, Sb) by solid-state reactions in evacuated silica glass tube and measured electromagnetic properties on the resulting materials. For device application, thin film fabrication by pulsed laser deposition method was examined.

Figure 33 summarizes the electromagnetic properties elucidated for LaFeAsPnO. The properties drastically vary with the 3d electron number in the TM: an anti-ferromagnetic semiconductor with Neel temperature (T\(_N\)) >400 K for the Mn system, a Pauli para metal with a superconducting critical temperature (T\(_c\)) at ~4 K for Fe, a ferromagnetic metal with a Curie temperature of ~60 K for Co, a Pauli para metal with T\(_c\) = 3 K for Ni, and a diamagnetic semiconductor for Zn. It is of interest to note that metal—superconducting transition is observed for T\(_M\)\(^{2+}\) with even number of 3d-electrons, i.e., Fe\(^{2+}\) with 3d\(^6\) and Ni\(^{2+}\) with 3d\(^8\). The optical properties of LaTMPnO rather differ from those of binary TMPn, but the magnetic properties are similar to each other: anti-ferromagnetic (AFM) for Mn and FeAs, ferromagnetic (FM) for Co, and Pauli para for Fe and Ni. Noteworthy is the striking difference between FeP and FeAs. FeP is a Pauli para metal, while FeAs is an AF metal.

#### 5.1.3. LaFePnO: Striking Difference Between LaFePO and LaFeAsO

Figure 34 shows ρ–T curves between LaFePO and LaFeAsO\(_{1-x}\)F\(_x\). The resistivity of the former simply decreases with temperature and suddenly drops to zero around 4 K [72], while the latter has higher resistivity by an order of magnitude than the former but does not change monotonically with temperature, showing a distinct break around 150 K and ρ increases with T. When electrons are doped via partial replacement of the oxygen sites by fluorine, this break disappears and instead the bulk superconductivity of T\(_{\text{onset}}\) of 32 K emerges [73]. This break in the ρ–T curve is not seen for LaFePnO and LaNiPnO with T\(_c\) = 3K [74]. After we submitted a communication reporting LaFeAsO\(_{1-x}\)F\(_x\) with T\(_{\text{mid}}\) = 26 K, we concentrated to examine the superconductivity in the compounds with multi Fe(Ni)Pn layers following a drastic jump of T\(_c\) in cuprates by the discovery of YBCO with bilayers of CuO\(_2\). The first material we chose was BaNi\(_2\)P\(_2\) [75]with 122 structures. We found a T\(_c\) but it remained ~4 K. The next compound we chose was BaFe\(_2\)As\(_2\) and tried to dope electrons by replacing oxygen sites by fluorine but the result was totally unsuccessful (it took almost 2 years to succeed in synthesizing superconducting Sr122 compounds by electron doping via...
aliovalent substitution of alkaline earth ion with a help of high pressure [76]). At this time, a paper reporting Ba$_{1-x}$K$_x$Fe$_2$As$_2$ with a maximum $T_c = 38$ K by the Jahrendt group of Munich [77] was posted on the cond-mat preprint-server. It is evident that BaFe$_2$As$_2$, the parent compound of $T_c = 38$ K, exhibits a break at $\sim$140 K like LaFeAsO but BaNi$_2$P$_2$ has no such anomaly as LaFePO and LaNiAsO ($T_c = 3$ K). So, it is natural to postulate that the presence of a break in the $\rho$–$T$ curve is requisite for the parent material for high $T_c$; but the invalidity of this idea will be shown later.

What happens around this break temperature? First, we suspected crystallographic transition and measured powder XRD data by using the SR-ring as a function of temperature. As a consequence, a phase transition from tetragonal to orthorhombic (not monoclinic from the extinction rule)
symmetry was clearly observed at ~155K [78]. Measurements of heat capacity and the internal field at the Fe-site monitored by Mossbauer spectroscopy revealed that crystallographic transition and anti-magnetic ordering occur at ~150 K and ~140 K, respectively. The most decisive information on the magnetic structure was elucidated by neutron diffraction. The spin configuration and the magnitude of local moment at a Fe ion at the AFM state in the parent compound were determined by a group of Oak Ridge National Lab (ORNL) [79]. Figure 35 shows electronic phase diagrams for the 1111-type and 122-type iron-based compounds. Superconductivity emerges by introducing carriers to the parent compounds which are anti-ferromagnetic metals. The major differences between these 2 systems are the presence of overlapped region between the AF phase and the superconducting phase and convergence of structural transition and magnetic transition temperature for the 122 system.

5.2. Advances in Materials [80–82]

When we discovered the high $T_c$ in LaFeAsO$_{1-x}$F$_x$ in October, 2007, the target was switched from LaFeAsO$_{1-x}$F$_x$ and found a raise of $T_c$ to 43 K under a pressure of ~4 GPa. This temperature exceeded the $T_c$ of MgB$_2$ and is next to the high $T_c$ cuprates. We submitted this paper on February 26, 2008 and started to synthesize LuFeAsO$_{1-x}$F$_x$ at the same time. It was evident that high pressure effectively works to raise the $T_c$ in this system. We thought if the La ion with the largest radius among rare earth ions could be replaced by nonmagnetic rare earth ions with the smallest radius, the highest $T_c$ would have been realized. This trial was very unsuccessful as well. The compound LuFeAsO itself was not obtained (even to date). During struggle with the synthesis of LuFeAsO, higher Tcs were reported on a condmat-server by several Chinese groups at USTC and IOP in early March. They obtained the high $T_c$ by replacing the La ion with Ce, Pr, or Sm ion with an open 4f-shell. These reports surprised us because we intentionally avoided the use of magnetic rare earth ion based on the experimental result (presented at the 15th Conference on Intercalation at Korea in May, 2007) that superconductivity disappears when the La ion in LaFePO ($T_c = 4$ K) was replaced by Ce. That is, we misunderstood without careful checking whether or not the 4f/5d orbitals of rare earth ions participate to the Fermi-level which are primarily composed of Fe 3d orbitals.

Figure 36 summarizes the progress in Fe(Ni)-based superconductors viewed from material aspects. The abscissa is the date of receipt in a journal or posted through the preprint-server. The $T_c$ reached a maximum of 56 K in a short period by replacing the La ion in LaFeAsO with an appropriate rare earth ion. Two types of new materials were reported in June: one is Ba$_{1-x}$K$_x$Fe$_2$As$_2$, 122-type, by Johredt’s group and the other FeSe$_{1-x}$, 11-type, by Wu’s group of Taiwan. Subsequently, Na(Li)FeAs, 111-type was reported by Chu’s group of Houston and Jin’s group of IOP, and in March 2009, Shimoyama’s group of Tokyo found Tc in FePn compounds with perovskite-like blocking layers. These 5 types of compounds summarized in Figure 37 have a common structural unit, a square lattice of Fe$^{2+}$ ions (Figure 38). The Fermi level is primarily composed of Fe 3d orbitals and the contribution of pnicogen/chalcogen orbitals is ~10% at the Fermi level. This situation is rather different from high Tc cuprates in which both orbitals of Cu 3d and O 2p participate to the Fermi level.

In June 2008, a unique phenomenon was posted by a group of ORNL. They found that superconductivity is induced by partially replacing the Fe-sites in the 1111 compounds by Co ions. The similar results were observed in the 122 phases. This result strikingly differs from that in

![Figure 35](image-url)

FIGURE 35  Phase diagrams for 1111 and 122 Iron-pnictide superconductors. $T_s$: crystallographic transition temperature, $T_N$: magnetic transition (Neel) temperature. QCP: quantum critical point (phase transition at 0 K). For color version of this figure, the reader is referred to the online version of this book.
FIGURE 36  Progress in iron (nickel)-based superconductors. The ordinate is the date of the paper accepted by the journal or posted on cond-mat preprint-server. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 37  Crystal structure of 5 representative types of parent materials. For color version of this figure, the reader is referred to the online version of this book.
high $T_c$ cuprates in which partial substitution of Cu sites by other ions diminishes $T_c$. The group led by Clearfield at Iowa found that $T_c$ in Fe-substituted 122 compounds with a 3d/4d transition metal could be scaled by the excess number of electrons per Fe ion.

5.3. Current Status

5.3.1. What is the Primary Factor Controlling $T_c$?

It was implied from striking difference in $T_c$ between LaFePO and LaFeAsO systems or BaFe$_2$As$_2$ and BaNi$_2$As$_2$ systems that materials with higher $T_N$ exhibit high $T_c$. However, this idea is discounted by a plot in Figure 39 showing the anti-correlation between $T_c$ and $T_N$ for each material system. If this anti-correlation is valid, highest $T_c$ should be derived from the parent material with higher $T_N$, but the fact is reverse. Thus, this implication is invalid.

Lee et al. [83] found a good correlation between $T_c$ and the bond angle of Pn(Ch)-Fe-Pn(Ch), i.e., high $T_c$ is observed as this bond angle approaches that (109.28') of the regular tetrahedron. Kuroki et al. [84] proposed a height ($h_{Pn}$) of pnictgen (chalcogen) from the iron plane as a measure of $T_c$ on the basis of calculation on spin-fluctuation-mediated superconductivity using five 3d orbitals of Fe. Since the energy level of a band derived from Fe 3d$_x$2/C0$_y$2 is most sensitive to $h_{Pn}$, the shape of the Fermi surface is rather changed with $h_{Pn}$ as illustrated in Figure 40. The $\gamma$-pocket at ($\pi$, $\pi$) that is hidden for low $h_{Pn}$ appears as $h_{Pn}$ is increased. As a consequence, nesting becomes satisfactory and $T_c$ is increased as $h_{Pn}$. Figure 41 shows a plot of $T_c$ vs. $h_{Pn}$. Although there are some exceptions such as the 11 system, a good correlation is observed as a whole.

5.3.2. Comparison with Cuprates and MgB$_2$

Table 2 summarizes the features of iron pnictide along with cuprates and MgB$_2$ (the representative of BCS superconductors). Three unique properties are evident for iron pnictides: robustness to impurity doping, very high upper critical magnetic fields, and low crystallographic anisotropy in physical performances. These features are favorable for application to superconducting wires.

5.4. Perspective

Iron, a representative magnetic element, was believed to be the last constituent for emergence of superconductivity because long range magnetic ordering competes with the
formation of a Cooper pair requisite for superconductivity. However, once LaFeAs(O, F) with $T_c = 26$ K was discovered, many iron pnictide (chalcogenide) superconducting materials have been found and the maximum $T_c$ reached 56 K, which is next to the high $T_c$ cuprates exceeding MgB$_2$. I think there are two significances to the discovery of iron-based superconductors. First, we realized that a magnetic element is not a hateful enemy but a powerful friend to realize high $T_c$ superconductors. Second, it provides a large opportunity to find new high $T_c$ superconductors because there exist several hundreds of layered compounds containing square lattice of transition metal cations taking tetrahedral coordination with nonoxide anions. We expect materials with higher $T_c$ and/or novel class of superconductors would be hidden among these. To our interest, the crystal structure of 122 is the same as that of a representative heavy fermion superconductor CeCu$_2$Te$_2$($T = Si, Ge$). One may expect some clue to bridge these two superconducting systems would be found.

What we should not forget is a historical fact that most groundbreaking materials including high $T_c$ superconductors have been discovered by serendipity in the course of a concentrated exploration effort. I am anticipating new material functions would be discovered as a result of concentrated material exploration with the help of theoretical modeling and advanced characterization. Iron is the most important element that led to a leap of civilization. I hope iron would serve the same role in the history of superconductivity. Strike while the iron is hot. I think this saying is still true for superconductivity research.

REFIGURE 40  Change in Fe 3d energy level and Fermi surface with the height of pnictgen from the iron plane ($h_{Pn}$). When $h_{Pn}$ increased, nesting becomes satisfactory via appearance of a $\gamma$-pocket. This result well explains the striking difference in $T_c$ between LaFePO and LaFeAsO$_{1-x}$F$_x$. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 41 Correlation between $T_c$ and $h_{Pn}$. For color version of this figure, the reader is referred to the online version of this book.
6. FUTURE CHALLENGE: UBIQUitous ELEMENT STRATEGY

Although more than 100 elements are known to exist to date, the number of elements which are practically available for materials are restricted to 60–70 due to toxicity, abundance, and radioactivity. Our territorial crust is primarily composed of oxides of representative metal, i.e., Si, Al, Ca, Na, Mg, and so on. Iron and titanium are only two transition metals included in the top 10 elements in the Clarke number. Human beings have created cultures by utilizing these abundant elements as represented by the Iron Age which was opened by creation of the plow leading to the development of agriculture. The present information era is supported by silicon (for memory and the central processing unit in the computer) and silicon dioxide (for optical fibers). It is natural to consider that the coming era, which is facing social difficulties such as energy, resource, and environmental issues, should be supported by innovative materials based on abundant elements as well. We call the approach to realize the important material function using abundant elements with high Clarke number, the ubiquitous element strategy. The major fraction of elements with high Clarke number belong to the light representative element group; this strategy may be regarded as innovative materials science based on light representative metal oxides which are widely used as the main ingredient for traditional ceramics such as cement, glass, and porcelain.

The function of materials is determined by combination of elements and structures. Since the number of abundant elements is limited to 20–30, the key to realizing the ubiquitous element strategy is how to find or create a structural element in Figure 42. What we can now raise as the structural element are nanostructure, surface and interface, and defects including nonconventional valence state. Materials scientists have an image for each element based on their accumulated experience and knowledge. How to realize a functionality which has been essential for a specific element using a totally different element by utilizing an innovative structural element and specific internal freedom of element (charge, orbital, spin, etc.) is the most successful example of this strategy. Prerequisite is a deep understanding of role of a key element in the material functionality in question. Novel information which can be probed by advanced characterization and theoretical modeling will be the platform for this understanding. In the last decade science and technology of nano-sized materials have been much advanced at the global scale under the extensive sponsorship of the nation. Ubiquitous element strategy is thus ranked as a real challenge of nanoscience and technology accumulated to date.

![Figure 42](https://example.com/figure42.png)
ACKNOWLEDGMENTS

This work was supported by the ERATO-SORST Project (October 1999—March 2010) sponsored by the Japan Science and Technology. A part of the research described herein was obtained by the Element Science and Technology Initiative Project sponsored by MEXT, JAPAN, and the FIRST Project by JSPS. The author appreciates the collaborators who participated in this research.

REFERENCES


1. INTRODUCTION
Rapid prototyping (RP) is a revolutionary and powerful technology with wide range of applications. The process of prototyping involves quick building up of a prototype or working model for the purpose of testing the various design features, ideas, concepts, functionality, output and performance. The user is able to give immediate feedback of the prototype and its performance. RP is an essential part of the process of system designing and it is quite beneficial as far as reduction of project cost and risk are concerned. Depending on the technologies involved, it is referred as Solid Freeform Fabrication (SFF) or freeform fabrication, digital fabrication, automated freeform fabrication, 3D printing, solid imaging, layer-based manufacturing, laser prototyping and additive manufacturing.

1.1. History in RP Techniques
The development of RP involves the development of computer applications in the industry. The increase in the use of computers has engaged the advancement in many computer-related areas including Computer-Aided Design (CAD), Computer-Aided Manufacturing (CAM) and Computer Numerical Control (CNC) machine tools. In particular, the apparition of RP systems is directly related to the CAD technology. However, many other technologies and advancements in other fields such as manufacturing systems and materials have played an important part in the development of RP systems.

CAD technology (Computer-Aided Design Technology—TCAD) is a branch of electronic design automation that mostly models semiconductor fabrication and semiconductor device operation [1]. The modeling of the fabrication is termed Process TCAD, while the modeling of the device operation is termed Device TCAD. CAM is the use of computer software to control machine tools and related machinery in the manufacturing of work pieces [2]. The CAM technology involves planning, managing, and control manufacturing operations. CAM systems have become possible through the development of computers, electronically operational, accurate, versatile, and productive machine tools, and through the development of sophisticated software technologies. Through the use of CAM systems it is possible to generate numerical control (NC) instructions to control a machine based on geometric information from a CAD database.

Computer-aided engineering (CAE) is the broad usage of computer software to aid in engineering tasks [3]. CAE technology is related to the use of computer systems in order to establish CAD geometry, and simulating the behavior of the product. Through a CAE system, the CAD geometry can be redefined and optimized if necessary. CAE tools are available for a wide range of analyses, such as: stress and strain evaluation, heat transfer analyses, magnetic field distribution, fluid dynamics, vibration and kinematics analyses, etc.

Finite element analysis tools usually solve each of these types of problems, by transforming a physical file into a simplified model made up of interconnected elements.

The historical development [4] of the technologies related to RP is as follows:

● Sixties: The first RP techniques became accessible in the later eighties and they were used for production of prototype and model parts. The history emerged in the
late sixties, when Professor Herbert Voelcker questioned himself about the possibilities of doing interesting things with the computer controlled and automatic machine tools. These machine tools had just started to appear on the factory floors then. Voelcker was trying to find a way in which the automated machine tools could be programmed by using the output of a design program of a computer.

- **Seventies**: Voelcker developed mathematical algorithms that clearly describe the three-dimensional aspects and resulted in the earliest theories of mathematical theories for solid modeling. These theories range from the basis of modern computer programs used to design almost all mechanical objects [5]. They changed the designing methods in the seventies, but, the old methods for designing were still very much in use. The old method revolves on a computer-controlled machine.

- **Eighties**: In 1987, Carl Deckard, a researcher from the University of Texas, came up with a good revolutionary idea. He pioneered the layer-based manufacturing, wherein he thought of building up the model layer by layer [6]. He printed 3D models by utilizing laser light for fusing metal powder in solid prototypes, single layer at a time. Deckard developed this idea into a technique called Selective Laser Sintering (SLS). The results of this technique were extremely promising. The history of RP is quite new and recent. As this technology of RP has such wide ranging scope and applications with amazing results, it has been tremendously developing over the years.

- **Present-day RP**: Today, the computer engineer has to simply sketch the ideas on the computer screen with the help of a design program that is computer aided. Computer-aided designing allows to make modification as required and it is therefore possible to create a physical prototype that is a precise, as well as proper 3D object. Voelcker and Deckard researches and innovations have given extreme development to this significant new industry known as RP or freeform fabrication. It has revolutionized the designing and manufacturing processes. There are many references of people contributing to the development of the RP technology such as Charles Hull’s patent of Apparatus for Production of 3D Objects by Stereolithography (SL) (Table 1).

### 1.2. Main Notions on Prototype

#### 1.2.1. Definition of a Prototype

Prototypes are created in order to realize the conceptualization of a design. It is often used as part of the product design process to allow engineers and designers the ability to explore alternatives, test theories and confirm performance prior to starting production of a new product. Engineers use their experience to tailor the prototype according to the specific unknowns still present in the intended design. Some prototypes are used to confirm and verify consumer interest in a proposed design; other prototypes will attempt to verify the performance or suitability of a specific design approach [8]. The fabrication of prototypes involves several steps—material removal, castings, molds, joining with adhesives etc. and with many material types—aluminum, zinc, urethanes, wood, etc. [9].

Prototyping processes have gone through three phases of development, the last two of which have emerged only in the last 20 years. Like the modeling process in computer graphics, the prototyping of physical models is growing through its third phase.

In general, an iterative series of prototypes are designed, constructed and tested as the final design is prepared for production. Multiple iterations of prototypes are used to progressively refine the design. A common strategy is to design, test, evaluate and then modify the design based on analysis of the prototype. In many products it is common to assign the prototype iterations Greek letters. For example, a first iteration prototype is called an “Alpha” prototype. Often this iteration is not expected to perform as intended and some amount of failures or issues are anticipated. Subsequent prototyping iterations (Beta, Gamma, etc.) will be expected to resolve issues and perform closer to the final production intent. In many product development organizations, prototyping specialists are employed—individuals with specialized skills and training in general fabrication techniques that can help bridge between theoretical designs and the fabrication of prototypes. An example of a resonator prototype destined for high frequency applications is shown on the Fig. 1 below:

#### 1.2.2. Types of Prototypes

The implementation of a prototype covers the range of prototyping the complete product (or system) to...
prototyping part of, or a sub-assembly or a component of the product. The complete prototype, as its name suggests, models most, if not all, the characteristics of the product. It is usually implemented full-scale as well as being fully functional. The main prototype categories are:

- **Proof-of-Principle Prototype (Model)** (in electronics sometimes built on a breadboard). A Proof of concept prototype is used to test some aspects of the intended design without attempting to exactly simulate the visual appearance, choice of materials or intended manufacturing process [11]. Such prototypes can be used to “prove” out a potential design approach such as range of motion, mechanics, sensors, architecture, etc. These types of models are often used to identify which design options will not work, or where further development and testing is necessary.

- **Form Study Prototype (Model)**. This type of prototype will allow designers to explore the basic size and the look of a product without simulating the actual function or exact visual appearance of the product. They can help assess ergonomic factors and provide insight into visual aspects of the product’s final form. Form Study Prototypes are often hand-carved or machined models from easily sculpted, inexpensive materials (e.g. urethane foam), without representing the intended color, finish, or texture. Due to the materials used, these models are intended for internal decision making and are generally not durable enough or suitable for use by representative users or consumers.

- **User Experience Prototype (Model)**. A User Experience Model invites active human interaction and is primarily used to support user focused research. While intentionally not addressing possible esthetic treatments, this type of model does more accurately represent the overall size, proportions, interfaces, and articulation of a promising concept [12]. This type of model allows early assessment of how a potential user interacts with various elements, motions, and actions of a concept which define the initial use scenario and overall user experience. As these models are fully intended to be used and handled, more robust construction is necessary. Materials typically include plywood, REN shape, RP processes and CNC machined components. Construction of user experience models is typically driven by preliminary CAID/CAD which may be constructed from scratch or with methods such as industrial CT scanning.

- **Visual Prototype (Model)** will capture the intended design esthetic and simulate the appearance, color and surface textures of the intended product but will not actually embody the function of the final product [13]. These models will be suitable for use in market research, executive reviews and approval, packaging mock-ups, and photo shoots for sales literature.

- **Functional Prototype (Model)** (also called a working prototype) will, to the greatest extent practical, attempt to simulate the final design, esthetics, materials and functionality of the intended design. The functional prototype may be reduced in size (scaled down) in order to reduce costs [14]. The construction of a fully working full-scale prototype and the ultimate test of concept is the engineers’ final check for design flaws and allow last-minute improvements to be made before larger production runs are ordered.

### 1.3. Classification of RP Techniques

RP systems are mainly classified on the nature of the material that the prototype or part is built with. In this manner, all RP systems are categorized into liquid-based, solid-based and powder-based. We will present in detail the most common 3D RP techniques used in the manufacturing of ceramic materials.

#### 1.3.1. Liquid-based

In liquid-based RP systems, the initial form of its material is in liquid state [15]. Through a process of photopolymerization or curing, the liquid is converted into the solid state. The following RP systems fall into this category:

- 3D Systems’ Stereolithography Apparatus (SLA)
- Solid Ground Curing (SGC)
- 3D Systems Microstereolithography Apparatus (μSLA)
- Fused Deposition Modeling
- Solid Object Ultraviolet-Laser Printer (SOUP)
- Soliform System
- Two Laser Beams

As is illustrated in Fig. 2, two methods are possible under the liquid polymerization processes: the SGC and the microfabrication (SLA/μSLA processes) which are most
commonly used due to a better spatial resolution of the processes. SL/μSL processes will be presented in Chapter 3.

1.3.2. Solid-based

Solid-based RP systems enclose all materials in the solid state and the shapes could be in the form of a wire, a roll, laminates and pellets [17].

The following RP techniques fall into this definition:

- Laminated Object Manufacturing (LOM)
- Fused Deposition Modeling (FDM)
- Paper Lamination Technology (PLT)
- 3D Systems’ MultiJet Modeling System (MJM)
- ModelMaker and PatternMaster
- Slicing Solid Manufacturing (SSM)
- Extrusion Modeling and Multifunctional RPM Systems
- Laser Engineered Net Shaping (LENS)
- Direct Shell Production Casting (DSPC)
- Multiphase Jet Solidification (MJS)
- Electron Beam Melting (EBM)
- Direct Metal Deposition
- 3D Printing Process

All the above RP systems employ the Joining/Binding method. The method of joining/binding differs for the above systems in that some employ a laser while others use a binder/glue to achieve the joining effect [18].

1.4. Advantages and Problems Encountered in RP Processes

RP has various advantages relative to regular, traditional prototyping and also to the technologies which have taken its concept further. The revolutionary capability which it introduced is the automatic creation of physical 3D forms from computer images [19]. Moreover, the additive fabrication technique which it uses has the ability to produce almost any geometric shape or feature with absolute fidelity to the original design. With suitable material, shapes can be highly convoluted and even include parts nested within
parts. The technique has had a significant impact on certain fields of production engineering (Fig. 3). The ability to produce an accurate model of a product or component direct from the design greatly facilitates communication between designer, manufacturer, marketer, purchaser and user, all of whom can visualize the final product and, where relevant, make sure they are entirely satisfied before manufacturing commences. This in turn decreases development time because changes can be made, and errors detected and rectified, before production, thus eliminating or decreasing costly mistakes. Projects are less risky, take less time, and their overall cost is less so output can be higher. Once the product has been manufactured and gone into use, the same time and cost savings apply when it becomes desirable to make adjustments to improve its effectiveness. Thus necessary features can be added and redundant features eliminated comparatively early in the life of the design. Additional advantages of RP include the nonindustry-specific nature of the technique and the facility for creating a large number of variants on a single model. Certain other processes have been developed taking the concept of RP a step further like rapid tooling and rapid injection molding, both of which have advantages and disadvantages compared with RP itself. RP is faster than the other two procedures mentioned and can produce complex shapes. If only a few parts are needed it is relatively inexpensive (in a production engineering context), but for a quantity of parts there is no saving because each one takes as long and costs as much as the first. The finish on the product tends to be rough, and there are only a few materials which can be used. If strength or heat-resistance need to be tested, for example, this can make the process a poor choice for assessing a part’s function. For rapid tooling there is still a limited choice of materials, and the need to make molds restricts the complexity of the designs which can be produced. Also the time taken is a little longer. However the quality of the parts is significantly superior to that produced by RP. Rapid injection molding also provides good quality components, and additionally, because of the wider variety of materials which can be used in the metal molds, the functionality of the parts can be more readily tested. The time it takes is comparable with

![FIGURE 3](#)  
**FIGURE 3** RP versus full production model [7]. For color version of this figure, the reader is referred to the online version of this book.
that of rapid tooling, and can be contrasted very favorably with that of traditional injection molding which, while providing the ultimate in quality, takes weeks or months as opposed to days.

This technique requires a specialist input at every stage, starting with the engineer who creates the CAD design. The costs are high. Then, the conversion of the CAD design into the STL format, and the creation of the physical prototype itself, which is likely made of resin, is of limited use for anything other than design and fit. The relative economies of these different styles of additive manufacturing are related. In one case the machine may cost more, while each prototype costs less; in another the material used to make the prototype may be more expensive, but the lead time required to finish the article is reduced thus saving worker time.

Despite all of the benefits derived from this technology, RP has several disadvantages: it sometimes fails in replication of the real product or system. It could happen that some important developmental steps could be omitted to get a quick and cheap working model. This can be one of the greatest disadvantages of RP. Another disadvantage of RP is one in which many problems are overlooked resulting in endless rectifications and revisions. One more disadvantage of RP is that it may not be suitable for large sized applications [20]. The user may have very high expectations about the prototype's performance and the designer is unable to deliver these. The system could be left unfinished due to various reasons or the system may be implemented before it is completely ready. The producer may produce an inadequate system that is unable to meet the overall demands of the organization. Too much involvement of the user might hamper the optimization of the program. The producer may be too attached to the program of RP, thus it may lead to legal involvement.

2. PROCESS CHAIN IN CERAMIC SOLID FREEFORM FABRICATION

SFF is an important developing technology that enables the fabrication of functional objects with complex properties directly from computer data. The basic operation of any SFF system consists of slicing a 3D computer model into thin cross sections, translating the result into 2D position information, and using this information flow to control the placement of solid material [21]. This process is repeated for each cross section and the object is built up one layer at a time. SFF has historically been associated with manufacturing technologies, and used for the rapid production of visual models, low-run tooling, and functional objects. The impact of SFF goes far beyond these applications, however; the additive nature of SFF techniques offers great promise for producing objects with unique material combinations and geometries which could not be obtained by traditional manufacturing methods.

Through the years, the use of SFF techniques has increased in the fields of biomedical engineering, electronics, aerospace, architecture, and archeology.

In recent decades the development of RP techniques has led to the production of a diversity of mechanical, optical and fluidic components and systems. As almost all of the patterning techniques are appropriate only for selected materials, as a consequence, material available for microsystems technology is inevitably limited. Examples are the LIGA technique (German acronym for X-ray deep lithography, electro deposition, and polymer molding) for the production of parts of polymers and selected metals, silicon technologies, micromachining of metals or patterning of UV-sensitive glasses. For some applications, however, the specific properties of ceramics, such as high hardness, high thermal and chemical resistance or dielectric properties are very important and selective to the used SFF techniques.

Unfortunately, the required properties of the materials used make the use of the established patterning techniques impossible for most ceramic materials. Therefore the development and application of ceramic microcomponents have been retarded.

2.1. Main Processes

Ceramic components are mostly formed in the green state by consolidating a ceramic powder with the help of more or less organic additives into the desired shape. A variety of processes were developed or known techniques adapted to the special requirements of the shaping of ceramic components with patterning details in the sub millimeter range. These techniques differ in design, and costs, but have in common that a patterned mold is needed [22]. As most of these techniques require metallic molds at least in one-step of the replication process due to the applied pressure and/or temperature, these metal molds are mainly produced by more or less expensive and time-consuming techniques like erosion methods, mechanical micromachining, or even the LIGA technique.

For faster supply of models and prototypes, a large number of RP methods have been developed in the last 15 years. In 1986 the stereolithography technique of polymers (Hull 1986) has improved the resolution and precision of these manufacturing technologies, as reported by Halloran et al. [23].

However, SFF techniques like MJS, SLS, LOM or fused deposition of ceramics are not well suited for the production of high resolution ceramic components, as they still do not possess sufficient accuracy or do not provide prototypes with the properties of conventionally shaped ceramics. In order to obtain a better resolution of the final product, stereolithography technique is mostly used
[24,25]. Fabrications of patterned ceramics or ceramic parts make high demands on the precision and resolution of the molding process. As the finishing process of miniaturized ceramic components is nearly impossible, shaping has to be done by a replication step in the unfired state. To avoid high tooling costs in product development, a RPPC has been established that enables rapid manufacturing of ceramic microcomponents from functional models to small lot series within a short time. This process chain combines the fast and inexpensive supply of master models by RP with accurate and flexible ceramic manufacturing by low-pressure injection molding (LPIM).

Besides proper feedstock preparation and sufficient small grain size, the quality of the final components is mainly influenced by the quality of the master model.

Hence, the RP method must be carefully selected to meet the requirements of the component to be fabricated.

The manufacturing of ceramic parts involves the building of 3D—CAD model of the desired part. The data are transferred to the RP equipment, where a model is fabricated out of polymers. The polymer master model is then molded with liquid silicon rubber, which is subsequently used as a tool in the LPIM. The final ceramic part is obtained after dewaxing and sintering.

2.2. Process Chain

There are three fundamental fabrication processes: subtractive, additive and formative processes [26]. In the subtractive process the material is removed from the final product until the desired shape is reached. The additive process is used to create a product larger than the starting material. A material is manipulated so that its successive portions combine to form the desired object. The formative process is a building sequence where mechanical forces are applied on a material to form it into the desired shape.

The main parts of the process chain are 3D modeling, data conversion and transmission, checking and preparing, building and postprocessing (Fig. 4) [27]. Depending on the quality of the model, the process may be iterated until a satisfactory model or part is achieved. However, like other fabrication processes, process planning is important before the RP commences. In process planning, the steps of the RP process chain are listed. The first step is 3D geometric modeling. In this instance, the requirement would be a workstation and a CAD modeling system. The various factors and parameters which influence the performance of each operation are examined and decided upon. For example, if a SLA is used to build the part, the orientation of the part is an

**FIGURE 4** Process chain in ceramic freeform fabrication. For color version of this figure, the reader is referred to the online version of this book.
important factor which would, amongst other things, influence the quality of the part and the speed of the process. An operation sheet used in this manner requires proper documentation and guidelines. Good documentation, such as a process logbook, allows future examination and evaluation, and subsequent improvements can be implemented to process planning.

2.3. Master Model Fabrication

The advantage of a RPPC consists in the flexibility of the model preparation (Fig. 5). In principle the full-range of RP methods could be used if it leads to accurate models. However, due to high processing time or costs, a careful choice of the RP method is strongly recommended to obtain the desired model properties. For most ceramic components a standard stereolithography, MJM and the RMPD technique (Rapid Micro Product Development) were used. With these three methods all applications could be realized with sufficient accuracy [29]. Each route starts with the generation of a three-dimensional CAD model of the ceramic component to be produced. The 3D model is subjected to triangulation, i.e. it is approximated by a structure consisting of triangles (Fig. 6). By varying the number of these triangles, the amount of data and the resolution of the component are influenced.

If the sintering shrinkage of the ceramic is uniform in all dimensions it can be compensated by a simple rescaling of the model size. However even an inhomogeneous shrinkage may be compensated by an anisotropic scaling. For the preparation of models with a lower resolution a commercial stereolithography machine could be used. Due to the large spot size, parts with a relatively large volume can be fabricated within short times.

For the production of small-sized holes by stereolithography, problems arise from the epoxides or acrylates, which are used as precursor resins. After removing the part from the resin bath these details cannot be cleaned sufficiently from the adherent resin residues, therefore the parts are replicated with poor accuracy. For the fabrication of such items an extrusion based RP method like FDM [31] or a ballistic method like MJM [32] is used. Although these methods are less suited for microfabrication because of the inherent limited accuracy, it has been demonstrated that they meet the requirements for applications where the high surface roughness can be tolerated.

For parts with fine details or high resolution the models could be made of acrylates using microstereolithography technique, which is suited for micro dimensioning, and allows reaching a precision of about 5 μm.

2.4. File Format

The information chain used in all additive processes encloses the following steps [33]:

- Creation of a CAD model of the design
- Conversion of the CAD model to STL format

![FIGURE 5 Process steps in RP process chain [28]. For color version of this figure, the reader is referred to the online version of this book.](image-url)
- Slicing the STL file into thin cross-sectional layers
- Construction of the model by stacked layers
- Cleaning and finishing the model

**CAD Model Creation:** First, the object to be built is modeled using a CAD software package. Solid modelers, such as Pro/ENGINEER, tend to represent 3D objects more accurately than wire-frame modelers such as AutoCAD, and will therefore yield better results. The designer can use a preexisting CAD file or may wish to create one expressly for prototyping purposes. This process is identical for all of the RP build techniques.

**Conversion to STL Format:** The various CAD packages use a number of different algorithms to represent solid objects. To establish consistency, the STL (stereolithography, the first RP technique) format has been adopted as the standard of the RP industry. The second step, therefore, is to convert the CAD file into STL format. An STL file is a triangular representation of a 3D surface geometry [34]. The surface is tessellated logically into a set of oriented triangles (facets). Each facet is described by the unit outward normal and three points listed in counterclockwise order representing the vertices of the triangle. While the aspect ratio and orientation of individual facets is governed by the surface curvature, the size of the facets is driven by the tolerance controlling the quality of the surface representation in terms of the distance of the facets from the surface. The choice of the tolerance is strongly dependent on the target application of the produced STL file. In industrial processing, where stereolithography machines perform a computer controlled layer by layer laser curing of a photosensitive resin, the tolerance may be in order of 0.1 mm to make the produced 3D part precise with highly worked out details. The native STL format needs to fulfill the following specifications: the normal and each vertex of every facet are specified by three coordinates each, so there is a total of 12 numbers stored for each facet. Each facet is part of the boundary between the interior and the exterior of the object. The orientation of the facets is specified redundantly in two ways which must be consistent. First, the direction of the normal is outward. Second, the vertices are listed in counterclockwise order when looking at the object from the outside (right-hand rule). Each triangle must share two vertices with each of its adjacent triangles. This is known as vertex-to-vertex rule. The object represented must be located in the all-positive octant (all vertex coordinates must be positive). However, for nonnative STL applications, the STL format can be generalized. The normal, if not specified (three zeros might be used instead), can be easily computed from the coordinates of the vertices using the right-hand rule. Moreover, the vertices can be located in any octant. And finally, the facet can even be on the interface between two objects (or two parts of the same object). This makes the generalized STL format suitable for modeling of 3D nonmanifold objects. The STL standard includes two data formats—ASCII and binary. While the ASCII form is more descriptive, the binary form is far more common due to the very large resulting size of the CAD data when saved in the ASCII format. The first line in the ASCII format is a description line that must start with the word “solid” in lower case, followed eventually by additional information as the file name, author, date etc. The last line should be the keyword “endsolid”. The lines in between contain descriptions of individual facets as follows:

```plaintext
facet normal 0.0 0.0 1.0
outerloop
  vertex 1.0 1.0 0.0
  vertex -1.0 1.0 0.0
  vertex 0.0 -1.0 0.0
endloop
endfacet
```

**Slice the STL File:** Once the STL files are verified to be error-free, the RP system’s computer analyzes the STL files.
that define the model to be fabricated and slices the model into cross sections. The cross-sections are systematically recreated through the solidification of liquids or binding of powders, or fusing of solids, to form a 3D model. In a SLA, for example, each output file is sliced into cross sections, between 0.12 mm (minimum) and 0.50 mm (maximum) in thickness. Generally, the model is sliced into the thinnest layer (approx. 0.12 mm) as they have to be very accurate. The supports can be created using coarser settings. An internal cross hatch structure is generated between the inner and the outer surface boundaries of the part. This serves to hold up the walls and entrap liquid that is later solidified with the presence of UV light. Preparing building parameters for positioning and stepwise manufacturing in the light of many available possibilities can be difficult if not accompanied by proper documentation. These possibilities include determination of the geometrical objects, the building orientation, spatial assortments, arrangement with other parts, necessary support structures and slice parameters. They also include the determination of technological parameters such as cure depth, laser power and other physical parameters as in the case of SLA. It means that user-friendly software for ease of use and handling, user support in terms of user manuals, dialog mode and online graphical aids will be very helpful to users of the RP system.

2.5. Object Setup

Using one of several techniques, RP machines build one layer at a time from polymers, paper, or powdered metal. Most machines are fairly autonomous, needing little human intervention. The building time can take from minutes up to several hours depending on the size of the object, the precision and number of layers for 3D processes [35].

2.6. Finishing Processes

The cleaning task involves the removal of excess parts which may have remained. Thus, for SLA parts, this refers to excess resin located in entrapped portion, as well as the removal of supports. Similarly, for SLS parts, the excess powder has to be removed. Likewise for LOM, pieces of excess wood-like blocks of paper which acted as supports have to be removed [36].

As shown in Table 2, the SLA procedures require the highest number of post-processing tasks. More importantly, for safety reason, specific recommendations for post-processing tasks have to be prepared, especially for cleaning of SLA parts. It was reported that accuracy is related to the posttreatment process. Most defects in SLA-built parts occur with the use of inadequate cleaning solvents. Parts are typically cleaned with low viscosity solvent to remove unreacted photosensitive resin. Depending upon the build style and the extent of cross-linking in the resin, the part can be distorted during the cleaning process. This effect was particularly pronounced with the more open build styles and aggressive solvents. With the build styles approaching a solid fill and more solvent-resistant materials, damage with the cleaning solvent can be minimized. With newer cleaning solvents, like TDA (tridecyl acrylate monomer) introduced by 3D Systems, part damage due to the cleaning solvent can be reduced or even eliminated.

3. 3D ADDITIVE MANUFACTURING TECHNIQUES

Very sophisticated techniques have been developed in the past 10 years in order to create 3D objects. These techniques have been integrated into a new technology known as “Additive manufacturing techniques” that use different materials as listed in the Table 3 below:

All these techniques use a common principle, the transformation of a CAD model into a physical one produced layer by layer. They build complex 3D parts in one sequence without human intervention. This chapter will present these techniques in general and most attention will be focused on one of the most sophisticated ones, the stereolithography and µSLA.

<table>
<thead>
<tr>
<th>Table 2: Finishing Processes for Different RP Systems [7]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RP Technologies</strong></td>
</tr>
<tr>
<td>Cleaning: Yes, Stereolithography: Yes, Fused deposition modeling: No, Laminated object manufacturing: Yes</td>
</tr>
<tr>
<td>Posttreatment: No, Stereolithography: Yes, Fused deposition modeling: No, Laminated object manufacturing: No</td>
</tr>
<tr>
<td>Setting: Yes, Stereolithography: Yes, Fused deposition modeling: Yes, Laminated object manufacturing: Yes</td>
</tr>
</tbody>
</table>
3.1. Selective Laser Sintering

3.1.1. History

SLS was created in the University of Texas, and is a popular process used in RP and product development. The SLS technology was brought commercialized by DTM Corporation which is now called 3D Systems. It is a layer additive production process that creates three-dimensional objects using a CO₂ laser to melt, or sinter, and fuse selective powder molecules based on information supplied by a CAD file (Fig. 7).

Table 3 3D Additive Manufacturing Techniques

<table>
<thead>
<tr>
<th>3D prototyping techniques</th>
<th>Base materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS</td>
<td>Thermoplastics, metals, powders</td>
</tr>
<tr>
<td>Inkjet printing</td>
<td>Almost any alloy, metal</td>
</tr>
<tr>
<td>FDM</td>
<td>Thermoplastics, eutectic metals</td>
</tr>
<tr>
<td>Stereo/microstereolithography (SLA/μSLA)</td>
<td>Photopolymer</td>
</tr>
<tr>
<td>LOM</td>
<td>Paper</td>
</tr>
<tr>
<td>3D printing (3DP)</td>
<td>Various materials</td>
</tr>
</tbody>
</table>

Since being patented in 1989, the SLS technology has become one of the most used processes for prototyping and product development in all industries. SLS equipment and machinery have also developed and expanded to meet the criteria of users. Currently, there are five different types of machines including high-speed systems. Some of them have round platforms while others have square and rectangular build areas. The build platforms range in size from small (12" in diameter), to medium (12 × 14” rectangle), all the way up to the largest platform (20 × 20” square). The different size build areas allow users to choose how parts are built and oriented depending on how large or small the parts are designed. If parts exceed the parameters of the build area it is not difficult to build parts in the SLS process in multiple sections and then bond them in postproduction. The variety of properties that the final materials offer makes it quite easy to bond sectioned parts together with an adhesive that is strong and tough.

3.1.2. Principle

The technique, shown in Fig. 8 below, uses a laser beam to selectively fuse powdered materials, such as nylon, elastomer, and metal, into a solid object [37]. Parts are built upon a platform which sits just below the surface in a bin of the heat-fusible powder. A laser traces the pattern of the first layer, sintering it together. The platform is lowered by

![CAD file of a craniofacial implant (3D Ceram, Limoges, France).](image)
the height of the next layer and powder is reapplied. This process continues until the part is complete. Excess powder in each layer helps to support the part during the build. The powder material that is fused during the process is commonly called thermoplastic material or, in some cases, thermoplastic binders for use in metals [38]. This technology allows for these materials to be fused together in thin layers ranging between 0.003” and 0.006”. This allows SLS to create parts with accurate details and tolerances comparable to stereolithography. However, it has an added benefit in that the strength and durability of the parts it creates is much better. Additionally, the SLS process makes parts that have longer stability than stereolithography.

3.1.3. Materials

There are a variety of different types of materials available for use in the SLS process [39]. The most beneficial characteristic is how durable and functional the materials are. They are made of nylon-based materials and create plastic prototypes. Other types of materials are used for investment casting patterns, rubber-like parts, and materials that make metal prototypes. Additionally, there is continual research and development going on to bring new materials to market. Each of these materials require little to no post build processing to be ready to use, which cuts out several steps in post-processing of parts as compared to stereolithography. However, all of the selective laser sintering materials can be finished in multiple ways to meet the desire or needs of users [38]. Among other types of post-processing, parts can be sanded, painted, plated, tapped, or even machined. This allows for a higher grade of smoothness and appearance to parts and assemblies and also gives users an unlimited number of ways to use these parts. Depending on the material, up to 100% density can be achieved with material properties comparable to those found with traditional manufacturing methods.

3.1.4. Applications

The SLS process is able to produce parts and/or prototypes that fit a wide range of different applications. The most widely used application of the process is functional
prototypes [40]. Some thermoplastics provide a choice of a durable and flexible plastic or a stiff and rigid plastic that have properties very close to an injection molded part. Parts created with these materials are ideal for end users to test for form, fit, and/or function. Other than functional prototypes, parts can be used as models, whether it is a presentation model for marketing purposes, or a show-piece for display. Yet another regular application that uses this technology is casting patterns i.e. investment casting because it is capable of making patterns with high accuracy and intricate detail. The SLS material could be a polystyrene that is coated with wax to solidify the pattern after being built in the machine. Once at the foundry, the pattern is melted away prior to pouring the molten metal. The parts made usually require a highly esthetic finish which is often called a master finish. Master finishing is a postproduction process that calls for the parts to be sanded extra smoothly and can even be primed, plated or painted to the end users desire.

3.1.5. Advantages
The advantages of SLS parts are [41]:

- Fast efficient technique to build complex geometries due to the layer wise building method;
- SLS is produced faster than conventional tooling, taking off as much as 80–90% of the time it takes to create first parts; parts can be created in days as opposed to weeks;
- It is typically delivered at a lower cost compared to conventional tooling—as much as 90–95% less;
- Parts and/or assemblies that move and work that have a good surface finish and feature detail;
- This technique gives the capability of flexible snaps and living hinges as well as high stress and heat tolerance;
- Wide variety of materials such as flexible and rigid plastics, elastomeric materials, fully dense metals and casting patterns;
- Tight dimensional tolerances all the way down to thousandths of an inch;
- Finishing capabilities that include painting for presentations, tapping or threading for use and inserts for assemblies;

3.1.6. Limitations
The limitations are [42]:

- Surface roughness caused by the grain like structure;
- Lower dimensional accuracy than (SLA/μSLA), two sided tolerance >0.1 mm;
- Low surface hardness of sintered materials (HV ≈ 170).

3.2. Fused Deposition Modeling

3.2.1. History
FDM is an additive manufacturing technology commonly used for modeling, prototyping, and production applications [43]. The technology was developed by S. Scott Crump in the late 1980s and was commercialized in 1990. FDM uses the extrusion process to build 3D models. Stratasys introduced its first RP machine, the 3D modeler in early 1992 and started shipping the units later that year. Over the past decade, Stratasys has grown progressively, seeing her RP machine sales increase from six units in the beginning to a total of 1582 units in the year 2000.

3.2.2. Principle
The principle of FDM revolves in surface chemistry, heat energy and layer deposition. The process constructs three-dimensional objects directly from 3D CAD data. A temperature-controlled head extrudes thermoplastic material layer by layer [44]. The FDM process starts with importing an STL file of a model into preprocessing software. This model is oriented and mathematically sliced into horizontal layers varying from ±0.127 to 0.254 mm thickness. A support structure is created where needed, based on the parts position and geometry. After reviewing the path data and generating the tool paths, the data is downloaded to the FDM machine. This machine operates in X, Y and Z axes, drawing the model one layer at a time. In this process, a plastic or wax material is extruded through a nozzle that traces the parts cross-sectional geometry layer by layer. The build material is usually supplied in filament form, but some setups utilize plastic pellets fed from a hopper instead. The nozzle contains resistive heaters that keep the plastic at a temperature just above its melting point so that it flows easily through the nozzle and forms the layer. The plastic hardens immediately after flowing from the nozzle and bonds to the layer below. Once a layer is built, the platform lowers, and the extrusion nozzle deposits another layer. The layer thickness and vertical dimensional accuracy is determined by the extruder die diameter, which ranges from 0.013 to 0.005 inches. In the X–Y plane, 0.001” resolution is achievable. Once the part is completed the support columns are removed and the surface is finished [45] (Fig. 9).

The build material is usually supplied in filament form, but some setups use plastic pellets fed from a hopper instead. Several materials are available with different compromise between strength and temperature properties. One of the main materials used are acrylonitrile butadiene styrene (ABS) polymer, polycarbonates, polycaprolactone, polyphenylsulfones and waxes. A “water-soluble” material can be used for making temporary supports while
manufacturing is in progress, this soluble support material is quickly dissolved with specialized mechanical agitation equipment using a precisely heated sodium hydroxide solution. As thermoplastic materials, PC/ABS (Polycarbonate-ABS) is one of the most widely used industrial thermoplastics. It offers the most desirable properties of both materials—the superior mechanical properties and heat-resistance of PC and the excellent property of ABS. PC/ABS blends are commonly used in automotive, electronics and telecommunications applications. It provides a good strength and rigidity in conjunction with toughness and temperature tolerance. The produced parts present a good surface appeal.

3.2.3. Applications

FDM technique is used for the following application areas [47]:

- The creation of concept models used in early stages of product development. FDM models reduce costs and shorten development timelines;
- The creation of functional prototypes for testing purposes. These prototypes allow to test in real world environments and make decisions that have a dramatic effect on the cost to manufacture to the product;
- Fabrication of end-use parts. Without the expense and lead time of traditional tooling or machining, FDM produces end-use parts tough enough for integration into the final product. Ideal for building small quantities of parts while waiting for tooling, FDM Technology makes it possible to get the products to market faster;
- Fabrication of manufacturing tools. FDM reduces the time it takes to create manufacturing tools by up to 85%. It produces manufacturing tools such as jigs and fixtures, tooling masters and production tooling in hours—without expensive machining or tooling.

3.2.4. Advantages

- Minimal wastage. The FDM process build parts directly by extruding semiliquid melt onto the model. Only those material needed to build the part and its support
are needed, and material wastages are kept to a minimum. There is also little need for cleaning up the model after it has been built [48];

- Ease of support removal. With the use of Break Away Support System and Water Works Soluble Support System, support structures generated during the FDM building process can be easily broken off or simply washed away. This makes it very convenient for users to get to their prototypes very quickly and there is very little or no post processing necessary;

- Ease of material change. Build materials, supplied in spool form (or cartridge form in the case of the Dimension or Prodigy Plus), are easy to handle and can be changed readily when the materials in the system are running low.

3.2.5. Limitations

- Limited accuracy. Parts built with the FDM process have limited accuracy due to the shape of the material used, i.e. the filament form. The filament usually used has a diameter of 1.27 mm and limits the accuracy of the built part;

- Slow building process. The building process is slow, as the whole cross-sectional area needs to be filled with building materials. Building speed is restricted by the extrusion rate or the flow rate of the build material from the extrusion head. As the build material used are plastics and their viscosities are relatively high, the build process could not be sped up [49];

- Unknown shrinkage. As the FDM process extrudes the build material from its extrusion head and cools them rapidly on deposition, stresses induced by such rapid cooling invariably are introduced into the model. As such, shrinkages and distortions caused to the model built are a common occurrence and are usually difficult to predict.

3.3. 3D Printing

3.3.1. History

3D printing is the process of creating three-dimensional objects from digital file using a materials printer, in a manner similar to printing images on paper. The term is most closely associated with additive manufacturing technology, where an object is created layer by layer. Since 2003, there has been large growth in the sale of 3D printers. Additionally, the cost of 3D printers has decreased. The technology is mainly used in the fields of jewelry, footwear, industrial design, architecture, engineering and construction (AEC), automotive, aerospace, dental and medical industries, education, geographic information systems, civil engineering, and many others.

3.3.2. Principle

Three-dimensional printing operates by building parts in layers. From a computer (CAD) model of the desired part, a slicing algorithm draws detailed information for every layer [51]. Each layer begins with a thin distribution of powder spread over the surface of a powder bed. Using a technology similar to inkjet printing, a binder material selectively joins particles where the object formed. A piston that supports the powder bed and the part-in-progress lowers so that the next powder layer can be spread and selectively joined. This layer by layer process repeats until the part is completed [52]. Following a heat treatment, unbound powder is removed, leaving the fabricated part. Resolution is given in layer thickness and X–Y resolution.

Example:

Elaboration of composites of piezoelectric ceramics and polymers with specific spatial arrangement of phases (connectivity) for the desired electromechanical performances (Fig. 10):

**FIGURE 10**  Electronic devices made by FDM technique [50]. For color version of this figure, the reader is referred to the online version of this book.
in dpi. Typical layer thickness is around 100 μm (0.1 mm), although some machines such as the Object Connex can print layers as thin as 16 μm. X–Y resolution is comparable to that of laser printers. The particles (3D dots) are around 50–100 μm (0.05–0.1 mm) in diameter (Fig. 11).

The building sequence is shown below:

### 3.3.3. Materials

The 3DP process combines powders and binders with high geometric flexibility. The support gained from the powder bed means that overhangs, undercuts and internal volumes can be created (as long as there is a hole for the loose powder to escape). Material options, which include metal or ceramic powders, are somewhat limited but are inexpensive related to other additive processes [54]. Further, because different materials can be dispensed by different print heads, 3D Printing can exercise control over local material composition. Material can be in a liquid carrier, or it can be applied as molten matter. The proper placement of droplets can be used to create surfaces of controlled texture and to control the internal microstructure of the printed part.

### 3.3.4. Applications

Standard applications include design visualization, prototyping/CAD, metal casting, architecture, education, geospatial, healthcare and entertainment/retail [55]. Other applications would include reconstructing fossils in paleontology, replicating ancient and priceless artifacts in archeology, reconstructing bones and body parts in forensic pathology and reconstructing heavily damaged evidence acquired from crime scene investigations. More recently, the use of 3D printing technology for artistic expression has been suggested.

### 3.3.5. Advantages

Compared to other 3D techniques, 3D printing is optimized for speed, low cost, and ease of use, making it suitable for visualizing during the conceptual stages of engineering design through to early-stage functional testing [56]. No toxic chemicals like those used in stereolithography are required, and minimal postprinting finish work is needed; one need only to use the printer itself to blow off surrounding powder after the printing process. Bonded powder prints can be further strengthened by wax or thermoset polymer impregnation. The main advantages of the technology are stated below:

- Increases Innovation. Print prototypes in hours, obtain feedback, refine designs and repeat the cycle until designs are perfect;
- Improves Communication. Creates a full color, realistic 3D model to impart infinitely more information than a computer image; create physical 3D models quickly, easily and affordably for a wide variety of applications
- Speeds Time to Market. Compress design cycles by 3D printing multiple prototypes on demand, right in your office
- Reduces Development Costs. Cuts traditional prototyping and tooling costs. Identifies design errors earlier. Reduces travel to production facilities
- Wins Business. Brings realistic 3D models to prospective accounts, sponsors and focus groups

### 3.3.6. Limitations

- Lack of material strength for large sized objects [57];
- Generated objects possess rough and ribbed surface finish due to plastic beads or large size powder particles;
- 3D printers are expensive.
3.4. Inkjet Printing

3.4.1. History

The additive fabrication technique of inkjet printing is based on the 2D printer technique of using a jet to deposit tiny drops of ink onto paper. In the additive process, the ink is replaced with thermoplastic and wax materials, which are held in a melted state. When printed, liquid drops of these materials instantly cool and solidify to form a layer of the part. For this reason, the process is often referred to as thermal phase change inkjet printing. Several manufacturers have developed different inkjet printing devices that use the basic technique described above. Inkjet printers from Solidscape Inc., such as the ModelMaker (MM), use a single jet for the build material and another jet for support material. 3D Systems has implemented their MJM technology into their ThermoJet Modeler machines that utilize several hundred nozzles to enable faster build times. Inkjet printing offers the advantages of excellent accuracy and surface finishes. However, the limitations include slow build speeds, few material options, and fragile parts. As a result, the most common application of inkjet printing is prototypes used for form and fit testing. Other applications include jewelry, medical devices, and high-precision products.

3.4.2. Principle

The inkjet printing process, as implemented by Solidscape Inc., begins with the build material (thermoplastic) and support material (wax) being held in a melted state inside two heated reservoirs. These materials are each fed to an inkjet print head which moves in the X–Y plane and shoots tiny droplets to the required locations to form one layer of the part. Both the build material and support material instantly cool and solidify. After a layer has been completed, a milling head passes over the previously created layer to produce a uniform thickness. The material particles created from this step are removed by a vacuum as the milling process is taking place. Before the next layer is created, the nozzles are checked to assure that the flow path is clear. If the nozzles do not need to be cleaned, the table will move down a set distance for the next layer to begin. Once the part is complete, the support structures are melted away. This type of machines is capable of a fast production of fine parts when using multiple jet heads. However, the accuracy is lower when using multiple jet heads, and usually additional operations are required in order to obtain a uniform layer thickness.

- **Suspension inkjet** [59]:

  The main components of these inks are volatile organic compounds (VOCs), organic chemical compounds that have high vapor pressures. Inkjet printing is performed by the deposition of ceramic suspensions which dry by evaporation of the solvent that uses systems initially devoted to ink-jet printing on paper. The high print speed of many solvent printers demands special drying equipment, usually a combination of heaters and blowers. The substrate is usually heated immediately before and after the print heads apply ink.

- **UV-curable inkjet**:

  These inks consist mainly of acrylic monomers with an initiator package. After printing, the ink is cured by exposure to strong UV-light. The advantage of UV-curable inks thermally insulated tubing to individual jetting heads. The jetting heads then disperse the material in the form of tiny droplets to create part geometry. As the jetting heads continue to lay the material droplets, layers are formed. The droplets begin to cool and harden immediately after leaving the jetting head. When one layer is complete, a milling head passes over the previously created layer to produce a uniform thickness. The material particles created from this step are removed by a vacuum as the milling process is taking place. Before the next layer is created, the nozzles are checked to assure that the flow path is clear. If the nozzles do not need to be cleaned, the table will move down a set distance for the next layer to begin. Once the part is complete, the support structures are melted away. This type of machines is capable of a fast production of fine parts when using multiple jet heads. However, the accuracy is lower when using multiple jet heads, and usually additional operations are required in order to obtain a uniform layer thickness.

- **Thermal Phase Change Inkjet**

  During this process, the inkjet machines hold the build and support materials at elevated temperatures in a reservoir until the fabrication of the part beings. Once the process has begun, the liquid material moves through thermally insulated tubing to individual jetting heads. The jetting heads then disperse the material in the form of tiny droplets to create part geometry. As the jetting heads continue to lay the material droplets, layers are formed. The droplets begin to cool and harden immediately after leaving the jetting head. When one layer is complete, a milling head passes over the previously created layer to produce a uniform thickness. The material particles created from this step are removed by a vacuum as the milling process is taking place. Before the next layer is created, the nozzles are checked to assure that the flow path is clear. If the nozzles do not need to be cleaned, the table will move down a set distance for the next layer to begin. Once the part is complete, the support structures are melted away. This type of machines is capable of a fast production of fine parts when using multiple jet heads. However, the accuracy is lower when using multiple jet heads, and usually additional operations are required in order to obtain a uniform layer thickness.
is that they “dry” as soon as they are cured, they can be applied to a wide range of uncoated substrates, and they produce a very robust image. Disadvantages are that they are expensive, require expensive curing modules in the printer, and the cured ink has a significant volume and so gives a slight relief on the surface. Though improvements are being made in the technology, UV-curable inks, because of their volume, are somewhat susceptible to cracking if applied to a flexible substrate. As such, they are often used in large “flatbed” printers, which print directly to rigid substrates such as plastic, wood or aluminum where flexibility is not a concern.

UV Curable Ink Properties and Functions:

- **Photoinitiators**: Absorb the UV energy from the light source on the print head. Chemical reaction occurs that converts the liquid ink into a solid film.
- **Monomers**: Used as solvents because of their ability to reduce viscosity (thickness) and combine with other ink components. Hundred percent solids and do not release VOCs (volatile organic compounds). Monomers also add improved film hardness and resistance properties.
- **Oligomers**: Determine the final properties of the cured ink film, including its elasticity, outdoor performance characteristics and chemical resistance.
- **Colorants**: Can be dye-based or pigment-based. Usually, pigment-based because of the greater light fastness and durability of pigments compared with dyes. Pigments used in outdoor advertising and display applications have similar requirements to those used in automotive paints. Consequently, there is some crossover of use. While a pigment is selected on the basis of the required application, size control and reduction along with dispersion technique are major components of ink formulation.

3.4.3. Applications

- Three-dimensional printing constructs a prototype by “printing” appreciably thick cross-sections of material on top of one another [60];
- They are used in the production of Microelectromechanical systems (MEMS)
- Inkjet printers are used to form conductive traces for circuits, and color filters in LCD and plasma displays;
- This technique is in fairly common use in many labs around the world for developing alternative deposition methods that reduce consumption of expensive, rare, or problematic materials. These printers have been used in the printing of polymer, macromolecular, quantum dot, metallic nanoparticles, carbon nanotubes etc. The applications of such printing methods include organic thin-film transistors, organic light emitting diodes, organic solar cells, sensors, etc;
- Inkjet technology is used in the emerging field of bio-printing (Fig. 13).

**Examples:**

3.4.4. Advantages

- Multimaterial process;
- Deposition of different materials on successive layers via a multinozzle system;
- No limitation relative to the material;
- High flexibility as the 3D part is defined by CAD files [61];
- High definition (≈50 μm) controlled by the aperture of the printing head.

**FIGURE 13** Microelectronic devices using inkjet technology (Ceradrop, Limoges, France). For color version of this figure, the reader is referred to the online version of this book.
3.4.5. Limitations
- The ink is very expensive;
- The lifetime of inkjet prints produced by inkjets using aqueous inks is limited; they will eventually fade and the color balance may change. On the other hand, prints produced from solvent-based inkjets may last several years before fading, even in direct sunlight, and so-called “archival inks” have been produced for use in aqueous-based machines which offer extended life;

3.5. Stereolithography
3.5.1. History
The term “stereolithography” was introduced in 1986 by Charles W. Hull. It was defined as a method for making solid objects by successively “printing” thin layers of the ultraviolet curable material one on top of the other [62]. Hull described a concentrated beam of ultraviolet light focused onto the surface of a vat filled with liquid photopolymer. The light beam draws the object onto the surface of the liquid layer by layer, causing polymerization or cross linking to give a solid. Because of the complexity of the process, it must be computer-controlled. In 1986 Chuck Hull founded the first company to generalize and commercialize this procedure, which is currently based in Rock Hill, SC. More recently, attempts have been made at constructing mathematical models of the stereolithography process, and designing algorithms that will automatically determine whether or not a proposed object may be constructed by this process.

3.5.2. Principle
SL is a RP technique that allows the fabrication of three-dimensional ceramic pieces with final properties (mechanical, thermal...) close to those obtained by classical processing techniques. Its spatial resolution is \( \approx 50 \mu m \). SLA machines have been made since 1988 by 3D Systems of Valencia, CA. To this day, 3D Systems is the industry leader, selling more RP machines than any other company. Because it was the first technique, it is regarded as a benchmark by which other technologies are judged. Early made prototypes were fairly brittle and prone to curing-induced distortion, but recent modifications have largely corrected these problems. This technology uses a UV laser beam that induces photopolymerization of a reactive system, consisting in the dispersion of ceramic particles into a green shape solid. The polymerization of patterns in stacked layers leads to the creation of complex 3D objects, as presented in the research of Halloran and Chartier [63,64]. The data of the object to build is transferred from a three-dimensional CAD file to the automated equipment which physically builds the green part. The interest in this fabrication technique appeared in different domains of applications: microelectronics, with the elaboration of devices (filters, resonators) with high dimensional resolution and density, biomedical implants of hydroxyapatite with controlled porosities, structural complex alumina 3D parts. The process begins with the vat filled with the photocurable liquid resin and the elevator table set just below the surface of the liquid resin. The operator loads a three-dimensional CAD solid model file into the system. Supports are designed to stabilize the part during building. The translator converts the CAD data into a STL (*bmp) file. The control unit slices the model and support into a series of cross sections from 0.025 to 0.5 mm (0.001—0.020") thick. The computer-controlled optical scanning system then directs and focuses the laser beam so that it solidifies a two-dimensional cross-section corresponding to the slice on the surface of the photocurable liquid resin to a depth greater than one layer thickness. The elevator table then drops enough to cover the solid polymer with another layer of the liquid curable suspension. A leveling wiper (or vacuum blade recoating system) moves across the surfaces to recoat the next layer of resin on the surface. The laser then draws the next layer. This process continues building the part from bottom up, until the system completes the part [65]. The part is then raised out of the vat and cleaned of excess polymer. The main components of the SLA system are a control computer, a control panel, a laser, an optical system and a process chamber as shown in the figure below (Fig. 14):

Parts are built from a photocurable liquid resin that cures when exposed to a laser beam (basically, undergoing the photopolymerization process) which scans across the surface of the resin [66]. The building is done layer by layer, each layer being scanned by the optical scanning system and controlled by an elevation mechanism which lowers at the completion of each layer. The building sequence is presented below (Fig. 15):

The layer thickness is controlled by a precision elevation mechanism. It will correspond directly to the slice thickness of the computer model and the cured thickness of the resin. The limiting aspect of the RP system tends to be the curing thickness rather than the resolution of the elevation mechanism. The important component of the building process is the laser and its optical scanning system. The key to the strength of the SLA is its ability to rapidly direct focused radiation of appropriate power and wavelength onto the surface of the liquid photopolymer resin, forming patterns of solidified photopolymer according to the cross-sectional data generated by the computer. In the SLA, a laser beam with a specified power and wavelength is sent through a beam expanding telescope to fill the optical aperture of a pair of...
cross axis, galvanometer driven, and beam scanning mirrors. These form the optical scanning system of the SLA. The beam comes to a focus on the surface of a liquid photopolymer, curing a predetermined depth of the resin after a controlled time of exposure (inversely proportional to the laser scanning speed). The solidification of the liquid resin depends on the energy per unit area (or exposure) deposited during the motion of the focused spot on the surface of the photopolymer. There is a threshold exposure that must be exceeded for the photopolymer to solidify. To maintain accuracy and consistency during part building using the SLA, the polymerized thickness \( (E_p) \) and the cured line width \( (L_p) \) must be controlled. As such, accurate exposure and focused spot size become essential. Parameters which influence performance and functionality of the parts are the physical and chemical properties of the resin, the speed and resolution of the optical scanning system, the power, wavelength and type of the laser used the spot size of the laser, the recoating system, and the post-curing process. The projected density of energy \( E_i \) (\( \text{mJ} \cdot \text{cm}^{-2} \)) is related to the nominal laser power \( (P_0) \), scanning speed \( (v_s) \) and beam radius \( (\omega_0) \) through the following equation:

\[
E_i = \frac{2P_0}{\pi\omega_0v_s}
\]

(1)

3.5.3. Materials and Photopolymerization Process

There are many types of liquid photopolymers that can be solidified by exposure to electro-magnetic radiation, including wavelengths in the gamma rays, X-rays, UV and visible range, or electron-beam. The vast majority of photopolymers used in the commercial RP systems, including 3D Systems are low viscosity acrylate oligomers and monomers.

SLA machines are curable in the UV range. UV-curable photopolymers are resins which are formulated from photoinitiators and reactive liquid monomers/oligomers. There are a large variety of them and some may contain fillers and other chemical modifiers to meet specified chemical and
mechanical requirements. They are therefore called reactive suspensions. The schematic illustration of a reactive suspension is presented Fig. 16.

The choice of formulation in SL is driven by the physical and chemical properties of the reactive monomer/oligomer. The main criteria of choice are:

- **Polymerization rate (reactivity)**. The threshold or critical energy of polymerization ($E_c$) should be sufficient enough to generate the photopolymerization process according to Beer–Lambert law [67]:

$$E_p = Dp \ln \left( \frac{E_0}{E_c} \right)$$

where $E_c$ is the critical energy for photopolymerization, which is the minimum input energy necessary to trigger the curing process, $Dp$ is the cure depth or beam penetration depth.

- **The rheology of the monomer/oligomer blend and of the suspension.** The viscosity of the organic medium must be lower than 100 mPa.s at low shear rates, in order to hold powder concentrations above 55 vol% and ensure good weatherability of the ceramic particles. In presence of a volume fraction $\Phi$ of ceramic powder, the increase of the viscosity of the reactive suspension can be evaluated by the Krieger–Dougherty equation:

$$\frac{\eta}{\eta_0} = \left( 1 - \beta \cdot \frac{\Phi}{\Phi_0} \right)^{-[\eta]/\Phi_0}$$

where $\eta/\eta_0$ is the relative viscosity of the resin, $\Phi_0$ the volume fraction of the filler for a close-packed particles corresponding to an infinite viscosity (no flow) and $[\eta]$ the intrinsic viscosity which depends on the shape of particles ($\eta = 2.5$ for spheres).

- **The mechanical properties of the polymerized layers (flexibility, planarity).** The polymerized layer should exhibit high mechanical strength and planarity (Table 4).

Figure 17 and Table 4 show the main resins used in SL process.

The UV curable system is composed of a photoinitiator dissolved in a reactive oligomer. The photoinitiator is chosen as a function of the laser wavelength. A reactive diluent may be added in order to improve the flow behavior of the suspensions, due to its low viscosity. A ceramic material used is then chosen depending on its specific area, refractive index, and mean particle size. In order to improve its dispersion in the UV curable suspension and the homogeneity of the microstructure of the green part, the ceramic powder is first deagglomerated by attrition milling. This deagglomeration is performed during several hours in solvent (ethanol) with the addition of a dispersant on the dry powder basis. It is then introduced in the reactive medium. The main steps involved in the elaboration of a reactive suspension are presented in the Fig. 18.

The process through which photopolymers are cured is referred to as the photopolymerization process. It is a chain

| Table 4 Physical Properties of Different Oligomer/Monomers used in SLA |
|---------------------|------------------|------------------|------------------|
| Viscosity (mPa.s at 25 °C) | Functionality | Mechanical properties $= f$ (functionality) | Name of product |
| HDDA | 12 | 2 | Strong | Hexanediol diacrylate |
| TDA | 5 | 1 | Flexible | Tridecyl acrylate |
| UA | 25,000 | 2 | Strong | Aliphatic urethane acrylate |
| PEAAM | 70 | 3 | Strong | Amine modified polyether acrylate |
| DTPTA | 700 | 4 | Strong | Di-Trimethylolpropane tetraacrylate |
of reactions and its initiation stage is of photochemical nature. Multifunctional monomers/oligomers lead to a dense cross-linked 3D network upon curing. Radical photopolymerization is the most used reaction for stereolithography applications. It takes place within the curing of high UV reactivity monomers/oligomers having a vinyl unsaturation, such as acrylates and methacrylates. Their nature impacts the final properties of the reticulated polymer: elastic and opaque in case of acrylates; hard and transparent in case of methacrylates. Due to the faster cure
The irradiation of the monomer/oligomer instantly leads to an increase of $v_p$ that reaches its maximum in a few seconds when most of the active double bonds are consumed \([68,69]\). This phenomenon results in an increase of the viscosity of the medium, leading to a gel formation at the beginning of the polymerization. This effect is called self-acceleration or gel effect. During this stage, the mobility of free radicals is limited in the reactive medium.

The self-acceleration is followed by a rapid decrease of the polymerization rate. The viscosity continues to grow and the mobility of reactive species greatly decreases. This stage is called self-deceleration or glass stage and marks the vitrification state of the polymer. At the end of the photopolymerization reaction, the conversion rate is lower than 100%.

The technique used to determine the degree of the reaction, respectively the rate of conversion, of an acrylate double bond is Real-Time Infrared Spectroscopy (RTIR). This technique is based on the variation in absorption of reactive functions at each time of the reaction. For one of the most used oligomer-amine modified polyether acrylate (PEAAM) in stereolithography formulations, the rate of conversion at each point of the photopolymerization reaction has been estimated using the following formula:

$$\% C_{\text{acrylate}}(t) = \frac{A_0^{1000 \text{or} 1636} - A_t^{1000 \text{or} 1636}}{A_t^{1720} - A_0^{1720}} \times 100 \quad (10)$$

where $A_0$ is the initial absorbance with no radiation and $A_t$ is the absorbance at a time $t$ of the radiation. The two bands B810 and B1636 were used to follow the absorption decrease of the acrylate function (Fig. 19).

The rate of polymerization ($R_p$) has been obtained by derivation of the rate of conversion versus time.

### 3.5.4. Applications

The SLA technology provides manufacturers with cost justifiable methods for reducing time to market, lowering product development costs, gaining greater control of their design process and improving product design. It offers a variety of real world applications in diverse fields from manufacturing to biomedicine \([70]\). Before taking a design into full-scale production, stereolithography allows businesses to evaluate their design for feasibility, manufacturability, ergonomics, and esthetics without slow and costly prototyping. Models can also be tested using Optical Stress Analysis to study the effects of external loadings, torsion, tension, pressure and placed in wind tunnels to measure aerodynamics. Manufacturing firms also use SLA models for casting, tooling and production line design. Businesses also use these models as a marketing tool since it enables
them to present a physical representation of the product before production is complete. SL prototyping lowers the cost of production while increasing product durability, benefiting both industry and consumers. SLA is also becoming prevalent in the field of biomedical engineering. Surgeons are able to make models of patients and implants to practice delicate routines in advance. These models are also being used to educate families, patients, and students.

Examples:

SL is used to fabricate different structural 3D parts used in different areas as shown Fig. 20:

It is particularly suited for the fabrication of complex 3D microdevices such as resonators and RF filters used in high frequency operation circuits, due the fabrication precision and spatial resolution of the technique [71]. SLA is widely used in biomedicine for bone scaffold tissue engineering and the fabrication of biocompatible implants used as bone substitute in the human body (3DCeram, Limoges, France) (Fig. 21).

3.5.5. Advantages

One of the appealing aspects about SL is that a functional part can be created within one day [72]. The length of time it takes to produce any one part depends on the size and complexity of the project and can take anywhere from a
few hours to more than a day. Most SL machines can produce parts with a maximum size of approximately 50 cm x 50 cm x 60 cm (20" x 20" x 24") and some are capable of producing parts of more than 2 m in a single piece (the Mammoth has a build platform of 210 cm x 70 cm x 80 cm). Prototypes made by stereolithography can be very beneficial as they are strong enough to be machined and can be used as master patterns for injection molding, thermoforming, blow molding, and also in various metal casting processes. Moreover this technology has one of the best surface finishes amongst RP technologies.

3.5.6 Limitations

The main disadvantage of stereolithography is that the process is often expensive—the photocurable resin costs anywhere from $80 to $210 per liter. The parts made by stereolithography often require post-curing and post-processing [73]. The last one includes the removal of the support and residual uncured suspension. This could damage the part and is time consuming.

3.6. Microstereolithography (μSLA)

3.6.1. History

Microstereolithography, a technique derived from the stereolithography process, has emerged as a new technique for the microfabrication of complex three-dimensional parts, based on layer by layer polymerization of a photosensitive resin [74]. This technique has significantly improved both lateral and vertical resolution of stereolithography. As the fabrication method of this technique relies on the space-resolved and light induced polymerization of a resin, the improvement of its resolution is mostly related to the reduction of the interaction volume between light and matter. This objects made are destined in the fields of microrobotics, microfluidics or microsystems. Although most research teams involved in the microstereolithography field have investigated its use as a microfabrication technique, the most promising application field of this technology is nevertheless the RP domain that faces an increasing demand of small-size high-resolution prototype parts. This unique technology provides building of complex microstructures used in electromechanical systems (MEMS). In principle, this technique uses an UV light beam (365 nm) that scans the surface of a digital mask (DMD) and by reflection, comes into contact with the surface of a photosensitive resin and induces a process of photopolymerization.

3.6.2. Principle

The fabrication process of a part by microstereolithography firstly consists in the creation of a CAD
file of the object to build. Then the CAD file is converted to STL file. This one contains the closed surfaces of the 3D model of the object, broken down into elementary triangles. Then the object is digitally sliced, allowing defining the forms for each layer to cure. The final file (with the process parameters—beam intensity, scanning speed) is finally sent to the automated machine to begin the process technology.

The high dimensional accuracy of this technique has been achieved using a dynamic mask composed of a matrix of \( 1024 \times 768 \) micromirrors (Digital Micromirror device, Texas Instruments) with a size of 14 microns \([75]\). Each individual micromirror can be rotated \( \pm 12^\circ \). The source of radiation consists of a UV light source (Hamamatsu) including an optical waveguide and a 365 nm filter. An integrated shutter allows the control of the exposure time and the projected energy density. The beam scans the surface of the DMD onto which is loaded the section of the object to build. The reflected image of the object passes through an optical system comprising a focusing lens that provides a transmission above 83%, and is then projected onto the work surface onto which a photosensitive resin is spread out. For a magnification of 1, the dimensions of the object vary between 5 mm and 30 \( \mu \)m (lateral resolution of the system). This is due to the light scattering contribution of the ceramic particles in a curable suspension. A spreading and deposit systems control the uniformity and thickness of the thin layers. The resolution of the technique (at SPCTS laboratory, Limoges, France) is determined by the size of the micromirrors (14 \( \mu \)m). There are two types of microstereolithography processes: laser and integral. Integral microstereolithography processes (also named projection microstereolithography) are based on the projection of an image created by a dynamic mask on the surface of the photosensitive resin, and consequently, the object to be built is described by a series of black and white bitmap files rather than by vectors as is the case in scanning methods. These bitmap files are then used to shape the light beam using a dynamic mask, such that the image of each layer can be projected on the surface of the liquid resin after being reduced and focused by an appropriate optical system. The projected image induces a solidification of the irradiated areas, and a shutter controls the duration of the irradiation step. The superposition of many layers of different shapes allows the fabrication of a complex object in a similar way as for conventional RP techniques.

Integral microstereolithography processes are inherently faster than scanning techniques, as the irradiation of a complete layer is done in one-step, regardless to its complexity. Additionally, the light flux density on the surface of the photopolymerizable resin is much smaller when projecting an image than when focusing a light beam in one point, which allows avoiding problems related to unwanted thermal polymerization.

The schematic diagram of integral microstereolithography (SPCTS Laboratory, Limoges, France) is shown on Fig. 22.

- **Digital Micromirror Device (DMD)**

The Digital Micromirror Device made by Texas Instruments, is widely used in video projection applications and has been used as dynamic mask in integral microstereolithography machines. This component is in fact an array of micromirrors actuated by electrostatic forces, and is used as a light switch in the \( \mu \)SLA \([76]\). Each 14 \( \mu \)m square mirror can be independently actuated either to reflect the incident light beam either into or out of the pupil of the optical system, such that the corresponding pixel of the projected image appears bright or dark. For video projection applications, gray levels and colored images can be produced by combining the rapid on/off movement of the mirrors with a rotating color wheel, but these features are not used in the \( \mu \)SLA apparatus. The first microstereolithography machine using the DMD component as dynamic mask was developed by Bertsch et al. and used a metal halide lamp combined with optical filters to select a band of visible wavelength for the irradiation of the resin. Different photopolymerizable resins were developed for this apparatus. In a first step, a high resolution resin reacting at 530 nm was used.

It allowed the manufacture many components, but showed important curl deformations. In a second step, resins reacting at 410 nm were developed. Their composition is a lot closer to conventional acrylate-based stereolithography resins, even if they do not react with UV light, but in the blue part of the visible spectrum.

A second machine was later built by the same research team with an improved resolution (XGA: \( 1024 \times 768 \) pixels), and an irradiation in the UV. An acrylate-based resin was formulated specifically for this machine, which was used in an industrial context to produce high-resolution prototypes.

### 3.6.3. Materials

The fabrication sequence, materials used and photopolymerization process are the same as in stereolithography ones. The curable suspension contains a reactive oligomer/monomer, a photoinitiator and a ceramic powder in concentration above 50 vol% in order to obtain a high density and good mechanical properties of the sintered part. Here, the viscosity of the suspensions should not exceed 5 Pa.s. Therefore it is possible to improve the quality and resolution of the polymerized layers by adding a high viscosity oligomer such as amine modified polyether acrylates.

### 3.6.4. Applications

As the market for miniaturized products grows rapidly, there is also an increasing need for high-resolution
prototype parts. When small size objects have to be built with dimensions of only a few millimeters or less, current RP technologies are limited with respect to the feasibility of small features: openings and small holes are difficult to make and have to be cleaned once the prototype is built, which is particularly difficult. Another aspect, which has to be taken into consideration for small components, is that manual surface finishing can be of a great challenge.

Microstereolithography is used for the manufacturing of small size high-resolution components [77]. There is a growing demand of high-resolution prototype objects in particular in the medico-technical industry. For example, medical devices in which optical and chemical sensors could be embedded require very high precision and definition that microstereolithography has to offer.

Hearing-aid manufacturers try to design lightweight products small enough not to be detected, comfortable, with rounded shapes to be close from the natural geometry of the ear canal. For such applications it is necessary to prototype small mechanical components with very specific details.

Most 3D models made by microstereolithography can also be used as a simulation tool for medical teams before exercising an operation.

In the electro-technical industry, microstereolithography can be used for the prototyping of small connectors, and other microdevices.

This technique has also been used for prototyping MEMS components and optimizing their geometry, before investing in photomasks and using more conventional technologies for mass production. In the future, the manufacturing of hybrid polymer structures by combining various types of polymers, such as conductive polymers, polymers of various refractive index or flexible polymers could also lead to the creation of new optical, chemical and biochemical microsystems with the microstereolithography technology.

Example:

An example of a complex 3D resonator microstructure produced by the process of microstereolithography is shown in Fig. 23 [78]. This resonant structure is composed of three different parts. The first part represents stacks of rods with a missing row in the middle. The other two parts are metallized alumina plates and placed above and below the stacks of rods, forming a sandwich structure. The dimensions of the cavity created are 1.33 mm × 1.33 mm × 0.65 mm and operate between 140–145 GHz in...
the fundamental mode TE101 (the electric field is polarized along the axis of the rods).

This low-loss alumina compact structure can be easily integrated into microwave circuits for signal filtering, while providing a relatively high quality factor (2500) and a reflection coefficient ≈ 100%.

3.6.5. Advantages

Unlike stereolithography, microstereolithography uses an integral irradiation of the layer section to be built thus creating less fabrication defects (the density of energy is evenly distributed along the surface of the represented section) [79]. By operating with higher volume rate of oligomer in the compositions, it is possible to improve the definition of the created objects and to reduce the lateral overcure phenomenon.

3.6.6. Limitations

One of the major limitations of the microstereolithography technology is related to the materials that can be used in this manufacturing technique: only a few polymers can be used, acrylates in general or eventually epoxies. Because of their three-dimensional geometry, most objects produced by microstereolithography cannot be molded, which implies that, for some applications, microstereolithography is no longer a RP technology but a manufacturing technique [80]. In this case, the produced objects need to have adequate mechanical, chemical or physical characteristics. As a result, studies on the use of new materials for microstereolithography have been started, and in particular on the use of composite materials made of ceramic particles embedded in a polymer matrix as reactive medium in this technology.

The ultimate goal is the production of microcomponents in ceramic materials, which can be obtained by sintering the composite. However, for sintering successfully such composite components, the load of particles embedded in the resin has to be sufficiently high, which increases significantly the viscosity of the chemical media.

4. DATA FILE FORMATS

4.1. STL File Format

The natural file format for RP techniques such as stereolithography would be a series of closed polygons corresponding to different Z-values. However, since it’s possible to vary the layer thicknesses for a faster though less precise build, it seemed easier to define the model to be built as a closed polyhedron that could be sliced at the necessary horizontal levels [81]. The STL file format is capable of defining a polyhedron with any polygonal facet, but in practice it’s only ever used for triangles, which means that much of the syntax of the ASCII protocol is superfluous. STL files are supposed to be closed and connected like a combinatorial surface, where every edge is part of exactly two triangles, and not self-intersecting. Since the syntax does not enforce this property, it can be ignored for applications where the closeness doesn’t matter. The closeness only matters insofar as the software which slices the triangles requires it to ensure that the resulting 2D polygons are closed. Sometimes such software can be written to clean up small discrepancies by moving endpoints of edges that are close together so that they coincide. STL is a file format native to the stereolithography CAD software created by 3D Systems. This file format is supported by many other software packages; it is widely used for RP and CAM. STL files describe only the surface geometry of a three-dimensional object without any representation of color, texture or other common CAD model attributes. The STL format specifies both ASCII and binary representations. Binary files are more common, since they are more compact. In 2011, ASTM replaced the STL format with the Additive Manufacturing File Format (AMF), which has native support for color, multiple materials, and constellations. An STL file describes a raw unstructured triangulated surface by the unit normal and vertices (ordered by the right-hand rule) of the triangles using a three-dimensional Cartesian coordinate system.
- **ASCII STL file.**

An ASCII STL file begins with the line:

```
solid name
```

where *name* is an optional string (though if *name* is omitted there must still be a space after solid). The file continues with any number of triangles, each represented as follows:

```
facet normal n_1 n_2 n_3
outerloop
vertex v1_1, v1_2, v1_3
vertex v2_1, v2_2, v2_3
vertex v3_1, v3_2, v3_3
endloop
eendfacet
```

where each *n* or *v* is a floating point number in sign-mantissa ‘e’-sign-exponent format, e.g. “-2.64800e-002”. The file concludes with:

```
endsolid name
```

The structure of the format suggests that other possibilities exist (e.g. facets with more than one ‘loop’, or loops with more than three vertices) but in practice, all facets are simple triangles. White space (spaces, tabs, new lines) may be used anywhere in the file except within numbers or words. The spaces between ‘facet’ and ‘normal’ and between ‘outer’ and ‘loop’ are required.

- **Binary STL file.**

Because ASCII STL files can become very large, a binary version of STL exists. A binary STL file has an 80 character header (which is generally ignored—but which should never begin with ‘solid’ because that will lead most software to assume that this is an ASCII STL file). Following the header is a 4 byte unsigned integer indicating the number of triangular facets in the file. Following that is data describing each triangle in turn. The file simply ends after the last triangle. Each triangle is described by twelve 32-bit-floating point numbers: three for the normal and then three for the X/Y/Z coordinate of each vertex—just as with the ASCII version of STL. After the twelve floats there is a two byte unsigned ‘short’ integer that is the ‘attribute byte count’—in the standard format, this should be zero because most software does not understand anything else.

```
UINT8 [80] — Header
UINT32 — Number of triangles
foreach triangle
REAL32 [3] — Normal vector
REAL32 [3] — Vertex 1
REAL32 [3] — Vertex 2
REAL32 [3] — Vertex 3
UINT16 — Attribute byte count
end
```

There are at least two variations on the binary STL format for adding color information:

- The VisCAM and SolidView software packages use the two ‘attribute byte count’ bytes at the end of every triangle to store a 15 bit RGB color:
  - bit 0—4 are the intensity level for blue (0—31)
  - bits 5—9 are the intensity level for green (0—31)
  - bits 10—14 are the intensity level for red (0—31)
  - bit 15 is 1 if the color is valid
  - bit 15 is 0 if the color is not valid (as with normal STL files)

- The Materialize Magics software does things a little differently. It uses the 80 byte header at the top of the file to represent the overall color of the entire part. If color is used, then somewhere in the header should be the ASCII string “COLOR=” followed by four bytes representing red, green, blue and alpha channel (transparency) in the range 0—255. This is the color of the entire object unless overridden at each facet. Magics also recognize a material description; a more detailed surface characteristic. Just after “COLOR = RGBA” specification should be another ASCII string, “MATERIAL=” followed by three colors (3 x 4 bytes): first is a color of diffuse reflection, second is a color of specular highlight, and third is an ambient light. Material setting wags are preferred over color. The per-facet color is represented in the two ‘attribute byte count’ bytes as follows:
  - bit 0—4 are the intensity level for red (0—31)
  - bits 5—9 are the intensity level for green (0—31)
  - bits 10—14 are the intensity level for blue (0—31)
  - bit 15 is 0 if this facet has its own unique color
  - bit 15 is 1 if the per-object color is to be used

The red/green/blue ordering within those two bytes is reversed in these two approaches—so whilst these formats could easily have been compatible the reversal of the order of the colors means that they are not—and worse still, a generic STL file reader cannot automatically distinguish between them. There is also no way to have facets be selectively transparent because there is no per-facet alpha value—although in the context of current RP machinery, this is not important.

In both ASCII and binary versions of STL, the facet normal should be a unit vector pointing outwards from the solid object. In most software this may be set to (0,0,0) and the software will automatically calculate a normal based on the order of the triangle vertices using the ‘right-hand rule’. Some STL loaders check that the normal in the file agrees with the normal they calculate using the right-hand rule and warn you when it does not. Other software may ignore the facet normal entirely and use only the right-hand rule. Although it is rare to specify a normal that cannot be
calculated using the right-hand rule, in order to be entirely portable, a file should both provide the facet normal and order the vertices appropriately. A notable exception is SolidWorks which uses the normal for shading effects.

4.2. Tessellated Model

Tessellation is the process of creating a two-dimensional plane using the repetition of a geometric shape with no overlaps and no gaps. Generalizations to higher dimensions are also possible. Tessellations frequently appeared in the art of M. C. Escher, who was inspired by studying the Moorish use of symmetry in the Alhambra tiles during a visit in 1922. Tessellations are seen throughout art history, from ancient architecture to modern art. In Latin, *tessella* is a small cubical piece of clay, stone or glass used to make mosaics [82]. The word “tessella” means “small square” (from “tessera”, square, which in its turn is from the Greek word for “four”). It corresponds with the everyday term *tiling* which refers to applications of tessellations, often made of glazed clay. Examples of tessellations in the real world include honeycombs and pavement tiling.

In the subject of computer graphics, tessellation techniques are often used to manage datasets of polygons and divide them into suitable structures for rendering. Normally, at least for real-time rendering, the data is tessellated into a triangle, which is sometimes referred to as triangulation. In computer-aided design the constructed design is represented by a boundary representation topological model, where analytical 3D surfaces and curves, limited to faces and edges constitute a continuous boundary of a 3D body. Arbitrary 3D bodies are often too complicated to analyze directly. So they are approximated (tessellated) with a mesh of small, easy-to-analyze pieces of 3D volume—usually either irregular tetrahedrons, or irregular hexahedrons. The mesh is used for finite element analysis as shown on the Fig. 24.

The mesh of a surface is usually generated per individual faces and edges (approximated to polylines) so that original limit vertices are included into mesh. To ensure that approximation of the original surface suits the needs of the further processing, three basic parameters are usually defined for the surface mesh generator:

- The maximum allowed distance between the planar approximation polygon and the surface (aka “sag”). This parameter ensures that mesh is similar enough to the original analytical surface (or the polyline is similar to the original curve).
- The maximum allowed size of the approximation polygon (for triangulations it can be maximum allowed length of triangle sides). This parameter ensures enough detail for further analysis.

Algorithm generating mesh is driven by the parameters. Some computer analyses require adaptive mesh, which is made finer (using stronger parameters), in regions where the analysis needs more detail.

4.3. Other File Formats

4.3.1. IGES Format

IGES (Initial Graphics Exchange Specification) is a format used to exchange graphics information between commercial CAD systems. It was setup as an American National Standard in 1981. The IGES file can precisely represent CAD models [83]. It includes not only the geometry information (Parameter Data Section) but also topological information (Directory Entry Section). In the IGES, surface modeling, constructive solid geometry and boundary representation are introduced. Especially, the ways of representing the regularized operations for union, intersection, and difference have also been defined. The advantages of the IGES standard are its wide adoption and comprehensive coverage. Since IGES was setup as American National Standard, virtually every commercial CAD/CAM system has adopted IGES implementations. Furthermore, it provides the entities of points, lines, arcs, splines, NURBS surfaces and solid elements. IGES is a generally used data transfer medium which interfaces with various CAD systems. Advantages of using IGES over current approximate methods include
precise geometry representations, few data conversions, smaller data files and simpler control strategies. However, the problems are the lack of transfer standards for a variety of CAD systems and system complexities.

4.3.2. HPGL Format

HPGL, sometimes hyphenated as HP-GL, was the primary printer control language used by Hewlett-Packard plotters. The name is a start for Hewlett-Packard Graphics Language. It later became a standard for almost all plotters [84]. Hewlett-Packard’s printers also usually support HPGL in addition to PCL. The language is formed from a series of two letter codes, followed by optional parameters. For instance an arc can be drawn on a page by sending the string:

```
AA100, 100, 50;
```

This means Arc Absolute, and the parameters place the center of the arc at absolute coordinates 100,100 on the page, with a starting angle of 50° measured counterclockwise. A fourth optional parameter (not used here) specifies how far the arc continues, and defaults to 5°. Typical HPGL files start with a few setup commands, followed by a long string of graphics commands. The original HP/GL-Language did not support definition of line width, as this parameter was determined by the pens loaded into the plotter. With the advent of the first inkjet plotters, line width for the “pens” specified within the HP/GL-files had to be set at the printer so it would know what line width to print for each pen, a cumbersome and error-prone process. With HP/GL-2, definition of line width was introduced into the language and allowed for elimination of this step. Also, among other improvements a binary file format was defined that allowed for smaller files and shorter file transfer times, and the minimal resolution was reduced.

4.3.3. SLC File

The SLC file extension is a CAD slice file that is used with CAD software such as Rhinoceros 3D. A file in the .slc file extension can be created using the Model works command in Rhinoceros 3D and sent to the Rhino model [85]. The software has an edit box for slice files, from which you can also enter the name for the file in the .slc file extension that you want stored. This can be sent to other 3D modeling machines such as Solidscape. The thickness of CAD slice files in the .slc file extension can be edited in Rhinoceros 3D, using a drop down menu that allow you to set the thickness of your build. Files in the .slc file extension can be generated from slice created by meshes from a NURBS model. NURBS is also known as Nonuniform rational B-spline and is a mathematical model, which is utilized in computer graphics for the representation and generation of surfaces and curves. NURBS represent in compact form geometrical shapes, and can be handled efficiently by programs while still allowing human interaction. The SLC file extension is also used by SolidTools for its solid imaging. This RP component uses a database in the translation of 3D geometry, to parts or physical models that uses a variety of materials and resins. Files in the .slc file extension are the format used for output files by StudioTools. A NURBS surface is translated first to a file in the .slc file extension, before it can be read by the software for the solid imaging machine. Furthermore, geometry from StudioTools can also be converted to a file in the .slc file extension. Files in this format are able to cut contours in 2D of the three-dimensional base. An advantage to using the SLC file extension is that the geometry description of NURBS in StudioTools is sliced directly, and so less iteration is needed between the primary geometry and the sent data to be built in the Solid Imaging machine.

4.3.4. LEAF File

The LEAF or Layer Exchange ASCII Format is generated by Helsinki University of Technology. To describe this data model, concepts from the object-oriented paradigm are borrowed. At the top level, there is an object called LMT-file (Layer Manufacture Technology file) that can contain parts which in turn are composed of other parts or by layers [86]. Ultimately, layers are composed of 2D primitives and currently the only ones which are planned for implementation are polylines. For example, an object of a given class is created. The object classes are organized in a simple tree. Attached to each object class is a collection of properties. A particular instance of an object specifies the values for each property. Objects inherit properties from their parents. In LEAF, the geometry of an object is simply one among several other properties.

5. APPLICATIONS OF RP TECHNIQUES

RP is widely used in the automotive, aerospace, medical, and consumer products industries. Although the possible applications are virtually limitless, nearly all fall into one of the following categories: prototyping, rapid tooling, or rapid manufacturing. By exchanging prototypes early in the design stage, manufacturing can start tooling up for production while the art division starts planning the packaging, all before the design is finalized. Prototypes are also useful for testing a design, to see if it performs as desired or needs improvement. Engineers have always tested prototypes, but RP expands their capabilities. First, it is now easy to perform iterative testing: build a prototype, test it, redesign, build and test, etc. Such an approach would be far
too time-consuming using traditional prototyping techniques, but it is easy using RP [87].

In addition to being fast, RP models can do a few things metal prototypes cannot. For example, Porsche used a transparent stereolithography model of the 911 GTI transmission housing to visually study oil flow. Snecma, a French turbomachinery producer, performed photoelastic stress analysis on a SLA model of a fan wheel to determine stresses in the blades.

5.1. Rapid Tooling

A much-anticipated application of RP is rapid tooling, the automatic fabrication of production quality machine tools. Tooling is one of the slowest and most expensive steps in the manufacturing process, because of the extremely high quality required. Tools often have complex geometries, yet must be dimensionally accurate to within a hundredth of a millimeter. In addition, tools must be hard, wear-resistant, and have very low surface roughness (about 0.5 μm root mean square). To meet these requirements, molds and dies are traditionally made by CNC-machining, electro-discharge machining, or by hand. All are expensive and time consuming, so manufacturers try to incorporate RP techniques to speed the process. Peter Hilton, president of Technology Strategy Consulting in Concord, MA, believes that “tooling costs and development times can be reduced by 75% or more” by using rapid tooling and related technologies. Rapid tooling can be divided into two categories, indirect and direct [88].

5.1.1. Indirect Tooling

Most rapid tooling today is indirect: RP parts are used as patterns for making molds and dies. RP models can be indirectly used in a number of manufacturing processes:

- **Vacuum Casting**: in the simplest and oldest rapid tooling technique, a RP positive pattern is suspended in a vat of liquid silicone or room temperature vulcanizing rubber. When the rubber hardens, it is cut into two halves and the RP pattern is removed. The resulting rubber mold can be used to cast up to 20 polyurethane replicas of the original RP pattern. A more useful variant, known as the Keltool powder metal sintering process, uses the rubber molds to produce metal tools. Developed by 3 M and now owned by 3D Systems, the Keltool process involves filling the rubber molds with powdered tool steel and epoxy binder. When the binder cures, the “green” metal tool is removed from the rubber mold and then sintered. At this stage the metal is only 70% dense, so it is infiltrated with copper to bring it close to its theoretical maximum density. The tools have fairly good accuracy, but their size is limited to less than 25 centimeters.

- **Sand Casting**: a RP model is used as the positive pattern around which the sand mold is built. LOM models, which resemble the wooden models traditionally used for this purpose, are often used. If sealed and finished, a LOM pattern can produce about 100 sand molds.

- **Investment Casting**: some RP prototypes can be used as investment casting patterns. The pattern must not expand when heated, or it will crack the ceramic shell during autoclaving. Both Stratasys and Cubital make investment casting wax for their machines. Paper LOM prototypes may also be used, as they are dimensionally stable with temperature. The paper shells burn out, leaving some ash to be removed. To counter thermal expansion in stereolithography parts, 3D Systems introduced QuickCast, a build style featuring a solid outer skin and mostly hollow inner structure. The part collapses inward when heated. Likewise, DTM sells Trueform polymer, a porous substance that expands little with temperature rise, for use in its SLS machines.

- **Injection molding**: CEMCOM Research Associates, Inc. has developed the NCC Tooling System to make metal/ceramic composite molds for the injection molding of plastics. First, a stereolithography machine is used to make a match-plate positive pattern of the desired molding. To form the mold, the SLA pattern is plated with nickel, which is then reinforced with a stiff ceramic material. The two mold halves are separated to remove the pattern, leaving a matched die set that can produce tens of thousands of injection moldings.

5.1.2. Direct Tooling

To directly make hard tooling from CAD data is the biggest achievement of rapid tooling. Realization of this objective is still several years away, but some strong strides are being made:

- **RapidTool**: a DTM process that selectively sinters polymer-coated steel pellets together to produce a metal mold. The mold is then placed in a furnace where the polymer binder is burned off and the part is infiltrated with copper (as in the Keltool process). The resulting mold can produce up to 50,000 injection moldings.

In 1996 Rubbermaid produced 30,000 plastic desk organizers from a SLS-built mold. This was the first widely sold consumer product to be produced from direct rapid tooling.

- **LENS**: a process developed at Sandia National Laboratories and Stanford University that can create metal tools from CAD data. Materials include 316 stainless steel, Inconel 625, H13 tool steel, tungsten, and titanium carbide cermets. A laser beam melts the top layer of the part in areas where material is to be added. Powder metal is injected into the molten pool, which
then solidifies. Layer after layer is added until the part is complete. Unlike traditional powder metal processing, LENS produces fully dense parts, since the metal is melted, not merely sintered. The resulting parts have exceptional mechanical properties, but the process currently works only for parts with simple, uniform cross sections. The system has been commercialized by MTS corporation.

- **Direct AIM (ACES Injection Molding):** a technique from 3D Systems in which stereolithography-produced cores are used with traditional metal molds for injection molding of high and low density polyethylene, polystyrene, polypropylene and ABS plastic. Very good accuracy is achieved for about 200 moldings. Long cycle times (~5 min) are required to allow the molding to cool enough that it will not stick to the SLA core. In another variation, cores are made from thin SLA shells filled with epoxy and aluminum shot. Aluminum’s high conductivity helps the molding cool faster, thus shortening cycle time. The outer surface can also be plated with metal to improve wear resistance. Production runs of 1000–5000 moldings are envisioned to make the process economically viable.

- **LOM Composite:** Helisys and the University of Dayton are working to develop ceramic composite materials for LOM. LOM Composite parts would be very strong and durable, and could be used as tooling in a variety of manufacturing processes.

- **Sand Molding:** at least two RP techniques can construct sand molds directly from CAD data. DTM sells sand-like material that can be sintered into molds. Soligen http://www.3dprinting.com/uses 3DP to produce ceramic molds and cores for investment casting. (Direct Shell Production Casting).

### 5.2. Rapid Manufacturing

A natural extension of RP is rapid manufacturing (RM), the automated production of market products directly from CAD data [89]. Currently only a few final products are produced by RP machines, but the number will increase as metals and other materials become more widely available. RM will never completely replace other manufacturing techniques, especially in large production runs where mass-production is more economical. For short production runs, however, RM is much cheaper, since it does not require tooling. It is also ideal for producing custom parts tailored to the user’s exact specifications. A University of Delaware research project uses a digitized 3D model of a person’s head to construct a custom-fitted helmet. NASA is experimenting using RP machines to produce spacesuit gloves fitted to each astronaut’s hands. From tailored golf club grips to custom dinnerware, the possibilities are endless.

The other major use of RM is for products that simply cannot be made by subtractive (machining, grinding) or compressive (forging, etc.) processes. This includes objects with complex features, internal voids, and layered structures.

### 5.3. Future Developments

RP is starting to change the way companies design and build products. There are several developments that improve manufacturing. One such improvement is increased speed [90]. RP machines are still slow by some standards. By using faster computers, more complex control systems, and improved materials, RP manufacturers are dramatically reducing build time. For example, Stratasys (since January 1998) has introduced FDM Quantum machine, which can produce ABS plastic models 2.5–5 times faster than previous FDM machines. Continued reductions in build time make it possible rapid manufacturing economical for a wider variety of products.

Another future development is improved accuracy and surface finish. Today’s commercially available machines are accurate to ~0.08 mm in the x–y plane, but less in the z (vertical) direction. Improvements in laser optics and motor control should increase accuracy in all three directions. In addition, RP companies are developing new polymers with better mechanical properties and. The introduction of nonpolymeric materials, including metals, ceramics, and composites, represents another much anticipated development. These materials allow RP users to produce functional parts. Today’s plastic prototypes work well for visualization and fit tests, but they are often too weak for function testing. More rugged materials would yield prototypes that could be subjected to actual service conditions. In addition, metal and composite materials will greatly expand the range of products that can be made by rapid manufacturing.

Many RP companies and research labs are developing new materials. For example, the University of Dayton is working with Helisys to produce ceramic matrix composites by LOM. An Advanced Research Projects Agency/Office of Naval Research sponsored project is investigating ways to make ceramics using FDM. Sandia/Stanford’s LENS system can create solid metal parts.

Another important development is increased size capacity. Currently most RP machines are limited to objects 0.125 m³ or less. Larger parts must be built in sections and joined by hand. To remedy this situation, several “large prototype” techniques are in the works. The most fully developed is Topographic Shell Fabrication from Formus in San Jose, CA. In this process, a temporary mold is built from layers of silica powder (high quality sand) bound together with paraffin wax. The mold is then used to produce fiberglass, epoxy, foam, or concrete models up to 3.3 m × 2 m × 1.2 m in size.
At the University of Utah, a team is developing systems to cut intricate shapes into 1.2 m x 2.4 m sections of foam or paper. Researchers at Penn State’s Applied Research Lab are aiming even higher: to directly build large metal parts such as tank turrets using robotically guided lasers. Group leader Henry Watson states that product size is limited only by the size of the robot holding the laser. All the above improvements are the key to the development of the RP industry worldwide.

One future application is Distance Manufacturing on Demand, a combination of RP and the Internet that will allow designers to remotely submit designs for immediate manufacture. Finally, the rise of RP has spurred progress in traditional subtractive methods as well. Advances in computerized path planning, numeric control, and machine dynamics have increased the speed and accuracy of machining. Modern CNC machining centers can have spindle speeds of up to 100,000 rpm, with correspondingly fast feed rates. Such high material removal rates translate into short build times. For certain applications, particularly metals, machining will continue to be a useful manufacturing process.

REFERENCES

Chapter 6.5 Rapid Prototyping of Ceramics


Chapter 7.1
Biomorphous Ceramics from Lignocellulosic Preforms

Peter Greil, Tobias Fey and Cordt Zollfrank
Department of Materials Science and Engineering (Glass and Ceramics), University of Erlangen-Nuernberg, Martensstrasse 5, 91058 Erlangen, Germany

Chapter Outline
1. Introduction 527
2. Lignocellulosic Cellular Preforms 528
3. Processing 531
4. Microstructure and Mechanical Properties 541
5. Applications 548
6. Conclusions 551
References 552

1. INTRODUCTION

Inspired by biological concepts of design, microstructure, and property optimization, new biomorphous inorganic materials with advanced functions and structures have attained increasing interest in materials science at the frontier between biology and chemistry [1,2]. The driving force to develop novel approaches for producing synthetic materials by use of biological processes and principles is for a great need for materials of enhanced performance and reliability, which provide better efficiency, specialization, and optimization of their properties [3,4]. Inorganic materials synthesized by biologically controlled growth processes are hierarchically organized microcomposite materials, which are characterized by an inorganic biomineral phase and a biopolymer phase. Prominent biominerals include the following: carbonates in mollusc shells, nacre, enamel, and corals; phosphates in bone and teeth; iron oxides in magnetotactic bacteria; or silica in diatoms and sponges [5,6]. Almost 70 different mineral types are known to be formed by living organisms. A bioorganic phase including various proteins and polysaccharides is a key element needed to control the directed growth of the biomineral in time and space (hierarchical and cellular structure anatomy) and morphogenesis (crystal habitus). Since biogenic inorganic composites materials are synthesized at mild conditions (ambient temperature), their limited temperature stability only allows applications at low temperatures due to the included bioorganic phase. Advanced functional and engineering materials, however, are often subjected to harsh environments such as elevated temperature, corrosion, and high mechanical loading. Thus, even small fractions of organic constituents that often form thin interface layers between the inorganic particles may limit the application of these materials at harsh environmental and application conditions. Furthermore, the low synthesis temperatures and often low concentrations of reaction solutions result in very slow (μm/h) growth rates of the biomineral so that an economical manufacturing process of bulk components could not be realized yet. Therefore, only low-dimensional products such as magnetic powders, size- and shape-specific silica for chromatography, and SiC whiskers from rice husks for ceramic cutting tools were fabricated using biological preform materials [7,8].

A significant increase of synthesis rate can be achieved when cellular biological structures with open porosity are used as a microstructural template for physicochemical infiltration and reaction techniques [9]. Thus, conversion of cellular tissue of naturally grown plants into inorganic materials (biomorphous ceramics) has gained an increasing interest in recent years [10–20]. Basic principles of conversion of plant-derived preforms (Lignocellulosics) into ceramic materials mimicking the initial template structure at various hierarchical micro- and macrostructural levels were discussed for example in [14]. Carbon derived from the natural cellular preform (biocarbon) may also serve as a template for a variety of infiltration and reaction processes. A comprehensive survey of the transformation
of Lignocellulosics into biocarbon can be found in [21,22]. Furthermore, plant tissue-derived products such as fiber paper and fiber boards may serve as preforms with a more uniform microstructure compared with naturally grown tissue, which may vary with seasonal and local growth conditions significantly (e.g. annual growth ring patterns, reaction tissue). A variety of biomorphous ceramics were fabricated from Lignocellulosic templates including fibers, wood tissue, and wood pulp paper and fiber board preforms, Table 1.

Mimicking the cellular design of natural tissue anatomy by material compositions relevant for functional and engineering applications (e.g. multicomponent oxides and nonoxides) offers a highly attractive approach for creating a novel class of biomorphous materials. The mechanical, thermal, or optical properties of the biomorphous ceramics are governed by the hierarchical cellular structure and the phase composition, which generally result in highly anisotropic materials. These materials have a high potential of achieving an excellent density-to-strength relationship and outstanding thermomechanical performance. In the following, a review on the current status of biomorphous ceramics processing, microstructure development, and property optimization will be given.

2. LIGNOCELLULOSIC CELLULAR PREFORMS

Lignocellulosics form the class of organic matter (biomass) produced by land-growing plants in the form of trees, shrubs, and agricultural crops. It is the essential carbon sink of the planet, which is formed by catalytic conversion of carbon dioxide to an organic mass mainly consisting of the elements C—O—H (—N, S, P) [14,38]. Lignocellulosic biomass is mainly composed of polysaccharides (cellulose, polyoses) and lignin biopolymers. A mature vascular plant contains several differentiated cell types grouped together in tissues. The tissues of a plant are organized into three tissue systems: the dermal, the ground, and the vascular tissue system, respectively. Phloem and xylem are the major cellular components of the vascular tissue system. The xylem conducts water and dissolved minerals from the roots to all the other parts of the plant. Two types of water-conducting cells of xylem are usually discerned: tracheids and vessel elements. Tracheids are individual cells tapered at each end, so the tapered end of one cell overlaps that of the adjacent cell. They have thick, lignified walls, which are perforated so that water can flow from one tracheid to the next. Vessels are thick-walled tubes that can extend vertically through several meters of xylem tissue. Their diameter may be as large as 0.7 mm. The inherent vascular transportation system provides a facile access to infiltration fluids for transforming biological template structures into inorganic materials.

Wood is a natural composite with partially crystalline cellulose fibrils embedded in an amorphous matrix of polyoses and lignin as the major biomolecular constituents. It exhibits a unique hierarchical cellular architecture with a remarkable combination of strength, stiffness, and toughness [39]. Wood cell walls provide skeletal support (mechanical stability), but they also play a dominant role in cell growth and morphogenesis, in cell recognition and signaling, in digestibility, and in herbivore nutrition. A single longitudinal tracheidal cell exhibits a layered wall structure, a thin primary wall (P) and a thicker secondary wall composed of sublayers (S1, S2), which vary in cellulose microfibril orientation and which play a key role for mechanical behavior of cellular tissue, Figure 1. Adjacent tracheids are joined together by a highly lignified layer (middle lamella, ML). Because of seasonal variations of growth conditions, growth ring patterns develop with low density/high porosity in the earlywood regions and high density/low porosity in the latewood regions.

Cellulose (C₆H₁₀O₅)n is the main load-bearing component in fibers. It is a linear polymer composed of

<table>
<thead>
<tr>
<th>TABLE 1 Examples of Biomorphous Ceramics Derived from Natural and Preprocessed Plant Preforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template</td>
</tr>
<tr>
<td>Fibers</td>
</tr>
<tr>
<td>Sisal</td>
</tr>
<tr>
<td>Cotton</td>
</tr>
<tr>
<td>Wood</td>
</tr>
<tr>
<td>Pine</td>
</tr>
<tr>
<td>Maple</td>
</tr>
<tr>
<td>Walnut</td>
</tr>
<tr>
<td>Beech</td>
</tr>
<tr>
<td>Oak</td>
</tr>
<tr>
<td>Balsa, Cypress</td>
</tr>
<tr>
<td>Other plants</td>
</tr>
<tr>
<td>Rattan</td>
</tr>
<tr>
<td>Sponge</td>
</tr>
<tr>
<td>Wood products</td>
</tr>
<tr>
<td>Wood fibers</td>
</tr>
<tr>
<td>Wood fiber boards</td>
</tr>
<tr>
<td>Fiber papers</td>
</tr>
</tbody>
</table>
glucose units joined together by β-1,4 glycosidic bonds. Cellulose is able to attain a high stiffness and tensile strength due to intra- and intermolecular hydrogen bonding. The degree of polymerization in a cellulose chain of a native softwood fiber is above 10,000 which is decreased after chemical pulping to 500–2000. The cellulose crystal is distinguished by a Young’s modulus of 145 GPa and a theoretical tensile strength estimated at 7500 MPa [40]. The orientation angle of the cellulose microfibrils with respect to the fiber cylinder axis therefore plays a crucial role in determining the mechanical properties of fibers.

2.1. Wood Preforms

In angiosperms (hardwoods), most of the water is conducted in the xylem vessels. Their walls are thickened with secondary deposits of cellulose and are usually further strengthened by impregnation with lignin. The secondary walls of the xylem vessels are deposited in spirals and rings and are usually perforated by pits. Xylem also contains tracheids. In ring-porous wood, such as oak and basswood, the spring vessels (earlywood) are much larger and more porous than the smaller, summer tracheids (latewood). This difference in cell size and density produces the conspicuous, concentric annual growth rings in these woods. Softwoods are gymnosperms, such as pine, spruce, and fir. Gymnosperms generally do not have vessels but produce only tracheids in their xylem. The tracheoidal tissue of gymnosperms shows a lower variation of cell pore sizes and hence a more uniform cellular microstructure compared with angiosperms. Table 2 summarizes some of wood templates used for processing of biomorphous ceramics.

The shape of vascular cells may vary from spherical (for example Meranti, Shorea spp.) over oval (walnut, Juglans regia.) to rectangular (pine — Pinus sylvestris), and the thickness of cell walls typically ranges from 4 µm (Okoumé, Aucoumea klaineana) to 18 µm (Ebony, Diospyros spp.). The density can be as low as 0.05 g/cm³ for balsa and as high as 1.03 g/cm³ for ebony. While small pores with diameters <10 µm (tracheids) may occur, the average cell diameter of vessels (tracheas) ranges from 50 to 100 µm in European wood and 150 to 200 µm in tropical wood species. Maximum vessel diameter of more than 350 µm (oak) are frequently observed.

Monocotyledons (monocots) are a subgroup of angiosperms (grasses, grains, and bamboo) in which the vascular tissue is scattered rather than arranged in concentric rings. Wood-like monocots such as palms (Arecaceae) and bamboos (e.g. Phyllostachys pubescens) show scattered vascular bundles embedded in tracheidal tissue. Rattan is the name for the roughly 600 species of climbing palms in the subfamily Calamoideae, native to tropical regions. Rattans (climbing palms) have the longest stems of any vascular plant, with some reported to be well over 100-m long. These climbing species also have the widest xylem vessels among palms with pore channel diameter ranging from 100 µm to more than 500 µm [42]. In dicotyledons (dicots), the xylem tissue appears like a 3-pronged or 4-pronged star. Storage parenchyma of Quiver tree (Aloe dichotoma) or Rush (Juncaceae), appears like a 3–6-pronged star, offering a high mechanical stiffness and stability at minimum weight (e.g. maximum porosity) (Figure 2).
<table>
<thead>
<tr>
<th></th>
<th>Ebony (Diospyros)</th>
<th>Beech (Fagus)</th>
<th>Oak (Quercus)</th>
<th>Maple (Acer)</th>
<th>Pine (Pinus)</th>
<th>Balsa (Ochroma pyramidate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porosity (%)</td>
<td>32</td>
<td>55</td>
<td>57</td>
<td>61</td>
<td>67</td>
<td>91</td>
</tr>
<tr>
<td>Density (g/cm³)</td>
<td>0.9–1.03</td>
<td>0.49–0.88</td>
<td>0.39–0.93</td>
<td>0.48–0.75</td>
<td>0.3–0.86</td>
<td>0.05–0.13</td>
</tr>
<tr>
<td>Composition (wt.%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cellulose</td>
<td>45</td>
<td>34–46</td>
<td>38–43</td>
<td>38</td>
<td>42–52</td>
<td>50–54</td>
</tr>
<tr>
<td>Hemicellulose</td>
<td>17</td>
<td>18–26</td>
<td>19–26</td>
<td>20</td>
<td>8–13</td>
<td>19</td>
</tr>
<tr>
<td>Cell types (vol.%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracheas</td>
<td>57</td>
<td>40</td>
<td>44</td>
<td>76</td>
<td>93</td>
<td>4</td>
</tr>
<tr>
<td>Tracheids</td>
<td>10</td>
<td>40</td>
<td>30–40</td>
<td>&lt;10</td>
<td>&gt;90</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Tracheas diameter (µm)</td>
<td>10–115</td>
<td>8–85</td>
<td>150–350</td>
<td>30–70</td>
<td>20–50</td>
<td>130–200</td>
</tr>
<tr>
<td>Drying shrinkage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axial</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.2–0.4</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Radial</td>
<td>8</td>
<td>6</td>
<td>4–5</td>
<td>3</td>
<td>3–4</td>
<td>2–3</td>
</tr>
<tr>
<td>Tangential</td>
<td>13</td>
<td>12</td>
<td>8–10</td>
<td>8</td>
<td>7–9</td>
<td>3–5</td>
</tr>
</tbody>
</table>

**FIGURE 2** Examples of plant tissue preforms used for synthesis of biomorphous ceramics.
3. PROCESSING
Conversion of the bioorganic template into inorganic material structure can be divided into direct replication of cellular anatomy and multiphase composite formation, Figure 3.

Direct replication involves a multistep process resulting in the mineralization of the three-dimensional template. Mineralization may be achieved by intercalation of the cell walls with liquid precursor systems, which may be an inorganic (metal salts) or a metal organic (metal alkoxides) sol. After chemical modification of the native cell wall structure (for example with maleic anhydride [43]) to facilitate penetration of the precursor sol into the cell wall, acid- or base-catalyzed hydrolysis reactions trigger gel formations via olation, oxolation, alcolation, or alcoxolation type of condensation reaction. Prior to using natural plant tissue for templating, the low-molecular weight extractives need to be removed. Otherwise, they may react with the infiltration solution, alter the pore structure, and decrease thermal stability of the template [44]. Extraction procedures generally include treating the wood pieces with organic solvents of varying polarity.

The conversion process follows two pathways, Figure 4:

1. pyrolysis of the biotemplate into biocarbon, which subsequently is infiltrated with gaseous or liquid reactants;
2. intercalation of the biopolymer cell walls with a metal organic liquid precursor and subsequent thermal treatment.

3.1. Biomorphous Carbon
Carbonaceous materials from wood have become of particular interest as cellular templates to fabricate hierarchically structured ceramic structures with anisotropic properties [14,45,46]. Pyrolysis of wood yields an anisotropic cellular carbon material where the cellular structure of the wood tissue can be retained during organic to inorganic transition, Figure 5. Despite a pronounced anisotropic shrinkage, the anatomical features of the wood tissue are replicated in the biomorphous carbon down to the submicron range. The cell walls of the tissue are converted into carbon struts. The resulting biomorphous carbon represents an open cellular structure with unidirectional pores corresponding to the former wood cells.

The molecular structure and chemical composition of the organic precursor directly affects the degree of alignment and growth of polyaromatic carbon stacks during carbon production [47]. A correlation between

FIGURE 3  Scheme of converting natural plant tissue into biomorphous ceramic.
oriented, highly crystalline cellulose (e.g. viscose rayon [48], lyocell [49], ramie [50]) and the resulting preferred orientation of the graphene layers and sheets (turbostratic carbon) is evident. This so-called *microfibril dominance theory* was first introduced by Tang and Bacon [48] and was discussed for the pyrolytic behavior of wood [11,45]. Additional evidence from X-ray diffraction [9,21,51] and Raman microscopy [52] shows that the turbostratic carbon layers were possibly oriented along the cell axis in pyrolyzed wood. The fate of the biopolymers within the wood cell wall and the resulting structural development of the carbonaceous material during pyrolysis, however, are still a matter of interest for engineering applications and subject of a continued scientific discussion [21,22,45,51–55].

During heating in inert atmosphere wood, pyrolysis causes a pronounced mass loss that is accompanied by an anisotropic shrinkage of the wood preform, Figure 6. Drying of wood is usually performed up to a maximum temperature of 200 °C, where only water evolves, and chemical and structural changes do not occur. Further thermal treatment of wood at temperatures from 200 to 350 °C is known as torrefaction, yielding wood materials with different properties than untreated wood due to chemical modification [56]. Torrefied wood can be considered as an intermediate between the original wood and charcoal [57]. Lignin is a complex, mostly amorphous polymer [58], with weak chemical bonds, which is degraded by pyrolysis starting at 260 °C and shows a smooth decay up to 500 °C, Figure 6a. The polyoses are the most thermally sensitive of the main wood biopolymers [38], which start degradation at 240 °C. They gave rise to most of the gaseous phases as well as the most important part of the obtained pyrolytic liquid (tar oils). The partly crystalline cellulose, which was reported to be stable under these temperatures, showed only little degradation up to 300 °C. At 350 °C, the degradation of cellulose is quite rapid and almost finished at 400 °C. The final carbon yield of the biopolymeric constituents of wood is 20 wt.% for
cellulose, 30 wt.% for the polyoses, and approximately 50 wt.% for the lignin. Similar to the mass loss, the dimensional changes of the specimens are predominately observed within the small temperature interval between about 250 and 350 °C. The thermal degradation curve of wood appears to be the sum of the degradations of the individual biopolymers; however, an influence on the pyrolytic behavior of the single biopolymer cannot be excluded. The dimensional shrinkage is strongly anisotropic in the axial (18%), radial (29%), and tangential (35%) directions.

The ultrastructure transformation of the wood biopolymers into biomorphous carbon is summarized in Figure 7. The principal structural and compositional changes during wood pyrolysis occur up to temperatures of 400 °C due to degradation, repolymerization, and carbonization of the wood biopolymers. At least seven different stages of the ultrastructural development during pyrolysis softwood cell wall can be described [18,21]:

(a) RT−200 °C: the original cell walls show nearly no ultrastructural changes up to temperatures of 200 °C. The cellulose microfibrils in the S2 cell wall are clearly visible as streaks in the Transelectronmicroscopy (TEM) micrograph. The cell-connecting compound ML is still intact, Figure 7a.

(b) 225 °C: depolymerization and degradation of the polyoses start around this temperature, and they are continuously degraded with increasing temperature. The resulting loss of material gave rise to consideration that the cellulose microfibrils were reoriented, taking over the free space. The ultrastructure of the cell walls (TEM) becomes indistinct with respect to the cellulose microfibril orientation Figure 7b. Simultaneously, to the decomposition of the polyoses, the amorphous fraction of the cellulose microfibrils are decayed. Until here, the lignin-rich cell substructures such as the CML remained nearly unchanged.

(c) 250−275 °C: at this stage, the decline of the cellulose crystallite size indicated a reduction of microfibril diameter in addition to a decrease of the microfibril length, Figure 7c [59]. This resulted in a complete loss of cellulose microfibril substructure, and the decomposition of the cellulose continued. The corresponding visible structures in the TEM micrograph appear blurred. As a consequence, an entire iso-orientation within the cell wall was observed. At the same time, the lignin-attributed substructure showed beginning of the decomposition until the CML was hardly detectable at 275 °C.

(d) 300−500 °C: proceeding degradation of the cell wall ultrastructure now leads to a substantial mass loss and dimensional changes. The crystal structure of cellulose [21] and the lignin substructures [22] were completely degenerated, and the former, structured wood cell appeared entirely disintegrated and fully random, Figure 7d. Therefore, the cell wall ultrastructural assembly was not retained within the developing carbon, which can be described as fully random, and without any visible structure. Once the principal degradation of the biopolymers and their carbonization are complete at around 500 °C, developments of the carbon structure will occur with increasing temperatures, finally yielding graphitized biomorphous carbon.

(e) 500−1000 °C: the carbon double bonds and aromatic rings were observed to form at a carbonization temperature of about 500 °C [60,61]. Density fluctuations and/or development of nanoporosity of increasing extension and amount, as well as beginning the stacking of two-dimensional ordered graphene sheets indicate the formation and successive
ordering of turbostratic carbon [21]. The corresponding Raman spectra indicate low ordering of the carbon, Figure 7e. The average number of stacked carbon sheets is three and remains constant, while their amount and their in-plane dimensions increase continuously up to 1000 °C. It was found that a slightly preferred orientation of the graphene sheets parallel to the cell axis redevelops, which improves with temperature.

(f) 1000—2000 °C: further condensation of aromatic rings proceeded resulting in an increased order of the graphene layers indicated by the Raman spectrum, Figure 7f. The average stack size of three-graphene layers was maintained, whereas the size of the graphene layers (lateral crystallite size) increases up to 4—6 nm corresponding to a stack of 13—20 layers. Carbonization up to 1800 °C resulted in partial graphitization. At the same time, oxygen-containing
functional groups diminished with the increase in carbonization temperature [61].

(g) 2000 °C and higher: at these temperatures, further lateral extension of the graphene layers occurs in the biomorphous carbon. The order of the carbon is increased, reflected by the sharp peaks in the corresponding Raman spectrum, Figure 7g. It could be shown that biomorphous carbon treated at 2200 °C is dominated by a microfibrillar pattern [62]. The lateral size of the graphitic crystallite size grows with increasing temperature. A layer size of 9 nm corresponding to a stack of 30 graphene sheets is observed at 3000 °C [63].

3.2. Reaction Infiltration into Biocarbon Templates (Silicon Carbide and Nonoxides)

Conversion of the biocarbon template into biomorphous SiC may be achieved by infiltrating a silicon-containing liquid (LSI) or vapor (LVI) precursor, which reacts with C to form SiC or a SiC-based composite material.

\[
\text{C} + M^{n+}(\text{OR})_n \rightarrow \text{MC}_{n/4} + \text{R}'' \\
\text{and} \quad \text{R}'' = \text{H}_2\text{O}, \quad \text{C}_6\text{H}_{2n}\text{OH}, \ldots
\]

\[
\text{C} + \text{Si}(1, g) \rightarrow \text{SiC/(Si)}
\]

\[
\text{2C} + \text{SiO(g)} \rightarrow \text{SiC} + \text{CO}
\]

The infiltration-reaction process involves long-distance transport of the Si-precursor through the vascular pore system (fluid flow and diffusion) interface reaction (nucleation and growth), and short-distance transport through the interface reaction product (grain boundary diffusion). Each of the processes may exhibit different rate dependence on temperature, pressure, and template geometry, which requires careful control of the process.

The long-range transport of liquid through the preform is governed by its capillarity and the resistance of the viscous and gravity forces. Based on Darcy’s model of unidirectional Newtonian fluid flow into porous media, the infiltration rate \( \frac{dh}{dt} \) (h is the infiltrated height and t the time) of an incompressible fluid is given by the modified Washburn model [64].

\[
\frac{dh}{dt} = \frac{\tau}{8\eta h} (p(r) - \rho g h) r(t)^2
\]

where \( \eta \) is the viscosity of the Si melt. For the case of spontaneous wetting, the infiltration is driven by the capillary pressure \( p(r) \) (\( p = 2\gamma \cos \theta / r \), where \( \gamma \) is the surface tension of the liquid, \( \theta \) is the wetting angle, and \( r \) is the capillary radius, respectively). Limitations to the infiltration process are imposed from the gravity \( g \), the density of the liquid \( \rho \), and the cellular channel geometry as expressed by the tortuosity factor \( \tau \) and the channel pore diameter \( r(t) \), which may change with time. Substituting the parameter values typical for Si-melt infiltration into C indicates that the second term (gravity) on the right is at least two orders of magnitude less than the capillary term at least for small values of \( h \) and is thus to be neglected [65]. Analysis of infiltration rate of liquid Si into porous carbon template was found to be reduced significantly when SiC was formed on the pore surface due a pronounced volume expansion of \( \Delta V/V_C = +58\% \) [66]. Coupling of transport and chemical reaction phenomena takes into account variation of pore radius due to formation of SiC on the pore surface [67].

\[
r(t) = r_0 - \sqrt{2D_{eff} \frac{M_C \rho_{SiC}}{M_{SiC} \rho_C} t}
\]

where \( r_0 \) is the pore radius in the biocarbon template, and \( M_i \) and \( \rho_i \) are the atomic weight and density of component \( i \), respectively. The effective diffusivity of material transport through the SiC reaction layer \( D_{eff} \) is given as \( D_{eff} = D_{eff}^{0} \exp (-E_D/RT) \) with \( D_{eff}^{0} = 2 \times 10^{-6} \text{cm}^2/\text{s} \), and the activation energy \( E_D = 132 \text{kJ/mol} \) [68]. Thus, integrating Eqn (2), an expression for the infiltration height in the reactive system was derived.

\[
h = \sqrt{\frac{2\gamma \cos \theta}{3 \eta} \left( r_0 t - \frac{d_{\text{strut}}}{3} \sqrt{2D_{eff} \frac{M_C \rho_{SiC}}{M_{SiC} \rho_C} t^3} \right)}
\]

Taking the physical properties for the system Si–C at an infiltration temperature of 1500 °C (\( \gamma = 0.725 \text{J/m}^2 \), \( \theta \approx 35^\circ \), \( \eta = 0.62 \text{mPa s} \) [69], \( d_{\text{strut}} = 10 \mu\text{m} \), \( M_{SiC} = 40 \text{g/mol} \), \( M_C = 12 \text{g/mol} \), \( \rho_{SiC} = 2.47 \text{g/cm}^3 \), \( \rho_C = 1.10 \text{g/cm}^3 \), \( \tau \approx 1 \)), the time for infiltration of a distance of \( h = 10 \text{ mm} \) was estimated, Figure 8.

Below \( r_{\text{min}} = 1 \mu\text{m} \), the infiltration rate is too slow to achieve reasonable manufacturing conditions (\( t > 1 \text{ min} \)) for 10 mm). It follows from Eqn (2) that for an initial pore radius, \( r_0 \), to strut thickness (\( d_{\text{strut}} \)) ratio, \( r_0/d_{\text{strut}} > 0.028 \), residual Si will always remain in pores larger than approximately 0.1–0.3 \( \mu\text{m} \) (for \( d_s \approx 5–10 \mu\text{m} \)), resulting in a two-phase SiC–Si composite material. Pores larger than \( r_{\text{max}} = 0.4 \text{ mm} \), however, will remain unfilled since gravity exceeds capillary suction. Though the complex three-dimensional cellular structure of biocarbon is characterized by a high degree of heterogeneity (variation of cell shape elliptical vs. rectangular, cell size, cell orientation longitudinal vs. radial, interconnectivity between cells), experimental observations generally confirm that pores below 1 \( \mu\text{m} \) could not be deeply infiltrated and pores with a diameter exceeding 100–300 \( \mu\text{m} \) remained free of residual Si. [9]. Depending on the template porosity and pore size distribution, the biomorphous SiC ceramics processed by liquid Si infiltration contained a residual Si content ranging from 20 wt.%
Instead of liquid infiltration, vapor-phase infiltration was applied to form single-phase SiC ceramics [25,29,71]. The gas transport fluxes, featuring ordinary diffusion, viscous flow (from Darcy’s law), and Knudsen diffusion, are evaluated using the Dusty-Gas Model [72]. Two parameters have been shown to pilot the vapor infiltration and reaction process, namely the Thiele modulus, $\Phi$,

$$\Phi = L \sqrt{\frac{kS_v}{\varepsilon D_{eff}}}$$  \hspace{1cm} (4)$$

and the characteristic infiltration time, $t_{inf}$,

$$t_{inf} = \frac{[\Omega C_0 S_v \varepsilon]}{\kappa}$$  \hspace{1cm} (5)$$

(maple) to 70 wt.% (balsa) [70]. Figure 9 shows two typical microstructures of biomorphous $\beta$-SiC formed by liquid Si infiltration process of softwood (pine) and hardwood (oak), which display pronounced transition from early- to latewood regions.

Figure 10 shows the variation of pore size distribution of pine and oak from the native template, the biocarbon template pyrolyzed at 800 °C and 1600 °C, and for Si-melt infiltrated SiC. At low pyrolysis temperature of 800 °C the biocarbon template retains the pore structure of the native template, increasing the temperature (1600 °C) leads to close the small pore fraction. All pores smaller than approximately 1 μm disappeared in the Si-infiltrated SiC reaction product and only pores larger than approximately 10 μm remain.
$L$ is the characteristic diffusion length, e.g. half width of component thickness, $S_v$ is the internal surface area, $\varepsilon$ is the porosity, $D_{\text{eff}}$ is the effective diffusion coefficient, $U$ is the deposit molar volume, $C_0$ is a reference precursor concentration, and $k(T)$ is a heterogeneous rate constant and follows an Arrhenius-like law. For low values of $\Phi << 0.01$, the deposition process is reaction-controlled because the reactant gases can easily reach the interior before reaction occurs. For high values of $\Phi$, the deposition process is diffusion-controlled since the mass transport of reactant gas is slow or is impeded by a constricted porosity and cannot easily reach the interior before reaction occurs. Thus, when the Thiele modulus becomes comparable to 1 or higher, then diffusional limitations will prevent the process from achieving a homogeneous enough infiltration [73]; this can be overcome by lowering the temperature or pressure but at the expense of increasing the total infiltration time. The effective diffusion coefficient $D_{\text{eff}}$, in the porous biocarbon template of tortuosity $\tau$ and porosity $\varepsilon$, is given by the molecular and Knudsen (constraint) diffusion coefficients [74]:

$$\frac{1}{D_{\text{eff}}} = \frac{\tau}{\varepsilon} \left[ \frac{1}{D_{\text{mol}}} + \frac{1}{D_{\text{Kn}}} \right]$$

Since the two diffusion coefficients are given by

$$D_{\text{mol}} = \left( \frac{kT}{\pi V_i} \right)^{3/2} \frac{1}{V_i^2 p} \sqrt{\frac{1}{M_i}} \quad \text{and} \quad D_{\text{Kn}} = \left( \frac{d}{3} \right)^2 \frac{8kT}{\pi M}$$ (7)

a minimum pore diameter, $d_{\text{min}}$, may be derived for the transition from molecular to Knudsen diffusion, below which infiltration rate becomes extremely slow and the Thiele modulus very high

$$d_{\text{min}} \approx \frac{6kT}{\pi \Omega^2 p}$$ (8)

Estimation of $d_{\text{min}}$ for Si and SiO vapor infiltration at 1400 °C–1600 °C suggested a critical minimum pore size in the range of 1 µm [14]. Thus, for Si infiltration via vapor-phase pore diameters larger than 1 µm are required to obtain reasonable infiltration rates. This may be the reason for residual C always observed in wood-tissue preforms manufactured by vapor-phase infiltration of Si-containing precursors. Nevertheless, vapor infiltration rate is significantly slower compared with liquid infiltration, and large pore diameters are required in order to achieve high infiltration rates and reasonable infiltration times [14].

Figure 11 shows a typical microstructure of biomorphous SiC ($P. sylvestris$) manufactured by Si-vapor infiltration at 1600 °C for 4 h.

Once Si has wetted or deposited on the C surface in the biocarbon template, nucleation and growth of $\beta$-SiC proceeds. Experimental observations indicate that in the case of liquid Si infiltration initially, a nanoscale $\beta$-SiC is formed, which subsequently transforms into a large-grain SiC. [71]. Grain growth was attributed to a dissolution and reprecipitation mechanism triggered by the Si melt saturated with carbon. Thus, a segmented layer structure was observed at later stage of reaction with small particle size at the C-SiC interface and a significantly larger size of well faceted particles at the Si-SiC interface. After formation of a continuous reaction layer product, grain
boundary diffusion is likely to control the conversion rate and the microstructure formation in the inorganic struts. Applying different processing conditions, biomorphous β-SiC was obtained with nanocrystalline structure, percolated porosity, and multilayer or gradient strut morphologies from a variety of hardwood and softwood species, Figure 12.

For the reaction mechanism, conflicting interpretations are available. Some authors reported continuous reaction product layers at the interface between C and Si that appeared to grow by a diffusion-controlled process [76]. No such continuous product layer was found by others [77], and it was suggested that the reaction product quickly spalled owing to volume misfit between SiC and C, leading to an interface-controlled reaction directly between Si and C. The other mechanism suggests repeating steps of dissolution of C in liquid Si, followed by precipitation of SiC from supersaturated solution of C in Si as the reaction front moves [78]. The faceted morphology of the layer of reaction-formed SiC grains observed in the study of Pancholi et al. [79] indicates that they are grown from solution and are consistent with the growth model via

![FIGURE 11 Typical microstructures of Si-vapor-infiltrated SiC derived from Pinus sylvestris.](image1)

![FIGURE 12 Strut microstructures(cell walls) of biomorphous SiC prepared according to various infiltration-reaction processes [75].](image2)
solution—precipitation [71]. The solution—precipitation and interface-controlled mechanisms have similarity with regard to linear dependence of reaction rate on carbon solubility in Si, the C—Si reaction being assumed to be of first order with respect to C concentration [77].

Orientation-imaging microscopy based on automated indexing of backscattered Kikuchi diffraction (EBSD: Electron Backscatter Diffraction) has frequently been applied to analyze the microtexture of crystal phase formed in biomorphic SiC. An earlier study on liquid Si-infiltrated biomorphic Si/SiC ceramics, processed by liquid Si infiltration into wood templates pyrolyzed at very high temperature of 1800 °C for 4 h, has reported (1 1 0)SiC//(0 0 0 1)C as a possible orientation relationship (OR) between SiC and graphite [9]. In cellular Si/SiC ceramics synthesized by replication of Indian dicotyledonous plants, EBSD confirmed a distinct orientation between the graphite crystals and the Si grains containing them [79]. No clear ORs, however, could be established between Si—SiC or SiC—graphite. Analysis of crystallite orientation in vapor-phase infiltrated biomorphic SiC indicated a whisker-like growth of $\beta$-SiC perpendicular to the template surface, e.g. cell wall surface with [111] as the dominating growth direction, Figure 13.

A variety of carbide-forming metals were used for biomorphic ceramics processing. For example, biomorphic TiC was processed by vapor-phase infiltration (CVI) of TiCl$_4$ into the biocarbon template at 1200 °C in a hydrogen atmosphere [80]:

$$C + TiCl_4 + 2H_2 \rightarrow TiC + 4HCl$$

A nanocrystalline TiC reaction product was obtained with residual carbon found in the core of struts.

3.3. Oxide Ceramics (Cell Wall Intercalation)

Chemical modification of the renewable natural raw material wood is an effective method for improving dimensional stability, decay resistance, and other properties [81]. Acetylation and propionylation of wood provides an enhanced dimensional stability and durability [82]. Another wood modification process is the combination of wood and inorganic substances to enhance the properties, particularly the flame resistance and decay resistance of wood. Silicic acid compounds impregnated in wood as colloidal silica could partially stabilize wood and were nonleachable. Wood—silica composites, prepared using the colloidal silica-boric acid, showed good decay, termite, and fire resistance properties [83]. Many plants absorb large quantities of dissolved silica, which may be deposited during growth of structural components such as rice husks, teak, bamboo, and other hardwood as well as grasses, palms, and nettles. The silica has often been isolated and characterized asopalized silica or nanoparticulate (<10 nm), arranged hierarchically in flakes and then sheets, which may be wrapped up to form spicules [84—86]. Submersion and underground embedding of wood initiate the very slow process of fossilization in which the cell wall substance is transformed into highly condensed compounds (charring) or is substituted by minerals (silification).

In higher plants, mineralization, such as for example calcium oxalate precipitation, typically develops within intravacuolar membrane chambers of specialized cells. The complex cellular features associated with calcium oxalate crystallization indicate that it constitutes a biologically controlled process, analogous to calcification processes that shape bones, teeth, and shells in animals [87]. Mineralization of the cellular template vacuoles in the cell wall has to be generated by chemical treatment prior to infiltration with precursor solution. Extraction procedures generally include treating the wood pieces with organic solvents of varying polarity. Mineralization may be achieved by intercalation of the cell walls with liquid precursor systems, which may be an inorganic (metal salts), metal organic (metal alkoxides), or organometallic (direct metal to carbon bond) sol. By embedding silica within the fiber cell wall and attaching a silane coupling agent to silanol groups,

![FIGURE 13](image-url) Cell wall structure of biomorphous $\beta$-SiC processed by Si-vapor infiltration.
stable Si–O–Si bonds would be produced. These bonds are more hydrolytically stable than the cellulose equivalent C–O–Si bonds (Figure 14).

After modification of the native cell-wall structure, immersion of the template in a low-viscous solution or colloidal dispersion causes impregnation of the cell walls. Acid- or base-catalyzed hydrolysis reactions trigger gel formation via olation, oxolation, alcolation, or alcoxolation type of condensation reaction [88]:

\[ \text{Cellulose} + M^{n+}(OR)_n + n/2H_2O \rightarrow \text{Cellulose} + M^{n+}O_{n/2} + nROH \uparrow \text{ with } M = \text{Al, Si, Ti, ...} \]

\[ >350 \, ^{\circ}\text{C, inert atmosphere} \]
\[ \text{Cellulose} + M^{n+}(OR)_n + n/2H_2O \rightarrow C + M^{n+}O_{n/2} + nROH \uparrow + \text{Gas} \uparrow \]

\[ >600 \, ^{\circ}\text{C, air} \]
\[ C + M^{n+}O_{n/2} + O_2 \rightarrow MO_{n/2} + CO_2 \]

Chelating agents forming polybasic complexes from inorganic precursors such as citric acid, acetylacetone, or ethylene glycol were also used to generate polymeric gels of multicomponent oxides [88]. Heating above the pyrolysis temperature of the hydrocarbons, forming the cell wall material in inert atmosphere (e.g. 800 °C), finally results in a positive replica of the cellular structure with a metal oxide/carbon composite forming the cell walls.

Amorphous, nano- or microcrystalline C/Si–O–C(–N) composite materials are formed by infiltration with polymeric precursor systems [RSiX_{1.5}]_n, where X = CH₂ (polycarbosilane), NH (polysilazane), O (polysiloxane), or a copolymer or mixture thereof.

\[ >800 \, ^{\circ}\text{C, inert atmosphere} \]
\[ \text{Cellulose} + (R-Si-X_{1.5})_n \rightarrow C + Si-O-C(–N) + \text{Gas with } R = H, \text{CH}_3, \text{C}_6\text{H}_5, ... \]

The Si–H functional group is able to react with OH groups of cellulose and lignin biopolymers to form Si–O ether bonds [43]. Annealing in air above 600 °C oxidizes the biocarbon residue leaving a highly porous ceramic residue.

Figure 15 shows the cellular microstructures of various biomorphous oxide ceramics derived from pine wood. Pine wood was vacuum infiltrated with different low-viscous alumina-, titania-, and zirconia-sol. Subsequent pyrolysis in inert atmosphere at 800 °C and annealing in air up to 1550 °C resulted in the formation of porous, microcellular α-Al₂O₃, TiO₂ (rutile), and c-ZrO₂ (stabilized by 8 mol% Y₂O₃), respectively [90]. The initial cellular anatomy was reproduced in the ceramic products. Mean particle size of the Al₂O₃-, TiO₂- or ZrO₂ grains is about 3–5 μm. The skeleton density of biomorphous Al₂O₃- and TiO₂ ceramics are very close to the theoretical density of Al₂O₃ (3.97 g/cm³) and TiO₂ (4.26 g/cm³), respectively, indicating only a small amount of residual strut porosity.

Ota et al. [91] produced biomorphous TiO₂ ceramics by infiltration of wood species with titanium isopropoxide.
Shin et al. [12] used a surfactant-templated sol–gel process. Rattan palm preforms were converted to biomorphous Al₂O₃, ZrO₂, and 3Al₂O₃2SiO₂ ceramics by a sol–gel process with metal alkoxides, metal oxide chlorides, and silica nanopowders [90]. Biomorphous ZrO₂ ceramics were manufactured by sol infiltration of wood [92–94]. Hydrothermal synthesis was applied to produce biomorphous NiO and ZnO from pine and fir wood, respectively [93–95]. Their porous structures were found to be hierarchical from 1 up to 25 μm (in micrometer scale) and from 2 to 60 nm (in nanometer scale). Porous silica (SiO₂) ceramic with a wood-like structure was prepared by wet impregnation tetraethyl orthosilicate into biological template that was derived from linden wood (Tilia amurensis) [96]. After repeated pressure impregnations, the subsequent annealing in air atmosphere at 800 °C resulted in oxidative removal of the template and consolidation of the oxide layers. It was found that the bioorganic structure was converted into oxide ceramics (SiO₂). At low temperature (800 °C), pore radius varied between 2 and 10 nm, indicating that the samples were mostly mesoporous. Samples treated at higher temperature (1300 °C) lost the mesoporous character; however, they were still porous, having the microstructural features of the biological preform.

4. MICROSTRUCTURE AND MECHANICAL PROPERTIES

The mechanical properties of biomorphous ceramics are dominated by the hierarchical and anisotropic structure of biological tissue. The representative volume element over which the average of mechanical properties are representative of the whole can be quite large (10³ mm³ and even equal to the component size) compared with conventional
engineering materials (down to $10^{-12}$ mm$^3$) [97]. Generally, averaging techniques from solid-mechanics theories (rule of mixtures for composites, laminate theories) are applied, which, however, is associated with the loss of some information of the subsystem in the process. This is particularly relevant in relation to mechanical events occurring at local levels (cellular level), such as damage initiation and fracture propagation, for example, as opposed to global events such as elastic behavior or structural instabilities of the kind associated with buckling and fracture [98].

The two main characteristics relevant for mechanical property validation are the density distribution of the biomorphous ceramic manufactured from the wood templates and the level of microstructural anisotropy between directions parallel (axial) and perpendicular (radial and tangential) to the growth direction that will correlate with anisotropy in the mechanical properties of biomorphous ceramics [99]. Furthermore, the cell shape and dimensions may vary significantly with growth conditions, Figure 16.

Periodical undulations of density generally occur in wood-derived ceramics due to seasonal variations of growth conditions. Figure 17 shows the radial distribution of fractional density $\rho^*$ as determined experimentally by scanning electron microscope (SEM) image analysis. While in the earlywood region $\rho^* < 0.4$, a pronounced increase of $\rho^*$ up to 0.9 was found to occur in the latewood regions.

As a consequence, nonlinear transitions of material properties (Young’s modulus, toughness, and strength) are expected to affect the fracture behavior in radial direction [98,99,101]. Local variation of fracture toughness is expected to occur when the cell size, strut thickness, and cell morphology are changed near the growth ring patterns. Figure 18 shows a typical fracture surface perpendicular to the longitudinal (axial) direction (L) of a Si-infiltrated SiC, derived from beech with longitudinal tensile loading.

4.1. Mechanical Properties of Models

Young’s modulus as well as strength models of porous ceramic materials available in literature mainly refer to two approaches:

1. Models for low-porosity materials (<50%), where strength is assumed to depend mainly on the minimum solid area perpendicular to the applied stress, generally express the property $E^*$, where $E^*$ stands for Young’s
modulus, strength or toughness, respectively, by an exponential scaling law of fractional density $\rho^*$ [102]:

$$E^* = \frac{E_{\text{cell}}}{E_{\text{strut}}} = \exp[-b(1-\rho^*)] \quad (9)$$

where $b$ was shown to be primarily a function of pore shape and alignment with respect to the stress axis but is independent on pore size, as long as that size is small with respect to the dimensions of the body. $b$ was found to equal 5 for spherical pores and higher values for elongated pores. From the analysis of strain and strain energy change associated with incrementally introducing pores into a loaded ceramic volume (incremental or differential scheme) $b(k, \cos \varphi)$ was extended to account for pores of nonspherical shape with $k$ the aspect ratio of an elliptical pore and for aligned or random pore orientation with $\varphi$ the misorientation angle of the elongated pore to the direction of applied load (e.g. for axial loading $\cos \varphi = 1$) [103].

$$(2)$$

Models for cellular materials with high porosities (>70%) where the mechanical behavior is mostly related with the bending and brittle failure of the cell walls, and $E^*$ is expressed by a power law of the fractional density $\rho^* (= \rho_{\text{cell}}/\rho_{\text{strut}})$ [100]:

$$E^* = \frac{E_{\text{cell}}}{E_{\text{strut}}} = C(\phi \rho^*)^n + C'(1-\phi) \rho^* \quad (10)$$

where $\phi$ represents the fraction of open porosity (interconnected) to closed porosity and defines the range of validity for the open cell and wood modification of the cellular model at $\phi \to 1$. $C$ and $C'$ are constants which depend on the loading direction with respect to the cellular pore orientation. Depending on the direction of loading, $n$ may attain characteristic values of 1—3, Table 3.

### 4.2. Strength

Brittle cellular ceramics of high porosity fail abruptly at a stress that is usually lower than the strength of the bulk material. The propagation of cracks will depend on, among other factors, the direction of loading to the cellular microstructure. In compression, the cells suffer progressive crushing; in tension, the cellular structure fails by fast brittle fracture. As with any brittle solid, fracture in tension is controlled by the largest defect (a crack, notch, pore, or cluster of damaged cells). Experimental values of...
strength of biomorphous ceramics were measured by four-point-bending (uniaxial loading) with longitudinal and radial loading directions or by ball-on-ring bending (biaxial loading), Figure 19. Applying central load in the longitudinal direction results in an in-plane biaxial tensile-stress state at the bottom of the cylindrical sample surface. Tensile loading in the radial and tangential directions is more sensitive to the cell microstructure, e.g. geometry, size, and packing arrangement, compared with the “strong” longitudinal cell-elongation direction. Generally, theoretical considerations predict that the in-plane strength, in every instance, is lower than that for out-of-plane loading because bending of the cell walls dominates in the first case and axial deformation in the second direction [100].

Due to the unidirected pore morphology, the highly porous, biomorphous SiC ceramics prepared by Si-vapor-phase conversion exhibit an anisotropic mechanical behavior with fracture stress in axial direction being 20 times higher compared with strength in radial direction [10,98,99]. The strength in axial direction mainly depends on the material fraction on the axial cross section similar to the strength of honeycombs [104]. In contrast, strength in radial direction strongly depends on cell morphology features, which include shape, size, and topology of pores and struts. Generally, the in-plane stiffness and strength (stress acting perpendicular to cell elongation) are the lowest, because in-plane loading makes the cell walls bend. The out-of-plane stiffness and strength (stress acting parallel to cell elongation) are much larger, because they require axial extension or compression of the cell walls. With increasing fractional density, cell walls (e.g. the struts) increase in average thickness, but cell shapes can change significantly and hence stiffness and strength increase in a complex manner.

Figure 20 summarizes strength data of biomorphous ceramics measured by four-point-bending and ball-on-ring bending.

Examinations of the fracture behavior of biomorphous SiC filled with high fractions of Si (>30%) have shown that the experimentally observed behavior can be modeled more realistically as a porous material rather than a cellular material, and that application of cellular model to strength of biomorphous ceramics is limited [66,99]. The minimum

![Figure 19](image1.png)

**FIGURE 19** Uniaxial and biaxial bending loading for strength measurements of biomorphous ceramics.

---

**TABLE 3** Power Law Coefficients $n$ of the Fractional Density of Selected Mechanical Properties of Cellular Materials According to [100]

<table>
<thead>
<tr>
<th>Direction of loading</th>
<th>Out-of-plane (axial)</th>
<th>In-plane (radial, tangential)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young's modulus</td>
<td>$n = 1$</td>
<td>$n = 3$</td>
</tr>
<tr>
<td>Strength</td>
<td>$n = 1$</td>
<td>$n = 2$</td>
</tr>
<tr>
<td>Toughness</td>
<td>$n = 1.5$</td>
<td>$n = 1.5$</td>
</tr>
</tbody>
</table>
solid-area model for close porosity is straightforward, and for connected porosity, it was calculated considering pores with truncated shapes [99]. The relative strength of the porous microstructure is equal to the relative minimum solid area in the direction perpendicular to the stress. The fracture stress of biomorphous Si prepared by liquid Si infiltration and a fractional density \( r^* < 0.5 \) can be approximated by Eqn (1) with \( b \) varying from 3–7 (\( \sigma_{\text{strut}} \approx 340 \) MPa). Vapor-phase reacted SiC with a low fractional density \( r^* > 0.5 \) shows a clear conformity with the model for cellular solids Eqn ((2)) with \( n = 2 \) found for the strength values measured by biaxial bending.

In a material with random distribution of pores, a sudden decrease of strength is observed, when the solid struts lose interconnection. Nature optimizes solid matter distribution by cellular anatomy and strength is maintained even at low fractional density. At low densities, the distribution of wall material available is optimized, and periodically packed structures like the one from pine wood are found. In pine wood, the number of neighbors is reduced to four and the cells have square shape. Thus, as the density of the structures is reduced, the microstructures of the biomorphous SiC materials change to maintain a good strength value (e.g. strut microstructure becomes more and more dominating the strength). In the case of axial compression of longitudinal tubular pores, the type of stacking and shape of pores do not have influence on the theoretical strength (the strength is simply equal to the volume fraction). For compression in radial direction, the models predict a strong dependence with the type of stacking of the tubular pores and a slight dependence with their shape.

4.3. FE-Calculations of Stress and Strain Distribution

Periodical undulations of fractional density due to variation of seasonal growth conditions give rise to a complex three-dimensional stress and strain distribution upon loading. Furthermore, transitions of cell shape such as for example from quadratic cells to elliptical cells, and cell arrangement may occur at the late-wood to earlywood transition. Numerical finite element methods (FEMs) were applied to simulate the stress distribution in ceramic honeycomb structures [105] and the effect of pore shape on Young’s moduli and Poisson’s ratio [106,107]. An overview about the use of FEM simulations for ceramics is given by Mackerle [108]. The FEM was selected as a vehicle for modeling the stress distribution of a representative volume segment of pine wood-derived SiC subjected under transverse (perpendicular to the axial direction) tensile loading [109]. Using a bottom-up method, the transition-region of wood-cell system was first divided into substructure I (square pores), and substructure II (elliptical pores), and then the complex structure with the earlywood to late-wood transition areas (combined structure III) was derived by combination of substructure I and substructure II. A 3D mesh was generated using Marc 7.3.2 and Mentat 3.3.2 (MSC.Software). A total number of \( 4 \times 10^5 \)-three dimensional 20-node brick elements were applied in the calculations. As boundary conditions, a uniaxial tensile stress of 10 MPa was applied to the surface, and the strut material was represented by the material properties of \( \beta \)-SiC (Young’s modulus 410 GPa; 

![FIGURE 20 Bending strength of biocarbon template and biomorphous SiC, derived from various kinds of wood for axial and radial/tangential loading directions.](image-url)
Possion’s ratio 0.22; density 3.1 g/cm³). Figure 21 shows the distribution of the normal stress component \(\sigma_{xx}\) parallel to the applied tensile stress for cell morphology transition, where a square cell geometry dominates in the earlywood area and an ellipsoidal cell geometry in the late-wood area. Strut thickness, cell size, and shape parameters were taken from experimental data.

Square cells loaded in transverse direction exhibit a significantly lower tensile stress at the same level of fractional density compared with ellipsoidal pores. At a fractional density of \(\rho^* = 0.35\), the maximum tensile stress between the nodes in the square cell pattern is less than 20 MPa, whereas tensile stresses exceeding 35 MPa were calculated at the sharp edges of the ellipsoidal cells. Despite the higher fractional density and strut thickness, the area at the transition from early- to latewood tissue anatomy may give rise to increased probability of crack propagation under the loading conditions considered. Crack advance is likely to be localized on the highly stressed areas, whereas regions of high density should yield higher fracture resistant segments in the loaded structure, Figure 22. Thus, the transition regions at the growth rings might act as crack deflection interfaces, which induce sequential crack propagation and hence a step-like stress–strain behavior. It would be of particular interest therefore to manipulate the growth of natural plants by chemical, physical, or genetic methods in order to tailor tissue structures with desired cellular pore shape and pore arrangement for improved fracture resistance and flaw tolerance.

While fracture propagation for longitudinal loading conditions is likely to be dominated by periodical density fluctuations, the strut microstructure of the cell walls was found to play a key role for in-plane loading situation. Figure 23 shows the stress vs. strain curves of vapor phase-processed biomorphous SiC with three different strut microstructures: a highly porous, single-layer SiC strut was formed by infiltration with SiO (\(2C + SiO(v) \rightarrow SiC + CO(v)\)); a dense, bilayer SiC strut was formed by infiltration with Si (\(C + Si(v) \rightarrow SiC\)) with a small interface layer of residual carbon often found as ML; and a three-layer strut structure was obtained after infiltration with CH₃SiCl₃/H₂ with an inner core of dense SiC and an outer layer of porous SiC [98]. All specimens exhibit a noncatastrophic failure behavior with a maximum failure strain of 0.6%.

FIGURE 21 Distribution of the normal stress component \(\sigma_{xx}\) parallel to the applied tensile stress for the cell morphology transition area in pine-derived SiC, calculated by FEM [98]. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 22 Crack propagation model for biomorphous ceramics dominated by late-/earlywood transitions (density and cell morphology) under longitudinal loading direction and cell-wall crack propagation under in-plane loading direction.

FIGURE 23 Stress vs. strain curves of vapor phase-processed biomorphous SiC with three different strut microstructures: a highly porous, single-layer SiC strut was formed by infiltration with SiO (\(2C + SiO(v) \rightarrow SiC + CO(v)\)); a dense, bilayer SiC strut was formed by infiltration with Si (\(C + Si(v) \rightarrow SiC\)) with a small interface layer of residual carbon often found as ML; and a three-layer strut structure was obtained after infiltration with CH₃SiCl₃/H₂ with an inner core of dense SiC and an outer layer of porous SiC [98]. All specimens exhibit a noncatastrophic failure behavior with a maximum failure strain of 0.6%.
Fracture strength optimization of the strut material \( \sigma_t \) involves minimization of flaws and porosity within the struts and maximize toughness (fracture-mechanics approach). An increase in strut strength would be expected as a result of reduced probability of finding a critical flaw in a smaller volume of material as predicted by Weibull’s weakest link hypothesis for strength variability. For the case of brittle materials, variation of mean strength, \( \sigma_c \), and toughness, \( K_c \), for a given loading direction were found to depend not only on the fractional density but on the Weibull modulus, \( m \), by

\[
\frac{\sigma_{c,1}}{\sigma_{c,2}} = \left( \frac{l_2}{l_1} \right)^{\frac{1}{2}} \left( \frac{\rho_{2*}}{\rho_{1*}} \right)^{\frac{1}{2}} \quad (11)
\]

and

\[
\frac{K_{c,1}}{K_{c,2}} = \left( \frac{l_1}{l_2} \right)^{1/2} \left( \frac{\rho_{1*}}{\rho_{2*}} \right)^{(1/2)} \quad (12)
\]

Thus, the fracture toughness can either increase or decrease with cell size \( l \) depending on the value of the Weibull modulus—for \( m > 4 \), \( K_c \) increases with increasing cell size, while if \( m < 4 \) it decreases. Evaluation of the ceramic microstructures indicates a reduction of the cell size \( l \) by a factor of three and an increase of strut thickness \( t \) by a factor of three at the early- to lateward transition of pine-derived SiC. Thus, for a reasonable Weibull modulus of SiC of 8, a variation of fracture toughness by 30% and of strength by 25% is estimated to occur periodically in the late- to earlywood cellular microstructure transitions.

4.4. Hardness and Erosion

Vickers hardness (HV) measurements have been performed in SiC samples fabricated from various wood precursors (pine, beech, and eucalyptus) [110]. It was found that hardness was highly dependent on the starting wood structure: a fully dense area corresponding to the wood growth rings and another that was mostly porous (porosity volume fraction between 30% and 50%). Hardness was found to increase with decreasing SiC/Si ratio. Solid-particle erosion studies were conducted on biomorphous SiSiC [111,112]. It was found that material loss occurred by brittle fracture involving the formation of both lateral and radial cracks [113]. Erosion rates in terms of volume (\( ER_V \)) were predicted from particle impact parameters and target material properties, based on the model of Wiederhorn and Lawn [114]:

\[
ER_V \propto V_p^2 R_p^3 \rho_p^{1.2} H^{0.11} \frac{K_{c,1}}{K_{c,2}}
\]

where \( V_p, R_p, \rho_p \) are the particle velocity, particle radius, and particle density, respectively. \( K_c \) is the target toughness and \( H \) is the target hardness. From experimental results, it was concluded that radial cracks able to penetrate through thin strut walls of the channel morphology of pores in the biomorphous SiC are likely to contribute to material loss by brittle fracture [99,111]. The erosion-induced surface damage causes a pronounced strength degradation. Most importantly, the erosion rates of the biomorphous SiC ceramics were found to be significantly higher than for the monolithic SiC ceramics.
4.5. Creep Behavior

Compressive creep experiments for Si-infiltrated SiC derived from a variety of wood template structures including eucalyptus (0.67 g/cm³), beech (0.49 g/cm³), and pine wood (0.33 g/cm³) were conducted at 1300 °C for constant pressures ranging from 400–900 MPa [99,115]. Creep strain rate was modeled in a similar way to systems, where creep is due to viscous grain boundary phase. The deformation of the intergranular Si allows limited grain boundary sliding between the SiC grains. For wood-derived SiC (63%) filled with high fractions of Si, an expression of the creep strain rate was derived [115]:

\[
\frac{d\varepsilon_z}{dt} = \frac{O_z}{\eta} \left[ \frac{h}{d} - \varepsilon_z \right]^3
\]

where \(\sigma_z\) is the applied (compressive stress), \(\eta\) is the effective viscosity of the Si (intergranular layer), \(\gamma\) is a shape factor for SiC grains (cubic \(\gamma = 2.38\)), \(h/d\) is the ratio of initial Si-intergranular layer width to the grain size \(d\), and \(\varepsilon_z\) is the creep strain. At 1300 °C and stresses of 250–290 MPa, strain rates smaller than \(10^{-10}\) 1/s were observed for 1% strains. Despite the high Si fraction, the creep resistance of biomorphous SiC compressed in the axial direction was found to be superior to that of conventional Si-infiltrated SiC materials, which was explained by an optimized distribution of SiC in the biomorphous material at least in the axial direction for better mechanical properties compared with conventional materials with a uniform distribution of SiC and Si. Table 4 summarizes selected mechanical properties of a variety of biomorphous ceramics derived from different wood templates.

<table>
<thead>
<tr>
<th>Wood species</th>
<th>Si content (vol.%)</th>
<th>Fractional density</th>
<th>Strength (MPa)</th>
<th>Young’s modulus (GPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oak</td>
<td>20</td>
<td>0.75</td>
<td>120</td>
<td>210</td>
</tr>
<tr>
<td>Maple</td>
<td>20</td>
<td>0.92</td>
<td>180–300</td>
<td>285</td>
</tr>
<tr>
<td>Beech</td>
<td>35</td>
<td>0.94</td>
<td>210</td>
<td>275</td>
</tr>
<tr>
<td>Pine</td>
<td>40</td>
<td>0.86</td>
<td>180</td>
<td>190</td>
</tr>
<tr>
<td>Paulowina</td>
<td>55</td>
<td>0.92</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Balsa</td>
<td>50</td>
<td>0.72</td>
<td>70</td>
<td>160</td>
</tr>
</tbody>
</table>

5. APPLICATIONS

5.1. Wood Tissue

The diversity of natural wood template structures offers a number of options in terms of material selections and applications. Potential applications of biomorphous ceramics derived from natural plant tissue are stimulated by a huge diversity of natural template structures with respect to pore volume (30–90%), pore diameter (<0.1–1000 µm), and pore shape (cylindrical, polygonal). While gymnosperms exhibit monomodal tracheidal pore size distribution such as in pine (\(P. sylvestris\)) with a maximum lumae diameter at approximately 20 µm, angiosperms typically contain a multimodal pore size distribution, as for example oak (\(Quercus Robur\)) with pore channel diameter maxima at 0.08 µm, 1–10 µm, and 200 µm (depending on the growth conditions). Due to the accessibility of large surface area coatings with functional layers including catalysts like zeolites and bioactive layers such as bio glass were explored. Table 5 summarizes a selection of biomorphous ceramics and their potential applications.

<table>
<thead>
<tr>
<th>Template Material</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pine ((P. sylvestris))</td>
<td>Al₂O₃ Microcellular filter</td>
</tr>
<tr>
<td></td>
<td>SrAl₂O₄:Eu Micropatterned phosphor</td>
</tr>
<tr>
<td>Beech ((Fagus sylvatica))</td>
<td>SiC Solar radiation heat absorber</td>
</tr>
<tr>
<td></td>
<td>SiC/Bioglass Bone regenerating scaffold</td>
</tr>
<tr>
<td>Rattan ((Calamus rotang))</td>
<td>SiSiC Zeolite catalyst support</td>
</tr>
<tr>
<td>Sponge ((Luffa aegyptiaca))</td>
<td>(Ca₃(PO₄)₂/3OH) Bone regeneration scaffold</td>
</tr>
<tr>
<td>Wood fiber board</td>
<td>Si–Cu/SiC High friction brake shoes</td>
</tr>
<tr>
<td></td>
<td>SiC/Si/C Light weight armor</td>
</tr>
</tbody>
</table>

TABLE 5 Potential Applications of Cellular Plant Tissue and Wooden Fiber Board-Derived Biomorphous Ceramics and Composites [75]
conditions, preprocessed fiber products offer more uniform microstructures and hence less variation of properties. Thus, for example, paper and preceramic paper were used as preforms for fabricating corrugated and laminated lightweight structures [37]. Preceramic paper is a multiscale composite material. It is constructed in the wet state from discrete tubular pulp fibers, which are bound to each other primarily through hydrogen bonds. A high loading of inorganic fillers may significantly modify the fiber bonding and hence the mechanical properties of preceramic paper. Pulp fibers can be made from a variety of wood plants like spruce, pine, eucalyptus, and many others. Pulp fibers are extracted from wood in chemical, mechanical, or chemimechanical processes. Typical fibers used in paper making are about 1–4 mm long and roughly 10–30 μm wide [117] (Figure 24).

Biomorphous ceramics were produced by liquid as well as vapor infiltration from paper preforms. Automotive particle filtration and catalytic exhaust gas purification are the fields of potential application of biomorphous high-temperature resistant SiC-based ceramics [118]. Selective leaching of the metal from biomorphous carbide matrices (SiC, TiC) in halogen atmosphere (Cl2/H2) was applied to manufacture carbide-derived carbon (CDC) substrates with a high active surface for catalytic applications [119]. CDC synthesized at low temperatures (400–800 °C) exhibits a microporosity with pore sizes <1 nm, which increases to mesoporosity range at higher temperature (>800 °C) and mean pore sizes of 4–6 nm. Specific surface areas of up to 1200 m²/g were measured at 800 °C, which further could be increased by thermal posttreatment (activation) at 850 °C in CO2 up to 2300 m²/g. Porous carbon scaffolds of high surface area derived from biomorphous TiC-fibrillar preforms were reported to show a high potential for antimicrobial efficiency after activation in Cl2 atmosphere at 800 °C [119].

Pyrolyzed fiber board templates were infiltrated with liquid Si to form dense biomorphous SiC/Si materials and investigated for their ballistic behavior as light weight armor [120]. Excellent resistance to single-hit ballistic impact of high velocity (>800 m/s) was demonstrated. Laminated cellulose fibre ceramics were manufactured from beech veneer by impregnation with phenol formaldehyde resin and airtight sintering [121]. The resulting laminated biocarbon material exhibited a layered structure and partially preserved microstructural characteristics of normal wood. The fracture toughness was reported to increase to 0.6–1.2 MPa m¹/² because of the laminated structure and the material exhibits a progressive failure behavior. Medium density fiber (MDF) board and wood powder were used to manufacture cellulose fibre derived ceramics [122,123]. Metal ceramic composites were fabricated by Si–Cu infiltration into porous biocarbon preforms derived from MDF board [35]. β-SiC and Si–Cu3Si were the major reaction products forming a biomorphous ceramic–metal composite. A high coefficient of friction of μ ≈ 0.6 and a good wear resistance were obtained with an optimized Si/Cu ratio of 3:1.

5.3. Nanopowders

Biomorphous β-SiC ceramics were produced at 1400 °C from pine wood impregnated with silica [124]. This one-step carbothermal reduction process decreases the cost of manufacturing of SiC ceramics compared with siliconization of carbonized wood in silicon vapor. After ultrasonic milling, the powdered sample showed an average particle size of <30 nm. The relatively small grain size in the synthesized SiC ceramics could be due to the low synthesis temperature, which impedes recrystallization and grain growth. A network of small pores existing in wood cells

FIGURE 24 Macroscopic preforms derived from preprocessed fiber preforms [37].
and transferred to the produced SiC ceramics make milling and separation into small grains relatively easy. The SiC nanopowder produced in this process may be used for manufacturing SiC ceramics for structural, tribological, and other applications.

5.4. Medicine  
Biomorphous silicon carbide ceramics were considered as a promising option for dental and orthopedic implants due to their unique mechanical and microstructure properties. The three-dimensional, highly oriented pore anatomy of native rattan (Calamus rotang) was used as a template to fabricate biomorphous hydroxyapatite ceramics design for bone regeneration scaffolds [33]. The porosity structure of theses scaffolds is similar to the anatomy of cortical bone with typical cell diameters ranging from 50 nm to 50 µm and long continuous pores up to 400 µm in diameter with a total porosity of 55–70%. In another study, it was concluded that the controlled porosity that biomorphous SiC can present is potentially excellent for osseointegration [125]. Enhanced bioactivity for accelerated osseointegration was achieved by coating the biomorphous SiC ceramic with silica-based bioactive glasses [126]. The biocompatibility and bioactivity were assessed by in vitro simulated body fluid (SBF) release tests and cytotoxicity studies.

Bioactivation of highly porous SiC ceramics by a multistep chemical surface activation process was reported by Will et al. [19] After creating a carbon-rich surface oxidation with mineral acids, (HNO₃) causes the formation of negatively charged carboxylic (−COO⁻) groups, which are able to effectively absorb Ca²⁺ ions from the body fluid. Absorbed Ca²⁺ provides ability to trigger in vitro formation of a bonelike apatite reaction layer as proofed by SBF exposure experiments. Variation of cellular anatomy of the native template with respect to porosity structure and fractional density is an advantage for adopting the biomechanical behavior of a biomorphous SiC implant to the surrounding bone tissue.

5.5. Energy  
The use of biomorphous high-temperature resistant silicon carbide is an alternative to metals for the fabrication of channeled as well as fibrillar solar radiation absorbers in the high-temperature solar-tower power generation [125]. The service temperatures of the absorber exceeds 800 °C and a specific energy absorption rate 1000 kW/m² is required.

Porous SiC with biomorphous structure was investigated for the thermoelectric properties [127]. The dielectric effect offers direct conversion of thermal energy to electricity, which has gained an interest for an ecological process for power conversion. Biomorphous SiC/Si demonstrated a negative thermal electromotoric force confirming the electronic mechanism of charge transfer. A high Seebeck coefficient (>140 V/K) was determined which increases with elevating temperature and demonstrated a certain correlation with the anisotropic cellular structure of the biomorphous SiC.

5.6. Catalysis  
Hierarchical porous structures are of major significance in catalysis, separation, and adsorption, where long-range and short-range material transports via liquid and vapor phases play a key role. Thus, biomorphous zeolite-based macrostructures with novel shapes, complex functional patterns, and hierarchical porosity are of interest for adsorbents, membranes, sensors, and catalysts of improved performance to be applied in reactor devices [128]. Slurry coating was applied to form zeolite coatings of various thicknesses on ceramic support structures derived from natural tissue templates [129]. A self-supporting MFI-type (ZSM-5) zeolite monolith derived from Luffa was manufactured by a hydrothermal surface crystallization process and tested in a fixed-bed catalytic reactor [130]. Despite the low density and high complexity Figure 25, the biomorphous zeolite reactor insert was demonstrated to achieve a reasonable mechanical stability and a high efficiency in n-hexane cracking reaction with a low deactivation tendency even at 550 °C.

Photocatalytic TiO₂ (anatase) presents a high photocatalytic activity for λ < 390 nm as well as an important chemical stability. Biomorphous TiO₂ processed from natural template, such as for example cedar wood [131] or leave surfaces [132,133], might be used in filtration and flow reactor devices. The photocatalytic activity of the biotemplated TiO₂ was demonstrated by degradation of methylene blue or rhodamine dye under UV and visible light irradiation.

5.7. Optics  
Biologically inspired optical science is a relatively new and expanding field [134]. Bioinspired optical materials with a microstructure adopted from a natural cellular template are intriguing since they possibly show increased sensitivity, improved spatial resolution, and accelerated signal processing in short-time domains [135]. Biomorphous phosphor materials with a unidirectional cellular microstructure may gain increasing interest for high-resolution screen and imaging devices. The hierarchical cellular tissue anatomy of the three-dimensionally structured biotemplate provides the microenvironment to achieve size, shape, and orientation control of the inorganic phosphor optical material. Phosphor materials exhibiting long-lasting phosphorescence or photostimulated luminescence (PSL) are interesting materials for accumulation and storage of photon-derived energy [136]. These materials have
potential applications in road markings, electrical power-
free illumination, information storage devices, and safety
labeling applications in the case of public-light blackouts.
Patterning of PSL phosphor materials is attractive for high-
resolution screen and imaging devices.

Biomorphous ceramics with a hierarchical cellular
structure were processed by liquid infiltration of inorganic
or metal-organic sols into the cellular template followed by
a controlled heat treatment (calcinations) at 400–1200 °C
to obtain oxide phases such as ZnO [95], Eu\(^{3+}\)-doped Y\(_2\)O\(_3\) [89], Eu\(^{2+}\)-doped SrAl\(_2\)O\(_4\) [27], or nonoxide phases like
Eu\(^{2+}\)-doped BaFBr [137]. The biomorphous optical
ceramics are characterized by a cellular pore structure with
pore channel morphologies and dimensions mimicking
the cellular anatomy of the template and a submicron
(nanoscale) crystallite microstructure in the inorganic
struts, Figure 26.

6. CONCLUSIONS
As the result of evolutionary processes, nature displays
a huge number of biological materials with a large variety
of cellular microstructures. Mimicking the cellular design
of natural tissue anatomy by material compositions relevant
for functional and engineering applications offers a highly
attractive approach for creating a novel class of bio-
morphous ceramic materials. The mechanical, thermal, or
optical properties of the biomorphous ceramics are gov-
erned by the hierarchical cellular structure and the phase
composition. These materials have a high potential of
achieving an excellent density-to-strength relationships and outstanding thermomechanical performance. Biomorphous ceramics may become one of increasing interest for emerging fields of application including catalysis, energy, photonics, and medicine.

REFERENCES


Chapter 7.1 Biomorphous Ceramics from Lignocellulosic Preforms


Application of a Quartz Crystal Microbalance with Dissipation for In Situ Monitoring of Interfacial Phenomena between Bioceramics and Cells

Motohiro Tagaya,* Cross J. Scott,† Toshiyuki Ikoma‡ and Junzo Tanaka‡

*Department of Materials Science and Technology, Nagaoka University of Technology, 1603-1 Kamitomioka-machi, Nagaoka, Niigata 940-2188, Japan. Former Affiliate: Department of Metallurgy and Ceramics Science, Tokyo Institute of Technology. †Department of Metallurgy and Ceramics Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

1. INTRODUCTION

When bioceramics are implanted into the body, proteins in the body fluid adsorb and subsequently cells adhere on the surface. Thus, investigation of the interfacial layers formed between the bioceramics and cells is of significant importance for controlling cell functions and understanding biocompatible phenomena. The interfacial layers can be divided into three layers, which include the ion/water adlayer, protein adlayer, and adherent cell layer. In order to in situ monitor the interfacial layers, a quartz crystal microbalance with dissipation (QCM-D) technique is one of the excellent analytical methods. The hydroxyapatite sensor surface deposited by an electrophoretic method applicable for the QCM-D technique has been recently studied for understanding the interfacial phenomena. In this chapter, the QCM-D technique that can monitor the interfacial phenomena between bioceramics and cells is described and suggests the importance of monitoring the interfacial phenomena.

2. INTERFACE BETWEEN BIOCERAMICS AND CELLS

2.1. Bioceramics: Current Status and Challenges

Ceramics used for the repair and reconstruction of diseased or damaged parts of the body are termed as “bioceramics.” Bioceramics are a class of advanced ceramics, which are defined as ceramic products or components used for the repair and replacement of diseased and damaged parts of the musculoskeletal system, employed in medical and dental applications, mainly as implants and replacements. They are biocompatible, and can be inert, bioactive, and degradable in a physiological environment that makes it an ideal biomaterial.

Bioceramics have become a diverse class of biomaterials presently including three basic types, i.e., bio-inert high-strength ceramics (alumina, zirconia, etc.), bioactive ceramics (bioactive glasses, glass-ceramics, hydroxyapatite (HAp), etc.) which form direct chemical bonds with bone or even with the soft tissue of a living organism, and various bioresorbable ceramics (tricalcium phosphate, etc.) that actively participate in the metabolic processes of an organism with predictable results. These characteristic materials applicable for biocompatible materials, such as artificial bone, dental roots, joints, blood vessels and trachea, and percutaneous devices and bone filling, have been widely studied. As future challenges, the surface property put on the biotissue should be investigated in order to elucidate the material design which promotes organizational cures.

It has been known that ceramics made of calcium phosphate salts can be used for replacing and augmenting bone tissue. The most widely used calcium phosphate based bioceramics is HAp. HAp has the chemical formula of...
Ca$_{10}$(PO$_4$)$_6$(OH)$_2$ and possesses a Ca/P ratio of 1.67, a hexagonal structure, and a space group of $P6_3/m$ with the cell dimensions of $a = 9.423$ Å and $c = 6.875$ Å [1]. The Ca(II) atom is surrounded by six O atoms belonging to the PO$_4$ groups and an OH group, whereas the Ca(I) atom is nearly octahedrally surrounded by nine O atoms belonging to the PO$_4$ groups. The Ca(II) atoms form a triangle normal to the $c$-axis, and the Ca(II) triangles stack along the $c$-axis, rotating mutually 60° from each other. The HAp surface has both positive and negative charges due to the charged Ca ion and the PO$_4$ groups, indicating the high biocompatibility [2] and characteristic property of protein absorption [3,4]. Thus, the HAp has been widely studied for application use as a bone-filling material [5–8], column-filling material [9–11], drug delivery carrier [12–15], and novel inorganic–organic composite material [16,17]. Although the applications of bioceramics have been widely studied, there is little information on the interfacial phenomena between the surfaces and biomolecules. Thus, the interfacial phenomena with the dynamic biomolecular behavior on the surfaces should be elucidated, and the future toward potential use of bioceramics should be studied for improving the biocompatibility and designing superior surfaces.

### 2.2. Importance of Interface Phenomena

Figure 1 shows the scheme of successive events after the implantation. When biomaterials are implanted into the body, three successive processes occur as follows [18]. The first substances to reach the surface are ions and water molecules, which are known to interact and bind depending on the surface properties. The water layers formed on the surface are an important factor influencing proteins, and then other molecules arrive a little later. Second, water-soluble proteins have hydration shells, and the interactions between the surface water layers with the protein water shells strongly influence the fundamental kinetics and thermodynamics at the interface to form protein adlayers. The interactions would determine the structures of the protein adlayers such as being denatured or not, orientation, and coverage, indicating that the surface properties such as topography, wettability, and charge are of great importance. Third, when cells arrive at the surface, they realize the structure of protein adlayers, which are initially determined by the water layers, to adhere, spread, and form the interface on the surface. Thus, the initial cell adhesion behavior is strongly affected by both the surface properties and structures of the protein adlayers, and the interface layers would be the dominant factors affecting the biocompatibility [18–21].

Figure 2 shows the scheme of a typical interface between the protein adlayer and cell on the biomaterial surface. After implantation into the body, multiple proteins in the body fluid immediately and competitively adsorb on the surface [19–21]. Albumin (Ab) and immunoglobulin (IgG) are the larger mass fractional protein components in the fluid, whereas extracellular matrix (ECM) proteins, which obligate the cell activities, are relatively minor components. In the protein adlayer, the structure of the extracellular matrix (ECM) containing the arginine–glycine–asparagine (RGD) sequence plays a role at the interface between the
surface and cell [24,30,31]. The conformation and denaturation of protein adlayers and the orientation of the RGD sequence strongly governs the biocompatibility [22,23], which is attributed to the cell activities and functions, such as adhesion, proliferation, migration, differentiation, expression, and survival. The ECM is classified into collagen, noncollagen glycoprotein, elastin, and proteoglycan groups [35–37]. The noncollagen glycoprotein includes fibrinogen (Fgn), fibronectin (Fn), laminin, vitronectin, thrombospondins, and tenascins, which significantly affect the initial cell adhesion behavior. On the other hand, integrins in the cell membrane directly mediate attachment between a cell and the RGD sequence of the protein adlayer, which may also be other cells or the ECM [25,26]. The cytoskeleton firmly combined with the integrin inside the cell membrane is changed, and subsequently forms adhesion points at the interface. Along with forming the points, signal transduction molecules through the cytoskeleton are successively communicated to determine the cell functions. Simultaneously, the cell produces ECM at the interface to consolidate the interfacial junction. The interfacial region above the material surface including the ECM—integrin—cytoskeleton significantly affects the cell activities and functions. Therefore, in situ monitoring technique of the interface between the material surface and cell is important for controlling the cell functions.

The surface morphology and protein structure are well known to be attractive parameters for controlling the cell functions. The surface morphology at the nano- and micrometer levels and subsequent protein adsorption are known to affect the later cellular activities (proliferation, survival, and gene expression) [27–29]. These later activities would be significantly affected by the initial cellular activities (adhesion and spreading), but there has been no clarification of the relationship between the physicochemical property and the initial behavior. The initial cell behavior on the adsorbed Fn on the self-assembled monolayer [32,33] and on poly (tetrafluoroethylene) surfaces [34] has been investigated, but the higher-order structure of the ECM proteins has not been significantly related to the initial behavior. It appears that techniques for in situ monitoring the interfacial phenomenon in an aqueous solution have not been established. Thus, it is indispensable to clarify the phenomena between the material surface and cell during the initial stage by new techniques for understanding the biocompatibility.

3. IN SITU MONITORING OF INTERFACIAL PHENOMENA

3.1. Various Current Techniques

Among the various means to investigate the biomolecular interactions on the surfaces, in situ monitoring and highly-sensitive detecting techniques for the adsorption and desorption on material surfaces in an aqueous solution have been widely developed. The in situ monitoring of the dynamic interaction changes by the calorific capacity [44,45], Fourier transform infrared spectroscopy with attenuated total internal reflectance (FTIR/ATR) [48,49], optical interferometry [42,43], ellipsometry [46,47], surface plasmon resonance (SPR) [38,39], total internal reflection fluorescence (TIRF) [40,41], and a quartz crystal microbalance (QCM) [50] have already been investigated. The evaluation of the absolute amount of the adlayer with the calorific capacity and FTIR/ATR techniques is difficult, and the refraction index parameters for the ellipsometry and optical interferometry and the refraction index parameter and density for the TIRF and SPR are needed for the evaluation, indicating the difficulty of monitoring an unknown adlayer.

Biomolecules have various functional groups that include hydrophobic groups (aromatic rings, alkyl chains, etc.), hydrophilic groups (−OH, −NH2, etc.), and dissociation groups (−COO−, −NH3+, = PO43−). The hydrophilic and dissociation groups interact to form water molecules with hydrogen bonds in an aqueous solution. Thus, the monitoring of biomolecules with a hydration structure is of great importance for investigating the interfacial structure. The QCM technique provides the mass containing the hydrated water molecules [51] to show the higher values as compared with other techniques, such as ellipsometry and SPR [52,53]. The QCM technique is also a reliable and suitable method with a high sensitivity due to its high response to mass changes of several nanogram order and for forming multilayered films, and can be used for evaluating the interface between the materials and cells.

3.2. Quartz Crystal Microbalance with Dissipation (Qcm-D) Technique

3.2.1. Fundamental Principles of Qcm-D

The mechanism of signal transduction in the QCM technique, which relies upon the piezoelectric effect in quartz crystals, was first discovered in 1880 by the Curie brothers via a pressure effect on quartz [54]. A change in inertia of a vibrating crystal has been shown by Lord Rayleigh to alter its resonant frequency (f) [55]. The subsequent developments for crystal stability were achieved through the use of electric resonators [56] and room-temperature stable AT-cut crystals [57]. For quartz crystals used in typical QCM applications, a few tenths of a mm in thickness at a 35° 10' angle from the Z-axis are cut as the AT form [58–60]. This geometry provides a stable oscillation with almost no temperature fluctuation in the f at room temperature.

A valuable advantage of the QCM technique for monitoring solid thin films is its capability to detect an adsorbed mass. For the pure elastic mass on the sensor surface, the well-known linear equation was first developed
by Sauerbrey [61]. Based on the assumption of a homogeneous foreign mass ($\Delta m$) with a spatially uniform density, the equation can be written including harmonic resonances as follows:

$$\Delta f = \frac{-2\Delta mnf_0^2}{A\sqrt{\mu q\rho q}}$$

(1)

where $\Delta f$ is the frequency shift of the quartz, $f_0$ is the fundamental frequency, $n$ is the overtone number, elasticity of quartz, and density of quartz, and $A$ is a metallic electrode area on the quartz. $\mu q$ and $\rho q$ are the shear modulus ($2.95 \times 10^{11}$ dyn/cm$^2$) and density (2.65 g/cm$^3$) of the quartz, respectively. $\Delta m$, in the nanogram range, uses a linear relationship between the added mass and the change in the resonant frequency, and an increase in the mass bound to the quartz surface causes the crystal’s oscillation frequency to decrease. However, the equation does not sufficiently fit for “viscoelastic adlayers” due to the strong damping of oscillation in liquid.

Konash [62] and Nomura [49] modified the oscillation circuit and covered the nonsensitive sensor surface to provide QCM monitoring with oscillation in liquid. Landau et al. [63] showed that the effective detecting height $l$ in liquid is

$$l = \sqrt{\frac{\eta_l}{\rho_l f_0}}$$

(2)

where $\rho_l$ and $\eta_l$ are the liquid density and viscosity of the liquid, respectively. Since the liquid is dragged by the oscillation of the sensor, there is a problem of an increase in the resonance series resistance.

Assuming the adsorbed liquid on the sensor surface and considering $l$ dependent on $\rho_l$ and $\eta_l$ of the liquid and $\Delta m (= \rho_l / A)$, the relationship between the $\Delta f$ and the viscosity-density product from Eqns (1) and (2) was represented by Kanazawa and Gordon as follows [64]:

$$\Delta f = -f_0^{3/2} \sqrt{\frac{\eta_l}{\pi \mu_q \rho_q}}$$

(3)

When the interface between the solution and the crystal surface is nonslip and the region at $l$ is a uniform viscoelastic layer, the relation between $\Delta f$ and $\Delta m$ in Eqn (3) is established. In other words, the equation for the locally different energy damping at $l$ on the surface has not been established. The bimolecular adlayer in an aqueous solution is complex with a hydration structure, and can be considered to be a viscoelastic layer. Thus, the accurate evaluation of the mass and viscoelasticity of the adlayer is needed.

Based on these points, Kasemo et al. suggested a quartz crystal microbalance with dissipation (QCM-D) technique for the in situ monitoring of the energy dissipation process with bimolecular adsorption [65–68], which provides a rapid evaluation of the mass and viscoelasticity of the hydrated bimolecular adlayers. The heart of the QCM-D is a 10-mm sensor which contains a thin disk of crystalline AT-cut quartz sandwiched between two gold electrodes, and the top one is the sensor surface. Any layer of molecules absorbing on the surface of the quartz crystal creates subtle changes in the frequency of oscillation and its duration.

Figure 3 shows the experimental setup for the QCM-D, and the typical rapid excitation of the QCM near resonance and subsequent damping. The measurement cell is connected to computer-controlled electronics. A control program tracks the $f$ and excites the crystal through a signal generator with an applied AC voltage for 10 ms. When the signal generator is switched off for 2 ms, an oscilloscope records the exponentially damped response of the oscillation from the quartz crystal in Figure 3(a). Thus, the computer repeatedly excites the crystal and records the changes in the $f$ and decay time ($\tau$) as shown in Figure 3(b), and the damping data are transformed from analog to digital. Based on Eqn (4), a numerical fit of an exponentially damped sinusoidal function to the recorded signal provides the $f$ and $\tau$:

$$A(t) = A_0 e^{-t/\tau} \sin 2\pi ft$$

(4)

where $A_0$ is the amplitude at $t = 0$, and $\tau$ is defined as the time when $A_0$ reaches $1/e$. These values are stored as one data point as a function of time.
The energy dissipation ($D$) is a dimensionless quantity defined by Eqn (5) [65–68]:

$$D = 1/Q = E_{\text{dissipation}}/2\pi E_{\text{stored}}$$

(5)

where $Q$ is the quality factor of oscillation in the equivalent electric circuit (Butterworth Van-Dyke element), and $E_{\text{stored}}$ and $E_{\text{dissipation}}$ are the energy stored and energy dissipation in the oscillation circuit during one period, respectively. The $Q$ value is represented as

$$Q = R_1/2\pi f L_1$$

(6)

where $R_1$ and $L_1$ are an inductance and resistance, respectively. The equivalent electric circuit corresponds to a mechanical vibration model, and $\tau$ equals $2L_1/R_1$. Thus, based on Eqns (5) and (6), $D$ is represented by $\tau$ and $f$ as follows:

$$D = 1/\pi f \tau$$

(7)

**Figure 4** shows the schemes of the damping with the biomolecular adsorption on the QCM-D sensor surface, and the $\Delta f$ and $\Delta D$ curves in situ measured with the changes in $f$ and $D$. When the changes from $f_0$ and $D_0$ in the time at $t = 0$ to $f_i$ and $D_i$ in the time at $t = i$ are observed with the formation of an adlayer, $\Delta f (= f_i - f_0)$ and $\Delta D (= D_i - D_0)$ were calculated. The $\Delta f$ based on Eqn (1) corresponds to $\Delta m$, and $\Delta D$ corresponds to $\Delta E_{\text{dissipation}}$ due to the constant $E_{\text{stored}}$ during each period as follows [69]:

$$\Delta D = \Delta E_{\text{dissipation}}/2\pi E_{\text{stored}}$$

(8)

The $\Delta f$ and $\Delta D$ curves can be obtained by measuring $f_i$ and $D_i$ in the time at $t = i$ for each time. Viscous adlayers effectively lose the oscillation energy by friction among the molecules, and then the oscillations are rapidly damped to produce higher $\Delta D$ values. Elastic adlayers effectively retain the oscillation energy, and the oscillations are slowly damped to produce lower $\Delta D$ values. Therefore, in situ monitoring of the mass and viscoelastic changes due to biomolecular adsorption can be successfully measured.

**3.2.2. Fabrication of Hydroxyapatite (HAp) Sensor Surface and its Reusability**

To understand the interaction between a biomaterial surface and biomolecules, the use of a biomaterial as the sensor surface has been investigated. The fabrication of a sensor surface with controlling the film thickness and adhesion is an important factor for the sensing response ability. The QCM-D sensor surfaces of gold, titanium, and poly(ethylene terephthalate) (PET) have been developed by q-sense Co. Ltd., but a bioceramic sensor has yet not been developed. To clarify the interactions between bioceramics and biomolecules, the development of a fabrication technique of the HAp surface similar to that in our body is of great importance. Furthermore, a reusability technique of the QCM-D sensor should be developed due to its expense.

Surface modifications with HAp formations on bio-inert materials have been widely studied to provide a biocompatibility. For example, magnetron sputtering [70,71], beam sputtering [72], plasma spray [73,74], laser ablation [75], and electrophoretic deposition (EPD) [76,77] have been developed. However, sputtering and laser ablation have several issues such as target fabrication and expensive equipment costs. The films formed with the sputtering, laser ablation, and plasma spray have different chemical compositions and morphologies as compared with the HAp in our body because they are high-energy and high-temperature processes. On the other hand, the EPD of synthetic HAp nanocrystals on a gold surface has been reported [73,78–81], since the synthesized chemical composition and morphology can be retained during the formation.

**Figure 5(a–c)** show the EPD system setup, the film formation process of HAp nanocrystals, and the HAp sensor. HAp surfaces with the EPD method applicable for the QCM-D technique have already been fabricated [83–87]. As the specific fabrication method, a suspension of HAp nanocrystals synthesized using a wet chemical method [86] was centrifuged at 2000 $g$ for 15 min, washed three times with ethanol, and ultrasonically dispersed in ethanol at 1 wt%. Before the EPD, the gold sensor surface.
was cleaned by immersing it for 10 min in an ammonia and hydrogen peroxide mixture (APM) solution (weight ratio of 5:1:1 mixture of Milli-Q quality distilled water, H₂O₂, and NH₃) at 70 °C and then dried by blowing N₂ gas. The cleaned surface was irradiated with UV light (λ_{irr} = 254 and 185 nm; UV/Ozone, Bioforce Nanoscience Co., Ltd.) in air for 10 min. A direct current voltage was applied at 100 V/cm for 1 min as shown in Figure 5(a). Any surplus nanocrystals were removed by a 1-min ultrasonic treatment (28 kHz, 100 W) in ethanol as shown in Figure 5(b), and the HAp sensor was successfully obtained as shown in Figure 5(c). As a result, the AFM topographic images of gold and HAp sensors in a 1 × 1 μm² area were observed as shown in Figure 5(d) and (e). The weight change of the HAp nanocrystals under air was 4.0 ± 0.2 μg/cm² by QCM-D, and the thickness was calculated at 12.9 ± 0.5 nm (based on the density of HAp as 3.14 g/cm³). The HAp nanocrystals deposited on the gold sensor had an RMS value of 4.2 ± 0.8 nm, which was almost the same as the gold surface with the values of 0.8 ± 0.4 nm.
The EPD mechanism has been described as follows. Alcohols are known to behave as a proton donor in the presence of bases, and pure alcohol can ionize in the following way: \( \text{RCH}_2\text{OH} + \text{RCH}_2\text{OH} = \text{RCH}_2\text{O}^- + \text{RCH}_2\text{OH}^+ \). Thus, the protonated alcohol is decomposed and leaves a proton on the surface, while the dissociated alcohol and alkoxide ion are desorbed into the solution. In ethanol, the ethoxide ion behaves as the counter anion, and then the HAp surface has a positive charge [80–82].

Based on the EPD method, the HAp nanocrystals with different crystallinities synthesized at 4 °C, 24 °C and 80 °C were precipitated, resulting in a deposited HAp nanolayer with a thickness of 10–20 nm consisting of the different crystal sizes [85]. The crystalline sizes of \( d_{100} \) and \( d_{002} \) were 22 nm and 67 nm at 4 °C, 28 nm and 88 nm at 24 °C, and 54 nm and 158 nm at 80 °C, respectively. The application of a DC bias of 10 and 50 V/cm resulted in inhomogeneous depositions with partly bare gold surfaces, and 100 V/cm of the DC bias ensured a homogeneous HAp covering over the gold surface. The density of the nanocrystals on the gold surface varied depending on the applied DC voltage, indicating the importance of the acceleration of the positively charged HAp nanocrystals to the cathode for controlling the deposition ability. Furthermore, the QCM-D measurement can reveal the real-time protein adsorption on the HAp sensor surfaces, and the adlayers were found to exhibit a flexible property [78,83,85,87,88]. The QCM-D sensor is very expensive, and its reusability should also be investigated for practical applications.

The cleaning methods of the sensor for organic chemicals [89] or proteins [90] adsorbed on metal sensors involve detergents and monitoring by the QCM technique. For example, the amount of tripalmitin or dotriacontane desorbed from the gold sensors was 88 or 78 wt%, respectively, when using octaethylene glycol mono-\( n \)-dodecyl ether as the detergent [89], and the removal of Ab adsorbed on the titanium and chromium sensors is 90 wt% when using an ionic surfactant of sodium dodecyl sulfate (SDS) at 37 or 90 °C [90]. On another HAp sensor (HAp/phosphate-terminated polymer/gold), the repeated adsorption and removal of a tea stain were studied using a flowing aqueous solution of sodium tripolyphosphate and QCM technique. A decrease in the frequency shift was observed during repeated testing [71]. Ionic surfactants, such as SDS, have been considered as candidates for removing chemicals adsorbed on bioceramics because of their strong exchange reactions; however, these surfactants were easily adsorbed on HAp [91–94]. Effective removal methods for HAp that do not damage or change the sensor surface properties and retain the reusability of the HAp sensors should be developed.

Q-Sense Co. Ltd. recommends a combined APM and UV treatment for cleaning the QCM-D gold sensor surface and its reusability. The treatment with APM removes dust and nanoparticles from the SiO\(_2\)/Si surface; it is a standard treatment by the Radio Corporation of America [95]. The APM generates the hydroxyl radical (\( \cdot\text{OH} \)) through the reaction of \( \text{H}_2\text{O}_2 \) with \( \text{H}_2\text{O} \); the hydrogen peroxide radical (\( \cdot\text{OOH} \)) is then produced by the reaction of the \( \cdot\text{OH} \) radical with \( \text{H}_2\text{O}_2 \) [96–98], and these radical species have been used as a \( \text{SiO}_2\)/Si etchant. The free-radical-mediated oxidation also leads to the hydroxylation of the amino groups of proteins, the conversion of amino acid residues to carbonyl derivatives, and the cleavage of polypeptide chains [99]. On the other hand, Bolon and Kunz [100] reported that exposure to ultraviolet light (UV) in an oxygen atmosphere leads to the depolymerization of photoresist polymers. The UV treatment has been used as an effective method of decomposing hydrocarbons on substrates [101,102]. The common light source has intense emission lines at 185 and 254 nm. The 185-nm photons induce the dissociation of \( \text{O}_2 \) by generating singlet oxygen (\( \text{O}^*(\text{D}) \)), which collides with inert \( \text{N}_2 \) to generate triplet oxygen (\( \text{O}^*(\text{P}) \)) and reacts with \( \text{O}_2 \) to produce ozone (\( \text{O}_3 \)). The UV also decomposes \( \text{O}_3 \) to \( \text{O}^*(\text{D}) \), which cleaves organic bonds such as \( \text{C}–\text{C}, \text{C}–\text{H}, \) and \( \text{O}–\text{H} \) [103,104].

The repeatability of the adsorption and removal of Fgn and fetal bovine serum (FBS) on HAp sensors using the APM and UV methods was investigated with an FT-IR spectroscopy and QCM-D monitoring [83]. Proteins adsorbed on the HAp sensors were removed by an ammonia/hydrogen peroxide mixture (APM), ultraviolet light (UV), UV/APM, APM/UV, and sodium dodecyl sulfate (SDS) treatments. The FTIR spectra of the reused surfaces revealed that the APM and SDS treatments left peptide fragments or proteins adsorbed on the surfaces, whereas the other methods successfully removed the proteins. The QCM-D measurements indicated that in the removal treatments, Fgn was slowly adsorbed in the first cycle because of the change in surface wettability revealed by contact angle measurements. The SDS treatment was not effective in removing proteins. The APM or UV treatment decreased the \( \Delta f \) for the reused HAp sensors. The UV/APM treatment did not induce the \( \Delta f \), but decreased the \( \Delta D \). Therefore, the APM/UV treatment was found to be the most useful method for reproducing the protein adsorption behavior on HAp sensors.

### 3.2.3. Protein Adsorption Studies on the HAp

#### 3.2.3.1. Mono-Component Adsorption

It is well known that protein adsorption depends on the surface properties such as wettability, free energy, charge, and roughness [105,106]. The mono-component protein adsorption process and any conformational changes by the QCM-D technique have been vigorously investigated [110,111]. The ions in the solvent are known to adsorb on surfaces to form hydration structures, which influence the fundamental protein adsorption kinetics [18]. Thus, the studies analyzed for the adsorption depending on the
solvent have been described by the batch method to clarify the Ab adsorption on HAp in phosphate-buffered saline (PBS) [107]. Ab adsorption on silica—titania in PBS and 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid [108], and Ab, IgG, Fgn, and lysozyme adsorption on germanium in PBS or Tris—HCl [109]. The conformation of the Ab adlayer on gold [112] and Fn adlayer on HAp [113] involving the binding of a monoclonal antibody were also discussed. However, in situ monitoring of the adsorption behavior and conformational change by the QCM-D technique on the bioceramics have not yet been reported. To understand the interfacial phenomena in vivo and in vitro, the various protein adsorption behaviors in body fluids should be precisely investigated.

Only a few studies of the saturated $\Delta D/\Delta f$ value ($\Delta D_{sat}/\Delta f_{sat}$) in the $\Delta D-\Delta f$ plot from the measured $\Delta f$ and $\Delta D$ curves have been used for evaluating the adsorption behavior and the conformation [69,78,114,115]. Figure 6 shows the schemes of the $\Delta D-\Delta f$ plots of different protein adlayers and the structures. Based on Eqn (8), the $\Delta D_{sat}/\Delta f_{sat}$ value corresponds to the $E_{diss}$ dissipation per unit adsorption mass, indicating the viscoelastic property of the adlayer. The higher or lower $\Delta D_{sat}/\Delta f_{sat}$ value donates the viscous or elastic property of the adlayer, respectively. Plot 1 showed a higher $\Delta D_{sat}/\Delta f_{sat}$ value when compared with plot 2 in Figure 6(a), indicating the relatively loose structure and viscous property of plot 1 in Figure 6(b). Ozeki et al. revealed a significant relationship between the $\Delta D_{sat}/\Delta f_{sat}$ value and the hydrated amount [69]. Rodahi et al. revealed by defining the protein adlayer as a Newtonian liquid that the $\Delta D_{sat}/\Delta f_{sat}$ value was significantly related to the inverse of the friction coefficient between the adlayer and the sensor surface [114]. Monkawa et al. reported that the $\Delta D_{sat}/\Delta f_{sat}$ value of the Fgn adlayer successfully corresponds to the conformation with an additional FT-IR analysis [78]. Based on the evaluation, Yoshioka et al. discussed the conformation of various acidic and basic protein adlayers on HAp using the $\Delta D_{sat}/\Delta f_{sat}$ values [115]. Therefore, the $\Delta D_{sat}/\Delta f_{sat}$ value is of great importance for evaluating the structure of the protein adlayer. The detectable height ($l$) of $D$ in the QCM-D system can be represented by Eqn (9) as follows [116]:

$$ l = \sqrt{\frac{\eta}{\pi \rho f}} \quad (9) $$

where the $\eta$ and $\rho$ values are the viscosity and density on the sensor surface, respectively. Thus, the viscous liquid adlayer on the sensor surface showed the higher $l$ value. When $1/\tan \delta$ is defined as $\chi$, the $\Delta f$ and $\Delta D$ values can be represented by Eqn (10) and Eqn (11), respectively [116]:

$$ \Delta f \approx \frac{1}{\rho_q d_q} \frac{d_{ad} \rho_{ad} f}{d_{ad} \rho_{ad} f} \left(1 + \frac{2 d_{ad}^2 \chi}{3 f^2 (1 + \chi^2)}\right) \quad (10) $$

$$ \Delta D \approx \frac{2 \rho_{ad} d_{ad}^3}{3 \pi \sigma_0 d_q} \frac{1}{f^2 (1 + \chi^2)} \quad (11) $$

where $d_{ad}$ is the thickness of the adlayer. Thus, the $\Delta D/\Delta f$ value is affected by only $d_{ad}$ and $\tan \delta$, and the adlayer of plot 1 also indicates the higher $d_{ad}$ and/or $\tan \delta$ values as compared with plot 2. Based on the viewpoints and the previous reports on hydration [69] and friction [78], the model structure of the adlayer of plot 1 and 2 can be represented as shown in Figure 6(b) and (c), respectively. Therefore, the $\Delta D_{sat}/\Delta f_{sat}$ value is one of the excellent tools to evaluate the viscoelasticity and structure of the adlayer.

Figure 7 shows the scheme of the secondary structure of Fgn. The Fgn, a structural glycoprotein in blood plasma with an isoelectric point at 5.5, is often employed for QCM experiments of protein adsorption [52,117–119] due to its moderate molecular weight (340 kDa) and size (45 nm length [120]). A central hydrophobic $E$ domain is connected to two hydrophobic $D$ domains with a coiled-coil chain. These hydrophobic domains are charged negative under neutral pH condition. The $\alpha$C domains, with Arg and Lys residues, are charged positive and are substantially more hydrophilic than the $E$ and $D$ domains. On the other
hand, Ab is globular with a pI at 4.7 and molecular weight (66.5 kDa), and has an asymmetric heart-like structure in which three main domains are divided into six subunit domains [121]. The protein surface with many carboxyl groups and 19 imidazole groups affects the hydrophilic property. The adsorption behavior depending on the characteristic protein morphology (fiber or globular) can be also detected by the QCM-D technique.

Figure 8 shows the ΔD−Δf plots and possible schemes of the conformational changes with the Fgn and Ab adsorptions on the HAp in PBS. The ΔD−Δf plots in Figure 8(a, c) show the 2-step change for Fgn and linear change for Ab, indicating the two-step conformational change and monomolecular adsorption behavior on the HAp surface. The ΔD sat/Δf sat value of Ab on the HAp sensors was 1.0 × 10−8, which is much lower than that of Fgn (34.5 × 10−8) [78,85]. Thus, the Ab was hardly absorbed on the HAp surface compared with the Fgn. The adsorption behavior of Fgn on HAp with its dumbbell-like structure has already been described [78,85]. One of the αC domains of Fgn, charged positive, is bonded to the negative sites of the HAp surface similar to phosphate and/or hydroxyl ions like in an “end-on” model as shown in Figure 8(b). On the contrary, the Ab has an asymmetric heart-like structure in which three main domains are divided into the six subunit domains already mentioned. The low adsorption amount of Ab as compared with that of Fgn indicates that the adsorption model of Ab could be “side-on” at the monolayer as shown in Figure 8(d), and the dissociated carboxyl and imidazole groups interact with the positively charged calcium ions on the HAp surface. Therefore, the different adsorption behavior of the Fgn and Ab with an almost similar pI value can be attributed to their secondary structure and realization of different adsorption models such as “side-on” and “end-on”.

3.2.3.2. Multicomponent Adsorption

When biomaterials are implanted into the body, multiple proteins in the body fluids immediately and competitively adsorb on the surfaces [19–21]. Since the protein adlayers govern the biocompatibility, the protein–surface interaction is important for designing biocompatible materials. From these points, the protein adsorption behaviors have been...
widely investigated [127–130], but the complex reaction at
the interface including multiple proteins is still controver-
sial. The “Vroman effect” is a phenomenon in which the
adsorbed Fgn on surfaces is replaced with other abundant
proteins in serum [124]. Multiple-protein adsorption for Ab
and IgG labeled with radioactive $^{125}$I and $^{131}$I [131,132] and
surface plasmon resonance for blood plasma solution [133]
have been described in recent studies.

Protein adsorption mechanism on a biomaterial surface
can be divided into the (i)–(iii) processes vs. time: (i) The
hydrated protein in the liquid interacts with the hydration
layer on the surface, and the adsorption on the surface
occurs with disruption of the hydration structure [18]. (ii)
The protein repeatedly adsorbs and desorbs on the surface
to thermodynamically stabilize the structure [122]. During
the adsorption of multicomponent proteins, the adsorbed
proteins simultaneously exchange with other proteins
[123,124,126]. (iii) The adsorbed proteins change their
conformation to become stable higher-order structures on
the surface, and the stabilization indicates an equilibrium
state [125]. During the adsorption of multicomponent
proteins, other proteins may subsequently co-adsorb on
the stabilized protein surfaces. These interfacial changes are
very important for understanding the phenomena in body
fluids, but only a few studies of the adsorption behavior and
conformational changes have already been reported
[127–130]. Therefore, the QCM-D technique is one of the
suitable techniques for in situ analyzing the interfacial
phenomena with the multicomponent protein adsorption.

In culture of cells in vitro, a culture medium is added for
the cell’s nature. The medium contains FBS of the basic
components, such as hormones and growth factors. The
FBS is blood containing many kinds of proteins without
cells, platelets, and clotting factors. The two major groups
of proteins are Ab and globulin. Ab is the largest mass
fractional protein component in the blood and is known to
eliminate cell attachment and block nonspecific binding
[134,135]. The other proteins in the blood are Fgn, Fn,
collagen (Col), and other subtle trace proteins (osteopontin, 
laminin, and vitronectin). These proteins are obligate
adhesive proteins for integrin-receptor-based cell attach-
ment and proliferation on surfaces, and are important
components of the ECM. Grainger et al. investigated the Fn
conformation adsorbed on polytetrafluoroethylene from
a double-component Fn–Ab mixture solution using an Fn
monoclonal antibody [34]. The antibody binding to the
compound adlayer was suppressed by the existence of Ab
to clarify Ab masking of the adsorbed Fn. Thus, the ratio of
the nonadhesive (e.g. Ab) to adhesive (e.g. Fn) proteins
from a multicomponent solution (e.g. FBS) selectively
adsorbed on the bioceramics is the key parameter to
understand and improve cell attachment on the surfaces.

The adsorption behavior of FBS on HAp was analyzed in
situ in PBS, minimum essential media (zMEM or DMEM),
and different carbonate-concentrated solutions using the
QCM-D technique [87,136]. Although the adsorption
behaviors of FBS on gold showed no dependence on the
solvents, those on HAp clearly depended on the adsorbed
ions in the solvents. Figure 9(a) shows the $\Delta D - \Delta f$
plots of the FBS adsorption on the HAp sensor in PBS, zMEM,
and DMEM. The adsorption amounts of FBS in PBS,
zMEM, and DMEM after 60 min were 1.57 ± 0.05, 0.79 ±
0.06, and 0.73 ± 0.06 µg/cm$^2$ on the HAp, respectively.
The $\Delta D_{\text{sat}}/\Delta f_{\text{sat}}$ values after 60 min were $-8.8 \pm 3.3 \times 10^{-8}$,
$-4.1 \pm 1.7 \times 10^{-8}$, and $-3.2 \pm 1.8 \times 10^{-8}$ on the HAp,
respectively. The adsorption amount and viscoelastic prop-
erty on the HAp in PBS are approximately two times larger
than those in the MEMs. The pre-adsorptive ions in the
solvents on the HAp and components of the surfaces induce
different adsorption behaviors. Acidic proteins with a nega-
tive charge at neutral pH of the buffers would preferentially
adsorb on the HAp surface. Figure 9(b) shows the scheme of
the FBS adlayers on the HAp sensor in the different solvents.

The structure of FBS adlayer on the HAp depended on the
solvent. The adsorption of carbonate ions from the MEMs on
HAp caused a decrease in the adsorption amount of FBS and
degradation of the viscoelastic property of the FBS adlayer.
Thus, the carbonate ions adsorbed on the HAp surface
strongly affect the FBS adsorption. As a consequence,
the adsorption amounts of carbonate ions on the HAp in the
carbonate solutions increased with increasing carbonate
concentrations, which inhibited the FBS adsorption and
decreased the viscoelastic property of the adlayers. Vroman
et al. described the exchange adsorption phenomenon
of proteins on a biomaterial: the order of the exchange
adsorption was related with the molecular weight of the
proteins [124]. The exchange adsorption from the multiple
proteins has not been clearly detected based on a change in
the $\Delta D - \Delta f$ plots by the QCM-D technique; however, the
$\Delta D_{\text{sat}}/\Delta f_{\text{sat}}$ values of the FBS adsorption were apparently
different from those of a single-protein adsorption. It is
speculated that the multiple proteins, such as Ab, IgG, etc.,
are simultaneously adsorbed on the HAp and the protein
species adsorbed on the HAp found to depend on the pre-
adsorbed ions and the solvent components.

3.2.3.3. Voigt-Based Viscoelastic Property
of Protein Adlayers

Figure 10 shows the schemes of the geometry and parameters
used to simulate the quartz crystal covered with a viscoelastic
protein adlayer, and the Voigt-based viscoelastic model
(infinite spring and dashpots model). Voinova et al. reported
that the measured $\Delta f$ and $\Delta D$ curves by the QCM-D technique
have been fitted by a Voigt-based viscoelastic model to
characterize the viscoelastic properties of the adlayers as
a Newtonian fluid as shown in Figure 10(a) [116,126,137].
The Voigt-model is assumed to be a linear spring in parallel to
the dashpot for the load shaft to simulate the viscoelastic behavior as shown in Figure 10(b). The viscoelastic property is represented by a complex shear modulus $G^*$ given by Eqn (12) based on the Voigt-based model:

$$ G^* = G' + iG'' = \mu_{ad} + i2\pi f\eta_{ad} $$

where $G'$ is the real part of $G^*$ (storage modulus), $G''$ is the imaginary part of $G^*$ (loss modulus), $f$ is the oscillation frequency, $\mu_{ad}$ is the elastic shear modulus, and $\eta_{ad}$ is the shear viscosity. The viscoelastic parameters, such as $\mu_{ad}$, $\eta_{ad}$, density ($\rho_{ad}$), and thickness ($d_{ad}$) of the adlayer, were determined using the density ($\rho_l$) and viscosity ($\eta_l$) of the

FIGURE 9 (a) $\Delta D-\Delta f$ plots of FBS adsorption on HAp sensors in different solvents (PBS, αMEM and DMEM) for 1 h. The arrows indicate the saturation direction in the plots. (b) possible schemes of FBS adlayer on HAp at 1 h. The adhesion amount of the proteins is in the order 10 mM carbonate solution > PBS > MEMs (αMEM and DMEM) and 120 mM carbonate solution. The $\Delta D_{sat}/\Delta f_{sat}$ value is in the order PBS > MEMs > 10 mM carbonate solution and 120 mM carbonate solution. (Reprinted with permission from [126], F. Höök, et al, Anal Chem 2001;73:5796–804. © 2001 American Chemical Society.)

FIGURE 10 (a) Scheme of the geometry and parameters used to simulate the quartz crystal covered with a viscoelastic protein adlayer having the thickness $d_{ad}$ in contact between the sensor surface and a semi-infinite Newtonian liquid. The adlayer is represented by $\mu_{ad}$, $\eta_{ad}$, and $\rho_{ad}$. The bulk liquid is represented by $\rho_l$, $\eta_l$. (b) Scheme of Voigt-based viscoelastic model with a parallel connection between the dashpot and spring. (Reprinted with permission from [126], F. Höök, et al, Anal Chem 2001;73:5796–804. © 2001 American Chemical Society.)
bulk liquid. The $\rho_1$ and $\eta_1$ values are fixed. The adlayer is situated between the QCM electrode and a semi-infinite Newtonian liquid under no-slip conditions. The changes in the $\Delta f$ and $\Delta D$ are fitted by the $\beta$ function as follows:

$$\Delta f = \text{Im}(\beta)/2\pi \rho_q$$

$$\Delta D = -\text{Re}(\beta)/\pi \eta_q$$

where $\beta$ is represented by Eqn (15) with $\alpha$, $\xi_1$, $\mu_{ad}$, $\eta_{ad}$ and $d_{ad}$ and $\xi_2$ and $\xi_1$ represented by Eqn (16) with $\mu_{ad}$, $\eta_{ad}$, and $\rho_{ad}$ as follows:

$$\beta = \frac{\xi_1^2}{2\pi \eta_{ad}} - i\frac{\mu_{ad}}{2\pi} \frac{1 - \alpha \exp(2\xi_1 d_{ad})}{1 + \alpha \exp(2\xi_1 d_{ad})}$$

$$\xi_1 = \sqrt{\frac{\rho_{ad}^2}{\mu_{ad}^2 + 4\pi^2 \eta_{ad}}}$$

where $\alpha$ is represented by Eqn (17) with $\xi_1$, $\xi_2$, $\mu_{ad}$, $\eta_{ad}$, and $\eta$, and $\xi_2$ is represented by Eqn (18) with $\rho_{ad}$ and $\eta_{ad}$ as follows:

$$\alpha = \frac{\xi_1^2}{\xi_2^2} \frac{2\pi \eta_{ad} - i\mu_{ad}}{2\pi \eta_{ad}} + 1$$

$$\alpha = \frac{\xi_1^2}{\xi_2^2} \frac{2\pi \eta_{ad} - i\mu_{ad}}{2\pi \eta_{ad}} - 1$$

$$\xi_2 = \sqrt{\frac{2\pi \rho_{ad}}{\eta_{ad}}}$$

The ratio of $G''$ and $G'$ can be calculated as a loss tangent delta ($G''/G'$) to evaluate the viscoelasticity of the adlayer as shown in Eqn (19):

$$\tan \delta = \frac{G''}{G'}$$

The $\tan \delta < 1$, $\tan \delta = 1$, and $\tan \delta > 1$ of the adlayer can be defined as elastic, viscoelastic, and viscous properties, respectively. In the energy dissipation process, the molecular structure of the adlayer changes with the oscillation (mechanical energy) to enhance the potential energy, which is an energy storage process. Next, the intramolecular and/or intermolecular interactions are not fixed to generate the sliding friction from the adlayer. This indicates that the elastic energy converts to molecular kinetic energy (thermal energy), which is a loss energy process. The elastic adlayer ($\tan \delta < 1$) induces the changed molecular structures with the mechanical energy that effectively returns to the original structure with a lower energy state. On the other hand, the viscous adlayer ($\tan \delta > 1$) induces the resistance force with molecular movements that effectively generates and converts to thermal energy.

Based on the Voigt-model, only a few studies of the viscoelastic property of the adlayer have been investigated, and the Ab [138] and laminin [139] adlayers on gold and Col (type I from calf skin) adlayer on poly(styrene) (PS) and oxidized PS (PSox) [140] have been discussed. The Ab had a $\mu_{ad}$ of 100 kPa, $\eta_{ad}$ of 6 mPa·s, and tan $\delta$ of 1.8 [138], whereas those of the laminin were a $\mu_{ad}$ of 7.6 ± 1.9 kPa, $\eta_{ad}$ of 1.83 ± 0.11 mPa·s, and tan $\delta$ of 7.7 [139]. Thus, the laminin, which is the ECM protein, clearly shows a more viscous structure on the gold based on the tan $\delta$ values. The Col shows a $\mu_{ad}$ of 125 ± 25 kPa, $\eta_{ad}$ of 4.0 ± 1.0 mPa·s, and tan $\delta$ of 1.15 ± 0.05 on PS and a $\mu_{ad}$ of 20 ± 10 kPa, $\eta_{ad}$ of 2.5 ± 0.5 mPa·s, and tan $\delta$ of 4.7 ± 1.5 on PSox [140]. Thus, the surface property also affects the viscoelastic property of the adlayer. The viscoelasticity of the adlayer on the bioceramics have not yet been reported. As compared with the viscoelastic property on the bioceramics, the evaluation of interfacial phenomena based on the Voigt-based viscoelastic model would be an effective tool.

### 3.2.4. Cell Adhesion Studies on the Hap

The adhesion behaviors of cells depend on the surface properties, such as topography, wettability, and charge, and on the protein adlayers [19–21]. The conformational change, denaturation, and RGD sequence of the adsorbed proteins on the surfaces govern the biocompatibility, which is attributed to the adhesion, proliferation, migration, and differentiation of cells [18,23]. The ECM with the RGD sequence affects cell adhesion [24]. The integrin of a cell binds to the RGD sequence in the ECM [25], and subsequent actin cytoskeletons are produced and the associated proteins form focal adhesion points on the surfaces [26]. With the interfacial reactions and morphological changes in Figure 1, the cells adhere and spread on the material surfaces. Thus, in situ monitoring of these interfacial phenomena with the initial cell behaviors on the surfaces is of great importance for controlling cell functions.

The protein adsorption and subsequent cell adhesion on the surfaces were reviewed by Anselme [21]. The adsorption of different amounts of FBS on nanophase ceramics caused different cell adhesion behavior [141]. The proliferation and mineralization of osteoblasts on a HAp sintered body with and without pre-adsorbed type I Col were investigated, revealing various phenotype and gene expression patterns that were different from those on PS dishes [28,29]. If the cell behavior is determined by the adlayers of the type I Col, the gene expression should be the same for coatings of the type I Col on the HAp and PS. Therefore, these results suggested that the cell functions were determined not only by the interfacial proteins but also by the substrate surface properties. The understanding of cell adhesion and spreading process on the interfacial protein layers adsorbed on a substrate surface was also of great importance.

Various physical parameters were measured to evaluate the interfacial phenomena with the initial cell adhesion,
such as resistance [68,142–147], impedance [148,149],
transient decay time, maximum oscillation amplitude
[150], and rheometer [163]. Li et al. recently described
the viscoelastic properties of a fibroblast cell monolayer
on gold using a thickness shear mode quartz crystal resonator
with a transfer matrix model [162]. They calculated the
\( \mu_{ad} \)
values of 21–39 kPa, \( \eta_{ad} \) values of 0.92–1.56 mPa·s, and
tan \( \delta \) values of 1.2–2.3 at a 5-MHz resonance frequency.
Fernández et al. investigated the viscoelastic properties
of a fibroblast cell monolayer with a rheometer and obtained
\( G' \), \( G'' \), and tan \( \delta \) values at 10 Hz of 400 Pa, 150 Pa, and 0.3,
respectively, indicating that the cell monolayer was
an elastic body [163]. Palmer et al. suggested that the elastic
shear modulus dominates the viscoelastic property of cells
due to their rigid cytoskeletons at lower frequencies, but
that the viscous modulus dominates the property due to the
cytosolic fluid at higher frequencies [164]. The actin
networks at the high frequency of 1 MHz showed a liquid-
like property and those at the low frequency of 10 Hz
showed an elastic property. The detected viscoelasticity
was found to be attributed to the measurable surface region
that depends on the excitation frequency. Thus, the QCM’s
high frequency caused the viscosity behavior of the adherent
cells due to the dependence of the actin network
property on the frequency. The different viscoelastic
properties of the adherent cells on the material surfaces
could be attributed to the different actin network structures
close to the surfaces.

The QCM-D technique based on the energy dissipation
shift [65,67,126,151] has allowed the monitoring of the cell
adhesion and spreading behaviors on PSox
[152–154,156,157], tantalum (Ta) [155–157], chromium
(Cr) [155], titanium (Ti), and steel [158]. The QCM-D
technique has also been used to understand the effect of
the pre-adsorption of proteins on the cell adhesion
[155–158], for example, on Ta and Cr with and
without the pre-adsorption of FBS [155] and on Ti, TiO2,
and steel with the pre-adsorption of Fn or Fgn [158]. The visco-
elastic property of the interface with the cell adhesion on
the HAp using the QCM-D technique should be
investigated.

Eqn (9) indicates that a higher \( \eta \) value for the adlayer
can detect a higher \( l \) region. The \( l \) of a resonating wave in
a culture medium at 37 °C with a \( \rho \) of 0.993 g/cm³ and an \( \eta \)
of 0.692 mPa·s [148] can be calculated to be ca. 100 nm at
15 MHz. The FBS adlayer with the viscoelastic values
measured shows an \( l \) of 100–200 nm at 15 MHz. The \( l \)
of the adherent cell layers with the measured viscoelastic
values is also 100–200 nm at 15 MHz. The values of \( l \)
were extremely lower than the actual height of cells, which
was on the order of micrometers as shown in Figure 11. This
suggested that the QCM-D measured the lower parts of the
cells close to the protein adlayer on the sensor surfaces.
Therefore, the interface between the material surface and
cell by the QCM-D effectively causes the differences in the
viscoelastic properties of the adherent cells. In situ moni-
toring of the pre-adsorption of proteins and subsequent cell
adhesion on HAp as compared with that on other surfaces
should be investigated.

HAp has a good biocompatibility for fibroblasts;
dowgrowth of the percutaneous device [165] and catheter
[166] is improved by using an HAp sintered body and
coatings. The biocompatible features are attributed to the
surface property of HAp. The pre-adsorption of proteins
and initial subsequent cell adhesion process on HAp using
QCM-D and confocal laser scanning microscopy (CLSM)
techniques have been reported [159–161]. The adhesion
and morphological changes of fibroblast NIH3T3 cells on
the HAp and PSox sensors with pre-adsorbed FBS have
been analyzed [159]. The \( \Delta f' \) and \( \Delta D \) curves on the HAp
showed a decrease in \( \Delta f' \) with increasing \( \Delta D \) for 80 min
and the subsequent increase in \( \Delta f' \) with decreasing \( \Delta D \),
while those on PSox showed a decrease in \( \Delta f' \) for 120 min
with increasing \( \Delta D \) for 50 min and then with subsequent
decreasing \( \Delta D \). The cells on HAp had rough fibrous
pseudopods and those on PSox had dense particulate
pseudopods. These different adhesion behaviors dependent
on the surface property were attributed to the cell—surface
interactions.

The pre-adsorption of three proteins (Ab, Fn, or Col)
and subsequent adsorption of FBS and adhesion of fibro-
blasts have been analyzed [161]. Figure 12 shows the
\( \Delta D - \Delta f' \) plots of the fibroblasts on the HAp modified
with the proteins for 2 h, and the tan \( \delta \) values and the aspect ratio
of the adherent cells on the surfaces at 2 h. Figure 12(a–d)
show the \( \Delta D - \Delta f' \) plots of the cells on the HAp modified
with the FBS, FBS-Ab, FBS-Fn, and FBS-Col for 2 h. The
Col-modified surface clearly showed an increase in \( \Delta f' \) with
a decrease in \( \Delta D \), while the other surfaces showed a
decrease in \( \Delta f' \) with an increase in \( \Delta D \) and a subsequent
increase in \( \Delta f' \) with a decrease in \( \Delta D \) that occurred at 75 min
for FBS, at 100 min for FBS-Ab and at 60 min for FBS-Fn.
The changes on the modified surfaces with ECM proteins (FBS-Fn and FBS-Col) indicate the interfacial cell—surface reaction such as binding to adhesion points during the fast stage. The $\Delta f$ and $\Delta D$ values at 2 h were $-10.3$ Hz and $+13 \times 10^{-6}$ on FBS, $-31.7$ Hz and $+21 \times 10^{-6}$ on FBS-Ab, $-3.4$ Hz and $+1.2 \times 10^{-6}$ on FBS-Fn, and $+54.8$ Hz and $-13 \times 10^{-6}$ on FBS-Col. The saturated $\frac{\Delta D_{\text{sat}}}{\Delta f_{\text{sat}}}$ values were 0.51, 0.14, 0.82, and 0.54 on the FBS, FBS-Ab, FBS-Fn, and FBS-Col, respectively, indicating the higher value of the ECM proteins (FBS-Fn and FBS-Col) with the binding between the modified surfaces and the cells. These results indicated that the different adhesion processes occur depending on the modified surfaces for 2 h, and the pre-adsorption of Col and Fn effectively affects the cell—surface interactions.

Furthermore, the different interfacial $\tan \delta$ and the morphology observed by the CLSM also depended on the surfaces as shown in Figure 12(e, f). The $\tan \delta$ values were calculated from the viscoelastic parameters based on the Voigt-based viscoelastic model, and were evaluated within
Chapter | 7.2 Application of a Quartz Crystal Microbalance with Dissipation

the detectable height of the D value [137]. The viscoelasticity (tan δ) and the spreading degree (aspect ratio) of the adherent cells on the surfaces modified with the ECM proteins showed higher values. The adherent cells on the surfaces modified with Col had expanded pseudopods, while those modified with Ab had round shape. This indicated that the interfacial viscoelasticity and the morphology clearly depended on the consisted proteins: the interfacial layer on the pre-adsorbed ECM proteins showed loose and flexible structures and uniaxially expanded morphologies. Therefore, the rearrangements of ECM and cytoskeleton changes at the interfaces would cause the different cell–surface interactions. The different cell adhesion process, viscoelasticity, and morphology depending on the surfaces were successfully in situ monitored and evaluated by the QCM-D and CLSM techniques.

4. SUMMARY

In situ monitoring of the interfacial layers formed between biomaterials and cells is crucial for controlling cell functions and understanding biocompatible phenomena. To detect and clarify the structural characteristics at the interface, a QCM-D technique is one of excellent in situ analytical methods. The QCM-D detecting the interfacial phenomena with protein adsorption and initial cell adhesion has been investigated. In the future perspectives for scientific challenges and opportunities, the exploration of the mutual interaction should be studied by developing novel detection techniques for the interface. Furthermore, superior biomaterials applicable for medical implants, biosensors and biochips for diagnostics, tissue engineering, bioelectronics, and biomimetic materials will be designed.

REFERENCES


Chapter | 7.2 Application of a Quartz Crystal Microbalance with Dissipation

Fohlerova Z, Skladal P, Turanek J. Adhesion of eukaryotic cell
wegener J, Seebach J, Janshoff A, Galla HJ. Analysis of the
Marx KA, Zhou T, Montrone A, McIntosh D, Braunhut SJ. A
Anticancer Diagnoses and Treatments Using Ferrite Nanoparticles and Bulk

Masanori Abe,* Tomoaki Ueda,† Takashi Nakagawa,** Mamoru Hatakeyama†† and Hiroshi Handa†††

* Professor Emeritus of Tokyo Institute of Technology, Japan, † Solutions Research Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan, ‡ Management of Industry and Technology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan, †† Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Chapter Outline
1. Introduction 577
2. High-Performance MRI Contrast Agent 578
3. Mediators for Self-controlled Induction Heating 578
4. Sentinel Lymph Node Mapping for Monitoring Cancer Metastasis 580
References 581

1. INTRODUCTION

Ferrite nanoparticles (FNPs) have a high biocompatibility and exhibit unique functionality when exposed to magnetic fields and electromagnetic radiations. Applying these features, some anticancer treatments have been developed using FNPs in vivo [1,2]. Among them, magnetic resonance imaging (MRI) contrast enhancement has reached a stage of standard clinical procedure [3]. It uses colloidal fluids, called Resovist®, Feridex®, etc., of superparamagnetic FNPs which are approved by U.S. Food and Drug Administration (FDA) for intravenous administration into human bodies. Other anticancer diagnoses and treatments using FNPs are in clinical test stage at highest. They include the following.

1. Magnetic drug delivery: Anticancer drugs are immobilized to FNPs, which are delivered to target tumors selectively using focused external magnetic field and sustained release is made. This allows reducing dosage of the drugs and thus suppressing the drug side effects [4]. However, high-performance magnetic drug delivery has not yet been realized, mainly because magnetic drug forces acting on FNPs become very weak when they are located in deep position.

2. Magnetic hyperthermia: As shown in Figure 1, FNPs are selectively targeted to tumors, either immobilized onto the surfaces of the tumors or uptaken by the cancer cells. In addition, they are induction-heated by an external alternating magnetic field to cause heat death of cancer cells selectively [5]. This suppresses unwanted heat damage to normal cells and allows noninvasive anticancer treatments. To realize the magnetic hyperthermia, however, we must resolve the problems remaining in attaining not only strong but also controlled heat rise by the FNPs. More expectation is placed on the magnetic hyperthermia at present, since it was found lately that anticancer immunity is enhanced by the hyperthermia; in animal experiments, tumors without injection of heat mediators were cured as well by the hyperthermia treatment [6]. It is suggested that the physiologically active substances (e.g., heat shock...
proteins) released from the cancer cells which underwent heat death enhance tumor immunity.

Utilizing particles and bulk ferrite, we have developed high-performance MRI contrast agent, fabricated thermal mediators for self-temperature-controlled hyperthermia, and proposed a novel type of “sentinel lymph node mapping” for staging metastasis of cancers. The following sections review our studies, explaining the background.

2. HIGH-PERFORMANCE MRI CONTRAST AGENT

The FDA-approved MRI contrast agents (e.g., Resovist®) are made of superparamagnetic particles (several nanometers in size) of spinel ferrites, Fe₃O₄, γ-Fe₂O₃, or the intermediate [3]. They enhance MRI contrast by shortening T2 relaxation time of protons nearby. They are coated with carboxyl dextran to form spheres of 18–65 nm diameter, which are uptaken by Kupffer cells in livers. Because cancer cells do not contain Kupffer cells, tumors are contrasted against normal cells containing Kupffer cells. Thus, the contrast agents are used to identify liver cancers.

Since brighter and clearer images are obtained by T1 shortening signals than by T2 shortening signals in MRI, we [7] have developed a novel T1-weighted contrast agent of FNPs. They were synthesized through a thermolysis of the iron–oleate complex organic solvent to spheres of 4.6 ± 0.5 nm diameter with narrow distribution. Then they were coated with citric acid by a two-step ligand exchange reaction to attain a high dispersibility in water.

Figure 2 shows that the T1-weighted images obtained by the citrate-coated FNPs were as bright as those obtained by Magnevist® (a T1-weighted MRI contrast agent of superparamagnetic gadolinium complex), and were much brighter than those obtained by Resovist®. This is because our novel FNPs have magnetization much stronger than Magnevist®, and have smaller particle size (4.6 nm) than Resovist® (18–65 nm).

Therefore, we have successfully fabricated T1 shortening contrast agent free from gadolinium, an element which may cause strong toxicity when solved out from the complex. Although small in number, deaths by dissolved gadolinium from the contrast agent have been reported. Developing gadolinium-free T1-weighted MRI contrast agent has been longed for. Our novel contrast agent of FNPs is promising to meet the needs. Also, the FNPs have an advantage that functional molecules can be easily immobilized onto their surfaces via citric acid to attain tissue specificity.

3. MEDIATORS FOR SELF-CONTROLLED INDUCTION HEATING

Usually, for magnetic hyperthermia FNPs smaller than several tens nm having single magnetic domains are used as heating mediators. Heat is generated by Néel relaxation losses (in spin dynamics) and Brownian relaxation losses (in mechanical dynamics) and secondarily by magnetic hysteresis losses [8]. We revealed that the FNPs exhibit the highest heating ability when their size falls in the range of 15–20 nm [9], as expected from calculation based on Néel and Brownian relaxations, as shown in Figures 3 and 4.

One of the major difficulties in developing magnetic hyperthermia is to strictly control the temperature to rise above the fatal temperature (ca. 42.5 °C) of cancer cells and yet keep it low enough to avoid thermal damages to normal cells. To solve this problem, we have fabricated smart heat mediators of a bulk ferrite in which the induced temperature rise is automatically stabilized by...
specific magnetic characteristics of the ferrite [10]. They are millimeter in size and of implant type which are inserted into targets.

To stabilize the induction heat rise, we chose a perovskite $\text{La}_{2/3}\text{Sr}_{1/3}\text{Mn}_{1-x}\text{Cu}_x\text{O}_3$ (LSMC) whose thermomagnetization curves sharply decrease to zero around their Curie temperature [11]. We prepared bulk samples by a polymerized complex method. As shown in Figure 5, their magnetization decreased sharply around the Curie temperature. Therefore, magnetic induction heating will be automatically balanced at the Curie temperatures; in the LSMC bulk samples, induction heating is caused mainly by the magnetic hysteresis losses and slightly by eddy current losses due to the conductivity accompanying the magnetization.

Figures 6 and 7 show that under an alternating magnetic field (1 MHz, 40 Oe) the temperatures of the samples were raised and balanced at temperatures roughly equal to the

**FIGURE 4** Heat power calculated to be induced by alternating magnetic field (900 kHz, 40 Oe) in ferrite nanoparticles. Results are plotted as a function of average particle size, assuming that the size is distributed with various values of standard deviation $\sigma$ [10]. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 5** Thermomagnetization curves obtained under 100 Oe static magnetic field for $\text{La}_{2/3}\text{Sr}_{1/3}\text{Mn}_{1-x}\text{Cu}_x\text{O}_{3-\delta}$ bulk samples [11]. Inset shows for example how Curie temperature $T_c$ of Cu044 ($x = 0.44$) sample was determined.

**FIGURE 6** Temperature rise profiles obtained for $\text{La}_{2/3}\text{Sr}_{1/3}\text{Mn}_{1-x}\text{Cu}_x\text{O}_{3-\delta}$ (e.g. Cu041 means $x = 0.41$) ferrite bulk samples under alternating magnetic field (1 MHz, 40 Oe) [11].

**FIGURE 7** Relation between balanced temperature $T_b$ and Curie temperature $T_c$ obtained for $\text{La}_{2/3}\text{Sr}_{1/3}\text{Mn}_{1-x}\text{Cu}_x\text{O}_{3-\delta}$ (e.g. Cu041 means $x = 0.41$) bulk samples [11]. Dashed line indicates $T_b = T_c$ relation.
Curie temperature [11]. Therefore, LSMC perovskite is a promising material for preparing heat mediators having desired self-controlled temperature.

4. SENTINEL LYMPH NODE MAPPING FOR MONITORING CANCER METASTASIS

Cancers metastasize to lymph nodes because cancer cells isolated from primary tumors drift along lymph flow. The first lymph node that filters the fluid draining away from the primary tumor, as shown in Figure 8, is called “sentinel lymph node (SLN).” If an SLN is not metastasized, cancer is not spread to downstream lymph nodes. Therefore, SLN diagnoses and treatments based on “SLN mapping” were established as follows [12]. A radioisotope marker is injected into a tumor and the isotope marker accumulated at SLNs is detected by a gamma ray counter. Then, the identified SLNs are surgically removed for biopsy to monitor metastasis of the cancer. If no metastasis is found in the SLNs, removal of the downstream lymph nodes is avoided and the patient’s QOL is improved much.

The SLN mapping has become standard treatment at present for breast cancers and melanomas and gaining popularity in other cancers of the stomach, colon, lung, etc. However, the isotope marker has a problem of strict law regulation, especially in Japan and Europe, and small medical facilities cannot afford. To solve the problem, SLN mapping by detecting FNPs by SQUID sensors has been proposed [13]. But SQUID sensors are expensive and complicated. Therefore, we are developing a novel type of SLN mapping based on “sonic-wave emission by magnetically stimulated particles,” a phenomenon which we found and called SEMP in short [14]. As shown in Figure 9, FNPs are injected into the tumor as a marker. The FNPs accumulated at the SLN are excited by an alternating magnetic field to generate sonic waves, which are detected by a microphone mounted on the bottom of the probe. We succeeded in detecting an “artificial SLN,” a polymer gel sphere (c.a. 3 mm in radius) which contained FNP fluid of Resovist® at an amount similar to that is accumulated in human SLNs.
REFERENCES


[11] Nguyen C, Pham QN, Hoang NN, Nguyen HL, Nguyen DT. Preparation and magnetocaloric effect of (La$_{0.67}$-$_x$Gd$_x$)$_2$Sr$_{0.33}$MnO$_3$ (X=0.1, 0.15) nanoparticles. Physica B 2003;327:214–7.


Chapter 8.1

Diesel Particulate Filters

Joerg Adler and Uwe Petasch
Fraunhofer-Institut fuer Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Chapter Outline

1. Introduction 585
2. Diesel Particulate Matter (DPM) 585
3. Limits 586
4. Working Principle of DPFs 588
5. Design and Properties of DPFs 590
6. Ceramic DPFs 596
7. Market Situation 601
8. Outlook 602
References 603

1. INTRODUCTION

Diesel engines have the highest efficiency of all combustion engines and are widely used in heavy-duty vehicles, agricultural machinery and stationary equipment. Their use in passenger cars has also risen greatly in Western Europe over the last few decades, already making up 50% of the market (approx. 7 M cars) in terms of new registrations in 2006 [1]. Rising concerns regarding fuel consumption and greenhouse gases should drive this trend further. Even in the US, where the share of diesel cars was less than 4% in 2005, a strong increase is expected in the future [2].

Extensive discussions revolving around particulate matter and particularly the harmful aerosol and particulate emissions from incomplete combustion processes (‘soot’) have made purification of diesel engine exhaust a priority in environmental laws and created public awareness of these problems. Factory equipping of diesel cars with particulate filters became technically and economically feasible in 2002. In 2007 the absolute number of ceramic filters in use was approximately 11.7 M. Introduction of the EURO 5/6 emission standards in the EU made equipping of new vehicles (cars, light-duty (LD) and heavy-duty (HD) vehicles) with high-efficiency diesel particulate filters (DPFs) in combination with NOx reduction systems mandatory. However, based on the environmental regulations already in place experts realize that nearly all diesel engines—from forklift trucks to diesel locomotives, from road sweepers to excavators and mobile cranes—in Western Europe and the US will have to be equipped with filters in the next few years.

Due to their commercial importance DPFs with a very wide range of materials (powders, fibers; ceramics, metals) and designs (deep-bed filters, surface filters) are now being developed and offered, with a large part of the alternative filter and material designs utilizing different regeneration principles. Typically monolith ceramic filters with parallel channels plugged at alternate ends are factory-installed in vehicles.

2. DIESEL PARTICULATE MATTER (DPM)

2.1. Origin and Structure of DPM

DPM is formed as a result of incomplete combustion of the diesel fuel in the engine and is composed of carbon and various accumulated compounds including water, inorganic oxides, sulfur-containing compounds and hydrocarbons such as the highly toxic polycyclic aromatic hydrocarbons (PAH). The exact composition depends on the fuel composition, the quality of the intake air for combustion, engine wear and the specific motor design and operating point.

The DPM is emitted from the engine as primary particles of size 5—20 nm and rapidly agglomerates to chains or clusters with typical sizes of 50—150 nm. Hence particles in the exhaust system are different regarding time and places. Specifically both the size and the composition of the DPM is influenced by the given engine, fuel and fuel additives [3—5]. After exiting the exhaust system the particles change again. Ultrafine particles of size <100 nm may coagulate or grow by condensation over a period of a few hours in the atmosphere.
The limits previously prescribed by emissions laws were primarily applied to the mass in the exhaust. The ultrafine particles <0.1 µm only make up a small portion of the overall particle mass, but make up the greatest share according to their numbers. Particulate matter (pollution) measurements in an urban location yielded a portion of 1% of the particle mass for size class 0.01—0.1 µm, but 73% in terms of particle count [6]. Because fine and ultrafine particles are deemed particularly harmful to health, a particle number-based limit was introduced with the EURO 5 emission standards.

Otto-cycle engines with direct injection also emit very fine particles [7] and have sparked discussions as to whether these emissions should also be regulated in the exhaust laws.

2.2. Effects on Health

The particulate matter (PM) is defined to be particles of size less than 10 µm (PM10) or 2.5 µm (PM2.5). It is difficult to establish relationships between methods to measure these particles, compliance with limits set for them and their impact because of the vague definition of ‘particulate matter’ which does not distinguish the various particle types and sources and hence does not allow for identification of the components that are particularly critical to health.

The effects of particulates and especially DPM on health have long been a controversial issue. A large number of studies carried out worldwide confirm that relationships between increased PM10 loads and consequences on health exist and that combustion aerosols with the highest toxic and carcinogenic potential play an important role in this (see, for example, summary report of WHO et al. [8]). Studies have shown that fine particulate matter (PM2.5) yields the most pronounced effects on health. Chronic effects of excessive long-term PM loads have been found to be more significant than acute effects of temporary load spikes.

The extensive medical research and measurements carried out over the last few years has strengthened the impression that DPM pollution is becoming the biggest health and mortality risk in industrialized countries. Driven by the nanotechnology wave of the last few years, more and more work is being done to determine the effect of technologically prepared and used nanoparticles on the human organism. The possibility of nanoparticles having an impact can no longer be ruled out, given the observed associated increase in brain diseases such as Alzheimer’s or Parkinson’s disease [9].

Even if the overall quantitative predictions of mortality rates due to particulate matter are definitely worthy of discussion, it is now indisputable that DPM, despite its low percentage in relation to the amount of overall particulate matter, represents the greatest risk to health and hence its emissions must be effectively limited.

Another important aspect is the direct effect of soot particles on the greenhouse effect. Soot particles in the atmosphere have 2000 times the effect (based on mass) of CO₂ [10]. Approximately 60% of these particles occurring worldwide originate from heavy-duty vehicles [11].

In diesel engine exhaust, apart from DPM, nitrogen oxides (NOx) form substances that are harmful to the environment and to health. Internal engine alterations can reduce particulate matter emissions, but at the cost of increased NOx emissions (so-called PM—NOx trade-off). Hence the laws limit both exhaust components simultaneously.

3. LIMITS

3.1. Emissions/Pollution

In the discussion of limits and the regulations prescribing them differentiation must be made between emissions laws and pollution laws.

Emission restrictions apply to the exhaust emitted by diesel engine systems and impose regulations on the manufacturer and, in part, the operator (owner). A clear measurable relationship between cause and effect can be identified here. However, one criticism is that filter certification and vehicle approval are carried out using test cycles that do not reproduce real driving conditions and emissions and hence higher emissions are possible in reality.

Pollution restrictions are focused on regulations prescribing limits in ambient air. Governmental authorities are responsible for monitoring and compliance. The widely varying local conditions (e.g. urban versus rural) and emission sources from which the pollution originates make the causes difficult to determine and effective restriction difficult to achieve.

In regulations on limits in the workplace pollution limits are also defined, but they usually allow the equipment/machines causing the pollution to be identified and the responsibility of the employer to be made clear. Governmental authorities have a monitoring function and can pass special technical regulations that the operators of equipment must comply with.

3.2. Emission Limits for Cars and Commercial Vehicles

The requirements of the EURO 5a standard for new car registrations in the EU became applicable on 1.1.2011 [12]. The limits were imposed on sport-utility vehicles (SUVs) and vehicles with four-wheel drive up to 2.61 t for the first time, not having been included in the previous regulations. Starting with EURO5b (introduced for type approvals between September 2011 and January 2013) a particle
number-based limit of 6 * 10^{11} particles/km will apply [13]. The EURO 6 standard will then come into effect on 1.9.2014 for type approval and on 1.1.2015 for approval and sales of new vehicles [14]. Since introduction of EURO 1 emission limits in 1992 the particle limits will thus have been lowered from 0.14 g/km to 0.005 g/km and the NOₓ limits from 0.5 g/km to 0.06 g/km.

The limits for heavy-duty vehicles are given in g/kWh following the EURO V (2008) to EURO VI standards (2013). A particle count-based limit is also under discussion. In addition starting with EURO VI there will be a limit for ammonia of 10 ppm in response to the widespread use of ammonia-based SCR-DeNOₓ technology.

In the US differentiation is made between the standards of EPA (Environmental Protection Agency), currently Tier 2, and the specific and usually more stringent requirements applicable in California and followed by other states (CARB = California Air Resources Board). In addition there are fleet mix requirements for manufacturers in various emission groups (‘bins’). Bin 8, the minimum standard, took effect in 2007. The complete CARB LEV2 programme has applied since 2007. The values generally correspond to those of EPA’s bin 5.

More stringent CO₂ emissions requirements are expected in the future, yielding new challenges for exhaust aftertreatment [15]. These consumption requirements will lead to a more pronounced trend toward use of diesel engines, including the exhaust aftertreatment strategies used up to now, with the downside of additional fuel consumption effects caused by exhaust aftertreatment or regeneration.

Diesel passenger cars play no real role in Japan, but the world’s most stringent limits for heavy-duty vehicles were introduced there in 2005 (PM: 0.027 g/kWh, NOₓ: 2 g/kWh) and lowered again in 2009 to correspond approximately to the EURO V limits. Regulations corresponding to EURO VI are expected to be introduced in 2016.

In newly industrialized and developing countries regulations limiting diesel particulate emissions and following the high standards of industrialized countries stepwise and usually with significant time delays are either in place or in the pipeline. In India and China for example, regulations analogous to EURO IV are planned for 2010 and 2012.

The emissions are determined gravimetrically in defined test cycles and expressed in mg/km (in g/miles in the US). A DPF is not explicitly prescribed. Internal engine alterations did allow some vehicles to comply with EURO 4 limits, but in general the EURO 5 limits cannot be complied without filters or DeNOₓ aftertreatment. Particle count-based limits will make so-called closed filters mandatory. Limits for direct injection petrol engines corresponding to EURO 6 in the EU and LEV 3 in the US will be specified by 2013 and may mean that particulate filters are introduced for a number of vehicles with these engines as well.

### 3.3. Limits for Nonroad Machinery and Stationary Equipment

There are special limits for nonroad machinery. Limits for mobile nonroad applications (construction machinery, forklift trucks, compressors) are regulated by EU directive 2004/26/EC and those for tractors are regulated by EU directive 2005/13/EC. Between 2011 and 2013, stage IIIb will be introduced with a particulate matter limit of 0.025 g/kWh, normally necessitating the use of a particulate filter. With introduction of stage IV in 2014 the NOₓ limits, staggered according to engine power, will become even more stringent, necessitating an additional NOₓ exhaust gas aftertreatment. Stage III/IV also sets a limit (of 25 ppm) for ammonia.

The limits of stage IIIb will apply to locomotives in 2012; stage IIIA has been applicable for inland ships since 2009 with particle limits of between 0.27 g/kWh and 0.5 g/kWh depending upon the capacity.

EU stages III and IV are harmonized with the American Tier 3 and 4 standards.

Stationary machinery is subject to national requirements. For example, the requirements in Germany (‘TA-Luft’) have become increasingly stringent since 2002, usually necessitating exhaust aftertreatment of large diesel engines.

### 3.4. Workplace Limits

Pollution regulations in the workplace represent a special case affecting different vehicles or machines depending on the application. In Germany and Switzerland the limits apply not to PM, but rather to elemental carbon (EC), thus eliminating the possibility of measurement errors through condensed water, sulfuric acid and hydrocarbons (HC); these volatile components are determined separately in the gas phase.

Because DPM is classified as ‘carcinogenic for humans’, the rule of minimization applies according to the principle of the ‘best available technology’ (BAT) [16]. In Switzerland diesel engine emissions are classified as carcinogenic substances in the limit list of SUVA (Swiss Accident Insurance Institute) and here as well the rule of minimization applies for occupational safety and health. Retrofitting of filters has been mandatory in underground construction and tunnel construction since March 2000, for large-scale construction projects since 1.9.2003 for engines > 37 kW and since 1.9.2005 for engines between 18 kW and 37 kW. The German TRGS 554 prescribes use of DPFs for operation of diesel engines in closed rooms and underground applications to reach a concentration of <0.05 mg/m³ (EC).
The future is expected to see regulations focusing on ultrafine particles to 200 nm and limiting particle counts.

In the US the OSHA (Occupational Safety and Health Administration) and the MSHA (Mine Safety and Health Administration) are the responsible authorities, the latter especially for mining. So-called threshold limit values (‘TLVs’ corresponding to the German MWC values) are published every year by the ACGIH (American Conference of Governmental Industrial Hygienists). In 1995/96 the limit was 0.15 mg/m³; later it was 0.05 mg/m³ (total DPM). In 2001 a limit of 0.02 mg/m³ (EC) was proposed, but it was retracted in 2003. Currently there are no legal limits for DPM in the general occupational safety and health regulations in the US.

In underground mining stepwise limits have been introduced since 2002; they were originally supposed to have reached 0.160 mg/m³ (TC = total carbon) by 2006, but they were retracted or delayed several times due to industry opposition. However, the MSHA did not introduce any limits for coal mines due to lack of suitable measurement techniques.

4. WORKING PRINCIPLE OF DPFs

4.1. Filtration of DPM

For filtration of DPM in the exhaust system both a filter and an apparatus or mechanism that cleans the filter are required because otherwise the backpressure in the filter will increase too steeply. The filter hence must also be able to withstand the conditions (e.g. high temperatures) occurring during regeneration.

Filtration or removal of particles from fluids (gases, liquids) is modeled according to the different mechanisms used (see Figure 1): sieve effect, depth filtration and filtration through filter cakes (‘floating media’ filtration).

Although filters with pore sizes in the nanometer range can be manufactured to remove the DPM using a sieve effect, the low permeability of such a filter leads to an extremely high backpressure in the exhaust system. Soot very quickly forms loose and relatively permeable layers (see Figure 2), allowing for very good filtration efficiencies (in most cases up to 99.9%) to be achieved quickly over the entire particle size range even with filters with micrometer-scale pores. Thus the filtration mechanism can be described to be a combination of depth filtration and floating media filtration, with the latter effect dominating after a short loading time and relatively thin filters (e.g. some 100 μm thick) being sufficient. These filters are hence also called wall flow or surface filters.

With coarse filters, e.g. felts or open-cell foams, DPM can also be removed through a pure depth filtration effect. At a constant filtration velocity and for a broad particle size distribution deep-bed filters exhibit good filtration efficiency below and above a specific particle size (so-called most penetrating particle, MPP) which is filtered relatively poorly. Cascading filters of differing fineness (MPP) can be used, for example, to clean the air supply to clean rooms (so-called HEPA (High-Efficiency Particulate Air) or ULPA (Ultra-Low Penetration Air) filters). The high permeability of these filters is a special advantage, whereas the required filter volume is a disadvantage. The permeability is also dominated more strongly by the filter structure than by the filtered particles during filtration and decreases much less quickly than is the case with wall flow filters. However, the maximum filtration efficiency also depends very strongly on the filter thickness. For a high filter thickness and adequate filter surface area deep-bed filters can also achieve very high filtration efficiencies (>90%) for DPM. Because the filtration and adhesion mechanisms are much more complex than in surface filters very large differences in filtration behavior as a function of details of the filter structure, soot particles, engine type and engine map can be expected.

The high filter surface areas and filter thicknesses required for high filtration efficiencies in deep-bed filters are not particularly compatible with the given boundary conditions in vehicles. The DPM deep-bed filters developed up to now hence have a much lower filtration efficiency than the above-mentioned wall flow filters (30–80% depending on operating conditions) and have only become established in niche markets. Introduction of particle count-based limits has intensified this situation.

4.2. Filter Regeneration

Filtered particles increase the filter backpressure. When a certain backpressure limit or the maximum amount of soot per filter volume according to the filter material (the ‘soot mass limit’ as described below) is reached the filter must be cleaned. Initiation of regeneration is controlled by a sensor that monitors the pressure drop in the filter as well as empirical values from driving cycles stored in the vehicle electronics. The pressure drop threshold for regeneration is determined as well by weighing the additional consumption due to the backpressure against the energy consumed during regeneration.
DPM is thermally removed, typically in periodic regeneration cycles. In the presence of air DPM has a spontaneous combustion temperature of approx. 600°C; a number of catalytically active substances (e.g. Cu, Fe, and Ce) can lower this temperature to approx. 300–400°C. Because these temperatures occur rarely in the exhaust system and only in certain parts of the driving cycle they must be generated actively to initiate burn-off.

When free oxygen radicals occur, soot oxidizes at lower exhaust temperatures (240°C and higher) than molecular oxygen. Free oxygen radicals are formed upon decomposition of the NO₂ formed from the nitrogen oxides in the exhaust on a special oxidation catalyst. This principle was developed by JOHNSON–MATTHEY and is called a continuous regeneration trap (‘CRT’). Once the so-called balance temperature is reached at least as much soot is burnt off as is emitted by the engine. However, this process requires low-sulfur fuel (≤50 ppm), reliable reaching of the balance temperature in the most commonly occurring operating conditions and sufficient NOₓ in the exhaust. Increased NO₂ emissions after the filter could be critical.

The choice of the regeneration process depends mainly on the position of the filter system in the exhaust system and the corresponding exhaust temperature. In the so-called underbody arrangement fuel-borne catalysts (FBCs), usually based on iron or cerium compounds, are frequently added as soluble compounds to the fuel and hence are separated uniformly with the particles on the filter. This results in very close contact between catalyst and particles, leading to very low ignition temperatures and uniform burn-off. The regeneration temperature is reached through a cyclic internal engine reinjection of fuel that burns on a noble metal oxidation catalyst (diesel oxidation catalyst, or ‘DOC’) positioned immediately in front of the filter and thereby increasing the exhaust temperature. Alternatively to reinjection, atomization or vaporization can be used to introduce fuel directly into the exhaust system in front of the oxidation catalyst; this eliminates the need for reinjection and may keep the particulate filter loading and hence thermal load low through more frequent regeneration [18,19].

In the close-coupled (CC) position near to the motor, higher exhaust temperatures frequently occur, making catalytically coated DPFs (CDPFs) also applicable here.

CRT is preferentially used in heavy-duty applications and for retrofitting. In principle the CRT effect also occurs on the oxidation catalysts used to increase the exhaust temperature and hence it is difficult to distinguish it clearly from the regeneration effects. Ultimately, however, the...
CRT effect is often not sufficient for achieving reliable and permanent filter regeneration in all operating conditions, making backpressure monitoring and forced regeneration through temperature increase necessary to prevent blocking of the filter.

Particles are burned off in an exothermic reaction, which can result in very high local temperatures in the filter. Following ignition the flame front spreads as a function of the loading, the ignition source and the exhaust conditions, besides the properties of the filter material (heat capacity and thermal conductivity) and the exhaust volume and temperature. The maximum thermal loading of a filter is reached when a fully loaded or possibly even overloaded filter is already undergoing regeneration and the engine is switched to idle or is completely switched off (so-called drop-to-idle scenario); temperatures much higher than 1000 °C can occur in certain locations in the filter.

In the retrofitting of older cars, intervention in the engine management is frequently impossible or no engine management exists to increase the exhaust temperature to the ignition temperature of the DPM. In such cases ‘open filters’ (with CRT catalysts) in which part of the exhaust is not filtered or is only coarsely filtered, but the filter is not blocked (i.e. completely clogged), were preferred. These ‘filters’ inherently have filtration efficiencies of <50% by particle mass. Depending on the engine outlet emissions, however, these filtration efficiencies may have been sufficient at the time to comply with the EURO 3 or EURO 4 limits.

With the help of burners or electric resistance heaters, filters can also be cleaned without catalysts. However, such heaters are technically complicated to implement and more suitable for special vehicles, large-scale machines or stationary diesel engines. The groundwork is still being laid in the development of alternatives to introduction of energy, specifically plasma and microwave processes, and it is not clear as to whether fundamental questions regarding functioning, applicability in mobile systems and system and operating costs can be solved to make application technically and economically feasible.

4.3. Ash Problem

The DPM also contains incombustible components originating from fuel and oil additives as well as salts from the ambient air and motor wear and remaining as ash in the filter after regeneration [20]. Ash is composed mainly of phosphates, sulfates and oxides of Ca, Zn, Mg and Fe, primarily CaSO₄ (anhydrite) and MgZn phosphate.

Ash particles can be accumulated in very different ways in different parts of the filter. With FBC catalysts and frequent forced cyclic regeneration, the ash particles presumably detach and concentrate in the low-flow (dead) zones of the filter, e.g. as ash plugs in the inlet channel ends of honeycomb filters. With catalytically coated filters and mainly constant regeneration, continuous ash layers form in the filter material or on its surface. In both cases the ash accumulation affects the filtration and backpressure behavior of the filter over its lifetime. Ash layers constrict the inlet channels and lower the permeability of the filter material and ash accumulation continuously reduces the overall filter surface area. Both of these factors increase filter backpressures and shorten regeneration cycles. At low ash quantities layered ash is advantageous because it can form a membrane and facilitate the depth filtration behavior of the filter material. At higher ash quantities accumulation of the detached ash in the form of plugs is better than layer formation in terms of the effect on backpressure [21].

Fuel-borne additives can contribute up to 2/3 of the total ash quantity. Since the introduction of these additives their effectiveness has been increased to a point where the required concentration in the fuel could be greatly reduced, e.g. to 3–5 ppm for a Fe additive. This additive only contributed 0.4 g of ash per 1000 km (total quantity: 0.8 g/1000 km) in a trial with a passenger car engine [22]. The ash quantity is otherwise mainly dependent upon the oil quality and the oil consumption of the engine. The average ash quantity for an HD vehicle is approx. 40 g per liter of filter volume after 6000 h of operation.

Ash deposits in the filter have a high pore volume (80–95%), but given the right composition and temperature (especially >900 °C) they have a strong tendency to undergo sintering and can corrode certain filter materials. The ash has a particle size of 0.5–50 μm. Compared with Zn ash, Ca ash appears to have a lower permeability and a greater sintering tendency.

A large part of the ash can be removed from the filter with water or chemicals, but this usually necessitates a service appointment and filter removal/exchange. The acceptance of such service intervals depends on the specific application and user. The first standard passenger car system from 2000 had a service interval of approx. 80,000 km, in the second generation it was approx. 120,000 km and directly catalyzed filters for cars later reached approx. 240,000 km.

A further possibility for decreasing the negative effects of ash on the filter life is through increasing or extending the filter inlet channels to increase ash collector volume, inlet channel cross-sectional areas and filter surface area. With a new generation of FBCs this resulted in an increase in ash-based service intervals to 390,000 km [22].

5. DESIGN AND PROPERTIES OF DPFs

5.1. Material Selection

The material used in a DPF must have the suitable chemical resistance and mechanical stability especially at high
temperatures to meet functionality and other requirements, particularly of thermal regeneration. The material should have a low Young’s modulus, a low coefficient of thermal expansion (CTE) and a high thermal conductivity to minimize the effects of thermal stress. Especially for applications in passenger cars the materials must be commercially available at low to moderate prices and lightweight. There is no material among known material groups that possesses all of these properties. The properties of different filter materials can only be compared directly if the materials have the same pore size distributions and pore volumes. Unfortunately, there is very little reliable data, especially regarding strength, in the literature. Most filters are processed using special methods into thin-walled channel structures, which cannot be measured using conventional standards; moreover the filters often differ in channel shape, size and wall thickness and hence cannot be compared properly. Therefore, Table 1 gives an overview of materials and their characteristic properties as fully dense materials only, which are used with appropriate porosities as DPFs.

A comprehensive report on material attributes concerning ceramic DPF materials can be found elsewhere [23,24].

5.2. Pore Size, Shape and Volume

The pore structure, i.e. pore size and pore size distribution, pore shape and number of pores (=pore volume) affect the filtration behavior as well as the permeability and material strength. Filters can be made from very differently structured porous materials such as sintered particles, short-fiber felts, long-fiber woven/knitted fabrics and open-cell foams (Figure 3).

For filters made from sintered particles the pore size is directly dependent upon the size of the sintered particles. Such filters typically have pore volumes of 40–50%, which can be increased to 60–70% if additional pore formers are used in the manufacturing process, although this is associated with a significant loss of strength. Typical DPFs have mean pore sizes of 10–15 μm and narrow pore size distributions.

### TABLE 1 Promising Materials for use in DPFs; the Properties of Pure and Dense Materials are Given Here. The High Filter Porosity Lowers the Thermal Conductivity, Young’s Modulus and Strength. Silicon is used in Combination with SiC

<table>
<thead>
<tr>
<th>Material</th>
<th>Density (g/cm³)</th>
<th>Cordierite</th>
<th>SiC</th>
<th>Silicon</th>
<th>Mullite</th>
<th>Aluminum Titanate</th>
<th>FeCrNi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature limit for application (air) (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrosion resistance*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*These properties cannot be compared directly. A rough estimate: ++ very good, + good, o neutral, – poor, --- very poor.

### FIGURE 3 Porous ceramic structures that can be used for particle filtration: sintered particles (left), fibers (middle) and open-cell foams (right). The size relationships can vary and are described in more detail below.
Filters made from fibers can have pore volumes above 90%, but they have a broad pore size distribution. Only a limited number of materials are applicable because the suitable fibers are either technically not available or difficult to process and prohibitively expensive (especially long fibers). Moreover wear of fine fibrous materials is classified as potentially harmful to health.

Open-cell foams manufactured industrially also have high pore volumes (85–93%) and uniform pore sizes (min. 100 μm). Finer open-cell foams have only been available thus far on the laboratory scale.

The technical availability of the filter materials is also restricted by the fact that they must be able to be manufactured in such a way that a high filter surface area is obtained in the smallest possible volume.

5.3. Backpressure

A low backpressure in the filter is very important for the overall effectiveness of the engine system. However, the entire exhaust system and the engine power must also be considered. In principle, however, any backpressure downstream of the engine leads to an increase in the fuel consumption.

The following factors contribute to the pressure loss for flow through the filter material: specific permeability of the material, thickness of the filter material and viscosity of the exhaust with a linear dependence on flow rate (exhaust volume/filter surface area). Added to this is a factor including the inertia coefficient of the material, the exhaust density, the thickness and the square of the flow rate (Eqn 1), so-called ‘DARCY–FORCHHEIMER equation’.

\[
\Delta p = \frac{\eta}{k} d \cdot v + \beta \cdot \rho \cdot d \cdot v^2 \tag{1}
\]

\(\Delta p\) = pressure drop [Pa], \(\eta\) = dynamic viscosity [Pa s], \(k\) = spec. permeability \(\left[\text{m}^2\right]\), \(d\) = thickness \([\text{m}]\), \(v\) = gas velocity \(\left[\text{m s}^{-1}\right]\), \(\beta\) = inertia coefficient \(\left[\text{m}^{-1}\right]\), \(\rho\) = gas density \(\left[\text{kg m}^{-3}\right]\).

The exhaust volume is determined by the engine size and the operating conditions as well as (through the exhaust temperature) the position of the filter and the thermal insulation between the engine and the filter. The viscosity and density of the exhaust also depend greatly on the temperature of the exhaust. Thus the above trend toward a close-coupled filter arrangement is relatively deleterious to the pressure drops.

The geometric design of the filter is included through the available filter surface area and the thickness of the filter material through which flow occurs. In principle the filter surface area must be large and the thickness low for a low backpressure to be achieved. Both of these factors have been nearly exhausted in wall flow filters (a few square meters of filter surface area in a volume of 2.5–3 l; approx. 0.35–0.4 mm filter wall thickness); for deep-bed filters a higher thickness is required and the inlet surface area is limited due to the limited available volume in the vehicle.

The material properties of specific permeability and the inertia coefficient are decisive for the pressure drop. Permeability is mainly a function of pore size, volume and shape. The shape also includes the effect of tortuosity, which describes the length and shape of the flow path through which the gas is forced. The inertia coefficient describes the additional increase in the magnitude of the pressure drop with flow velocity due to inertia effects.

A number of very complex methods can be used to predict or calculate the permeability. However, the equation from ERGUN (2) yields a rough estimate.

\[
k = \frac{e^3}{150 \cdot (1 - e)^2} D^2 \tag{2}
\]

\(e\) = pore volume \([-\)], \(D\) = avg. pore diameter \([\text{m}]\).

Wall flow filters have a mean pore size of about 10–15 μm at an open pore volume of 50%, for which reason the specific permeability is much (approx. 3 orders of magnitude) lower than, e.g. for open-cell foams with a pore size of 80–100 μm and 85–90% open pore volume.

These influencing factors apply to an unloaded filter; loading with particles causes the backpressure to increase pronouncedly. This loading effect is superimposed on the pressure drop according to the filtration mechanism. For surface filters the permeability of the soot layer formed is then the dominant factor determining the pressure drop of the overall system and falls to zero (i.e. the pressure drop tends to infinity) if the filter is not regenerated. For loading of fresh wall flow filters typical pressure drop curves such as that shown in Figure 4 arise. The brief initial phase corresponds to the primary penetration of particles into the volume of the filter and is associated with a greater decrease in permeability than is the case in the subsequent surface or cake filtration phase. The accumulated soot layers have a specific permeability of around 2–4 \(10^{-14}\) m². Over the course of filtration additional contributions to the pressure drop are yielded based on the channel cross section and filter length as a result of the soot layer (Figure 5) slowly growing inwards from the walls.

In the deep-bed filtration phase individual particles can pass through the filter, causing the filtration efficiency during this time to decrease slightly as a function of filter material and wall thickness.

The pressure loss limit at which forced regeneration is initiated can be determined from the engine manufacturer’s specifications and/or the filter loading capacity (see below). During regeneration the pressure drop nearly returns to the original value prior to loading if the soot in the filter is completely burned out. Incomplete regeneration can yield
a hysteresis curve, making it difficult for the loading condition of the filter to be clearly determined from the increase in pressure drop. A general upward shift in the baseline of the unloaded filter occurs because of the above-mentioned ash problems.

Deep-bed filters exhibit a slower increase in pressure drop and can become blocked after a longer loading time. In very coarse deep-bed filters or ‘open’ filters with low filtration rates the backpressure increases to an equilibrium level at which the particle deposition rate equals the particle removal rate.

Irrespective of the filter material and loading, geometry-dependent factors, e.g. inlet and outlet flow losses and channel friction, also contribute to the backpressure of a filter.

Typical pressure losses at an exhaust volume of 900 kg/h lie at around 3500–5000 Pa for an unloaded wall flow filter and 30,000–40,000 Pa for a loaded wall flow filter (7 g/l soot loading) [25]. This value is reached in passenger car wall flow filters after 500–1000 km depending on the engine operation (and soot formation).

**5.4. Filtration Efficiency**

The filtration efficiency is a key parameter of the filter and describes the relationship between particles in the untreated gas upstream of the filter and purified exhaust downstream of the filter. It is currently based on particle mass because these values are easy to measure and the corresponding limits have been prescribed. When particle number-based limits are introduced, the filtration efficiency will also be considered as the quotient of the particle count upstream and downstream of the filter. The efficiency of a filter can vary with time depending on the vehicle operation and loading condition and is hence determined for defined driving cycles.

The weighting of the filtration efficiency in the consideration of the overall system varies. The filter efficiencies required for compliance with the limits differ greatly depending on the untreated emissions; in some cases it is as low as 30–50%. On the other hand, today’s wall flow filters are capable of reliably reaching filtration efficiencies much greater than 95% over the entire particle size range, adding an environmental and prestige component to the prescribed limits. This weighting, in turn, depends greatly on the specific application (e.g. Original Equipment Manufacturer (OEM) installation or retrofitting, passenger cars or commercial vehicles, nonroad machinery, stationary equipment etc.).

**5.5. Load Capacity**

The load capacity describes the maximum amount of soot that a filter can accommodate without being destroyed in the subsequent cyclic regeneration. It exhibits a pronounced dependence upon the material properties of the filter (thermal expansion, thermal conductivity, heat capacity, thermal stability limit etc.) and is determined experimentally. Typical load capacities for cordierite wall flow DPFs lie at 6–7 g/l (soot amount/filter volume) and for SiC at 9–12 g/l [26,27]. The precondition for exhausting the soot loading capacity is that this amount of soot does not already cause the maximum allowable backpressure to be exceeded (see above).

The significance of the load capacity lies in the fact that it determines the number of cleaning cycles per a given duration of filter operation and represents the maximum allowable soot amount for safe operation. Because the cyclic cleaning process is initiated by a temperature increase, frequent cleaning leads to additional fuel consumption.

Uniform distribution of the particles over the volume is important for low stressing of the filter material during regeneration. In wall flow filters this depends strongly on a uniform flow over the filter channel faces. In homogeneous deep-bed filters, however, cross mixing in the volume is always possible and leads to a more uniform radial distribution of the particles.
5.6. Coating with Catalysts

5.6.1. Catalysts to Support Filter Regeneration

In some filter systems the catalyst is applied to the ceramic filter surface as an alternative to fuel-borne systems to reduce the soot ignition temperature, primarily to reduce fuel additive expenditures and lower the ash quantity (and hence increase the filter life). The (typically Pt−Rh) oxidation catalyst needed for increasing the exhaust temperature or for the CRT system is applied as a coating in the front segment of the filter to eliminate the need for a separate oxidation catalyst (so-called CDPF). Continuous regeneration occurs more frequently and active regeneration less frequently in combined catalyzed filters.

A normal washcoat and catalyst coating on a fine-pored wall flow filter material blocks the pore channels and lead to high backpressures. For this reason the pore size and the pore volume in the filter are increased (to approx. 20 μm and 60−65%) and, as far as possible, the catalyst coating is only applied to the pore channels while the permeability is maintained. However, the increase in pore size and volume considerably reduces the strength of the ceramic material and nevertheless the amount of catalyst that can be applied is limited.

Further development of the particulate filters or the strategies for reducing PM will involve the regeneration processes. The primary goal is to achieve continuous removal of particulate in the exhaust system under all conditions found in the driving cycle, thus eliminating the need for accumulation and cyclic cleaning in a filter medium.

Despite intense R&D efforts directed toward the catalyysts the particle ignition temperatures still have not been able to be reduced to such a point that the particles are completely oxidized in the exhaust system under easily achievable conditions. No surprising developments are expected to occur in the near future. Most of the exhaust aftertreatment concepts use precious metal catalysts or coatings which are far more expensive than the ceramic substrate itself. The price of platinum has tripped since 2001 to about 30 Euros/g (2007). A typical CDPF is coated with 3−4 g per liter of filter. Therefore focus must be on concepts with very low amounts of precious metals.

5.6.2. DeNOx Catalysts

Especially for the integration of DeNOx systems, use of the filter as a catalyst support is important as it allows for downsizing through combining of functions. The first sequential combination of DPFs and DeNOx systems in cars was introduced in 2006 for US bin 8 and EURO 5 [28,29] and consisted of four separate exhaust aftertreatment systems: oxidation catalyst, lean NOx trap, DPF and SCR catalyst. This was an indication of the important role integrated systems would play in the future.

Combining of the filter and the DeNOx catalyst requires a look at the different competing DeNOx strategies:

- Selective NOx reduction with fuel using a lean NOx catalyst (LNC)
- Selective catalytic NOx reduction (SCR) with ammonia
- NOx storage and subsequent regeneration using a lean NOx trap (LNT).

Exhaust gas recirculation (EGR) and internal engine techniques are also available for reducing NOx emissions.

Of the catalytic methods SCR and LNT today appear to have the most potential for achieving the limits prescribed for the future. The LNC method has a relatively limited selectivity of reaction and the catalysts have limited working temperature ranges, but it is associated with the lowest expenditures.

The SCR technology is based on very successful DeNOx systems for stationary applications such as coal-fired power plants and has been adapted to exhaust aftertreatment for diesel engines. It is now the leading DeNOx technology for exhaust aftertreatment of diesel engines. The efficiency can reach 90−95% and above with suitable control and design adapted to system conditions. However, the supply of ammonia represents a challenge. Over the last few years a urea solution, which must be provided in a separate tank in the vehicle, has become established as an ammonia source. The urea is decomposed and hydrolyzed to form ammonia, which is then converted on the catalyst by NO and NO2 to nitrogen and water.

V2O5 on TiO2/WO3, Fe-, Cu- and La-doped zeolites and zirconia-based mixed oxides [30] are being used or developed as catalysts which differ in terms of performance, thermal stability and sensitivity to sulfur and HC poisoning. Some of the catalysts are manufactured as fully extruded components, but most are applied as coatings on cordierite or metal substrates.

The LNT technology is based on modified three-way catalysts (TWCs) containing additional special alkali/earth alkali storage materials. NOx is stored on the alkali and earth alkali metal oxides in the catalyst during the very economical lean engine operation. The NOx storage material is then regenerated at periodic intervals by brief injection of excess hydrocarbons into the exhaust. Typical operating cycles are 60 s lean operation for NOx storage and 1 s ‘enriched’ operation for NOx reduction. The efficiency, at about 70%, is lower than for the SCR technique.

A significant problem of the lean NOx trap is its susceptibility to sulfur poisoning and the limited thermal stability, besides the noble metal content (e.g. 10−12 g for a 2 l engine).

The LNT technology is expected to hold greater potential for application in compact car and LD engines,
whereas the SCR technology should predominate in large engines. A combination of LNT and SCR in front of the DPF in which ammonia is generated in the LNT and greater NO\textsubscript{x} reduction is possible without the need for the elaborate ammonia dosing is currently under development [31]. Up to now SCR has been used nearly exclusively, although in some cases LNC has also been used, in the HD and non-road segments. Various sequential arrangements of DeNO\textsubscript{x} systems and DPFs downstream of the DOC are possible, each with advantages and disadvantages [32]. In general DeNO\textsubscript{x} systems also lead to higher costs for exhaust aftertreatment systems because of the additional catalysts that are required and the complexity of the overall systems. Hence direct coating of DeNO\textsubscript{x} catalysts on the particulate filters is attractive for the downsizing of future systems.

In 2003 TOYOTA introduced the so-called DPNR system, which consists of a PM trap made of cordierite coated with platinum and a NO\textsubscript{x} storage catalyst [33]. DPNR systems were in use in more than 100,000 vehicles as of the end of 2006. Improvements were reportedly obtained through combination of a sulfur trap catalyst in front of the DPNR and a CeO\textsubscript{2}–ZrO\textsubscript{2}–Y\textsubscript{2}O\textsubscript{3} washcoat support to prevent the sintering of the Pt catalyst [34].

A more recent LNC system with DOC–DPF–DOC for medium vehicles uses a special noble metal filter coating with a special oxide washcoat to reduce the poisoning tendency [35]. For HD applications additional SCR coatings (Cu zeolites) on cordierite filter substrates are increasingly being used upstream or downstream of the standard SCR catalyst [36–38] to improve the DeNO\textsubscript{x} efficiency at the same unit size and to lower energy requirements for active filter regeneration for certain constellations. The prerequisites for future use are a high catalyst thermal stability and an optimized filter coating for a low backpressure.

### 5.7. Mechanical Stability

In mechanical loading differentiation is made between loading during transport/support/canning and loading in operation. The filters experience the highest static loads during canning. The methods are the same as used for catalyst supports. Pressure is exerted on the outer casing and for some methods the face.

In operation the filter is subjected to loads occurring during driving (acceleration, vibration) and thermomechanical loads during regeneration. The mechanical stability of the filter in the presence of these loads is the prerequisite for use.

In principle the material strength of a denser filter with less porosity is considerably higher than that of a material with larger pore sizes and volumes; this makes wall flow filters superior to many alternative deep-bed filter materials. However, the strength of a filter in the overall system must be considered and strongly depends on the design and the peripheral equipment/mounting/support. Despite their higher CTEs, metal filters possess advantages in terms of mechanical stability when embrittlement can be lastingly avoided.

### 5.8. Heat Capacity and Regenerability

The heat capacity describes the amount of energy required to change the temperature of (i.e. heat/cool) the filter. The weighting of the heat capacity for a DPF system depends on the regeneration strategy, the thermal conductivity and the load capacity of the filter. At high stability and conductivity levels, as found in SiC, a low heat capacity is advantageous. The filter can be heated rapidly, keeping the thermal gradients and energy requirements low for thermal regeneration. In cordierite filters the heat capacity is increased through increasing of the mass in the filter because this allows temperature peaks to be reduced.

Many materials show negligible differences in volumetric heat capacity and hence the type of material has no great influence. The high open porosity lends deep-bed filters a relatively low heat capacity, which is an advantage for filter materials with high thermal conductivities.

The regenerability is important for the overall function of the filter because nonuniform and incomplete filter regeneration can result in overloading and, in turn, destruction of the filter.

### 5.9. Filter Manufacturing Costs

The cost factor for an introduced system is obviously weighted very heavily. Shortly after introduction in passenger cars an uncatalyzed, averagely sized OEM car wall flow filter made from SiC cost between 120 and 160 Euros, approx. 25% of the cost of the overall filter system (excluding canning and oxidation catalyst); bigger ceramic filters for commercial vehicles or nonroad applications cost more than 300 Euros.

However, the cost consideration must include the costs of the overall system, i.e. with the catalyst/coating, canning, suspension, sensors, and regeneration system. As examples, more effective exploitation of catalyst coatings and consistent employment of the noise reduction potential can generate a considerable cost advantage.

For ceramic shaping the complicated and (in the case of SiC) wear-prone extrusion technology and the expenditures needed for cell plugging (and the above-mentioned after-treatment) are heavily weighted. For SiC and mullite there is an additional need for gluing of segments.

Metal filters are less attractive due to the high price of powders and the higher weight in comparison with
ceramics metal filters. The raw material costs for SiC filters are 3–8 Euros/kg and for cordierite less than 2 Euros/kg.

5.10. Volume and Mass

For the above-mentioned reasons of the pressure drop/filtration efficiency and the limited available space in mobile applications attempts are being made to introduce the largest possible filter surface area in the given (smallest possible) volume. Large filter volumes also increase the canning requirements. For wall flow filters very small channels with small wall thicknesses can be extruded to achieve a large filter surface area (e.g. 0.8 m²/l for a 200 cpsi filter), allowing the typical car filters to provide enough filter surface area with 2.5 l (5.66" diameter, 6" length) to maintain a low backpressure.

Herein lies one of the main disadvantages of deep-bed filters—it is very difficult to get the filter thickness necessary for good filtration efficiency into a small volume and still provide the high filter surface area required for a small pressure drop. Complex geometries mean additional expenditures for tailoring and canning of the shaped foam part.

The volume problem is the biggest challenge in passenger cars; in commercial vehicles it is of lesser importance; for nonroad and stationary equipment large filter volumes are not as much of a problem. All DPFs also contribute in a non-negligible way to noise reduction, freeing up the space of silencer components (e.g. front exhaust silencers).

The weight of a DPF is critical for mobile applications, mainly in terms of the efforts required for secure mounting/support of the DPF system. Metals have the disadvantage of a high density (density of steel is about twice that of SiC). The heat capacity of the filter (see Table 1) is also included through the weight. The weight is about 0.7 (SiC) or 0.5 (cordierite) kilogram per liter of filter.

5.11. Flexibility of Shaped Parts

A large range of achievable filter sizes and outer geometries are needed to accommodate the large number of existing engines and automobile types. In segmented SiC wall flow DPFs these outer geometries are prepared by mechanical machining of blanks at high amounts of waste (up to 20%) and high costs. The alternative approach of adaptation of the extrusion heads would also generate a large range of products at a high cost.

The drawback of extrusion is the limited geometric variability in the third spatial dimension, making it very difficult to produce complex contours in shaped parts. Additive manufacturing processes for DPFs with complex shapes, e.g. using Desktop printing or screen printing techniques, are not relevant for mass-produced components due to the insufficient productivities and prohibitive costs.

Metal filter materials are first made into semi-finished products and then shaped into filter elements (e.g. wound or pleated) and welded. This yields greater flexibility in design of complex filter geometries.

5.12. Simulation Tools for Development of DPF

Computational methods are valuable tools for evaluation and optimization of DPF. At first work was concentrated on the pressure loss behavior and the thermal stresses occurring during regeneration of DPFs e.g. [39–41]. Recently, nearly all effects connected with aspects of DPF have been topics of simulation attempts at the different scales from the filter material to the whole exhaust system; like e.g. asymmetric cells, soot morphology effects, ash behavior, catalyst function, and flow maldistribution effects see e.g. [42–45]. Different program suites are available now, both at scientific and commercial scale.

6. CERAMIC DPFs

6.1. Design and Manufacture of Ceramic Wall Flow Filters

Ceramic parallel-channel monoliths exhibiting a high surface area per component volume can be manufactured using extrusion. Such honeycombs have been standard in cars as substrates for the familiar oxidation catalysts since the seventies. For this, ceramic raw materials are mixed with plastic binders and pressed through the extrusion heads of continuously running extruders to form quasi-endless parallel-channel monoliths. These are then cut to length and fired to remove the binder and produce the desired ceramic material. Different cross-sectional shapes can be produced, but the variance is limited to the cross-sectional shape. Cell sizes and wall thicknesses are given in CPSI (cells per square inch) and mil (inch/1000). Extremely fine cell sizes (to 1600 cpsi) and wall thicknesses down to 2 mil (approx. 50 μm) are technically feasible, but irrelevant for use as DPFs.

In building on this technology, filters can be produced if the channels are sealed at alternate ends (so-called plugging). Thus arise inlet and outlet channels, and the exhaust flowing in is forced through the (open-cell) wall [46]; see Figure 6. DPFs with cell sizes of 170–400 cpsi and wall thicknesses of 350–400 μm have thus far been able to be manufactured. Typical car filters have diameters of 5.66” and lengths of 6”–9” (corresponding to volumes of approx. 2.5 l–3.7 l). Such filters possess a filter area of about 1.9/2.8 m² (200 cpsi) resulting in filtration velocities
of a few cm/s. Filters for commercial vehicles can be considerably bigger depending on the exhaust volume.

The channel cross sections are usually square, but can also be triangular [47] or hexagonal. In more recent versions larger inlet flow channels and smaller outlet flow channels (asymmetric filters) are manufactured to increase the ash collection volume in the inlet flow channels [48–50]; see Figure 7.

6.1.1. Wall Flow Filters with Membranes

Asymmetric ceramic membranes have been used in liquid filtration and dust collection from hot gases for a long time now. A thin ceramic coating with very fine pores is applied to a coarse-pored substrate to enable surface filtration and limit the backpressure of the fine-pored ceramic.

Initial trials using an RSiC membrane as a DPF based on tape casting technology with the goal of increasing the filtration efficiency for ultrafine particulates are described [51]. More recent developments using a membrane in the inlet or outlet channels of honeycomb wall flow filters reduce both the depth filtration phase during initial loading and the pressure drop hysteresis in incomplete regeneration [52]. One example is a 50 μm membrane with a pore size of 2 μm and 48% porosity on a 250 μm SiSiC substrate with a pore size of 14 μm and 42% porosity [53].

A membrane with similar properties is also yielded by a special catalyst coating, e.g. a 10 μm coating of 300 nm CeO₂ particles on SiC (250 μm wall thickness, 11 μm pore size, 42% porosity) [54] or a 20 μm coating of an 80 nm/500 nm mixture of SiC nanoparticles with 60% porosity on a conventional 300 μm substrate with 42% porosity [55]. Another advantage is the closer contact of the soot particles removed from the soot layer with the catalyst [56]. However, a membrane with a considerably higher pore volume as described in the literature [57] would be more favorable.

In all cases a membrane allows for variable engineering of the filter for the respective application because the filter properties (for the membrane) and the load-carrying capacity and robustness of the filter material etc. (for the substrate) can be optimized separately. In this way the soot mass limit and the regeneration efficiency can be improved in catalyzed filters [58,59]. A disadvantage is the added expenditures for manufacturing the membrane due to the need for separate steps for application and sintering of the membrane.

6.1.2. Cordierite Wall Flow Filters

Cordierite is a Mg–Al silicate made from the naturally occurring raw materials kaolin and talc. Cordierite in the form of a porous ceramic material is only produced upon sintering at approx. 1250 °C in air. Cordierite honeycombs have long been used as catalyst substrates and hence there is a wealth of experience in their fabrication and processing and the corresponding production capacities are available. Catalyst substrates have been optimized to have extremely low CTEs in a direction parallel to the extrusion direction; this requires use of very special raw material qualities (especially in terms of purity and particle structure) and fabrication processes. In use as DPFs the pore size and structure had to be tailored through addition of special burnout agents.

Already in 1978, a wall flow filter made from plugged parallel-channel monoliths was developed for diesel particulate filtration [60]. The main advantage of cordierite is its extremely low CTE (to 0.4 * 10⁻⁶ 1/K), maintaining low thermal stresses during regeneration and making cordierite filters able to be manufactured as compact monoliths, i.e. not segmented like SiC filters are. Another key positive aspect is the low manufacturing costs resulting from the relatively inexpensive raw materials and the low costs for sintering. The starting raw materials have a natural
plasticity and low hardness, keeping the wear low during extrusion.

Significant disadvantages of cordierite for use as a DPF are the low thermal conductivity and the limited high-temperature stability. Both lead to temperature spikes at high soot loads during regeneration and can potentially result in formation of molten phases and local destruction of the pore structure. Another problem is the tendency of cordierite to react with Ca–Mg–Zn phosphate ash at temperatures >1000 °C [61,62].

The first DPFs of the seventies and eighties had a pore size of 12–13 μm and a pore volume of 50% and were mainly employed in special nonroad applications (e.g. for forklift trucks). DPFs began to be used as standard features of passenger cars in California between 1985 and 1987, but they had to be taken off the market due to their limited regenerability.

Primarily driven by the potential cost advantages, constant efforts were made to improve the material or reduce the regeneration loads, e.g. to optimize the heat capacity, pore structure, channel structure and filtration behavior [50,63,64]. For the above reasons cordierite has hardly been used in cars up to now, although it has found widespread use in HD applications.

Efforts made regarding the cordierite filters were accompanied by concentrated efforts to develop alternative materials. Besides SiC and aluminum titanate (see below), the group of sodium zirconium phosphates (SZPs) were focused on for a long time due to the similarly low CTEs as that of cordierite combined with a higher thermal stability [65]. However, the negative aspects of a similarly poor thermal conductivity and the complicated and expensive manufacturing thus far have prevented industrial use of SZP filters.

### 6.1.3. Aluminum Titanate (tialite) Wall Flow Filters

Aluminum titanate ceramics are commercially used in the casting industry. In the automotive industry they are also sporadically used as claddings of port liners and spacer segments between catalyst substrates. The outstanding characteristic of aluminum titanate ceramics is their excellent thermal shock resistance, the result of the very low thermal expansion coefficient and Young’s modulus. The cause is the pronounced anisotropy (directional dependence) of the thermal expansion of the single crystallites, leading to a network of microcracks in the polycrystalline material. Through this, however, the mechanical strength of the aluminum titanate ceramics is poor. This is exacerbated by further reduction in load-carrying capacity by the porosity required for filtration.

Pure aluminum titanate decomposes into its components Al₂O₃ and TiO₂ between 800 °C and 1300 °C, but can be stabilized using various additives. In 2005 an aluminum titanate-based DPF material was introduced especially for applications in light-duty vehicles [66]. The material is composed of approx. 70% aluminum titanate and contains Ca/Sr feldspar and mullite. The published properties include outstanding thermal shock characteristics, high-temperature stability and chemical stability, but an even lower mechanical strength than that of cordierite [67,68]. The soot mass limit was given as approx. 8 g/l [69]. The filters have been used commercially in passenger cars since around 2007 with asymmetric cells (asymmetric cell technology, or ACT) and catalytic coatings in the close-coupled arrangement [70]. Thanks to optimization analogous to that carried out for cordierite specialized materials with improved backpressure or soot mass limit are now available [71].

An interesting alternative technology for manufacturing tialite filters through coating and subsequent sintering of corrugated paper filters is described by Franz et al. [72]. The process features a large range of possible geometric designs and is expected to be commercialized on a large scale.

### 6.1.4. Silicon Carbide (SiC) Wall Flow Filters

A number of different grades of SiC ceramics are available for industrial applications. The particles are sintered using different processes and mechanisms, yielding widely varying manufacturing processes and properties. Some grades, e.g. recrystallized SiC (RSiC), clay-bonded SiC (KSiC, CBSiC) or reaction-bonded SiC (RBSiC), possess inherent porosity; others, including sintered, liquid phase-sintered and silicon-infiltrated SiC (SiSiC, LPSSiC and SiSiC) were developed as dense construction materials. In principle all of the mentioned SiC variants can also be produced with controlled porosity to enable them to be used as DPFs.

As can be seen in Table 1 SiC has a high thermal conductivity, moderate thermal expansion and outstandingly high thermal and chemical stabilities, all of which are utilized to varying degrees in the above-mentioned material variants. The thermal expansion is higher than that of cordierite and aluminum titanate, posing a detriment to DPF application as in combination with a high stiffness (Young’s modulus) it can lead to high thermal stresses during regeneration. For this reason SiC DPFs must be segmented; i.e. smaller parallel-channel monoliths (e.g. with 35 × 35 mm cross sections) are joined with a cement having a low modulus of elasticity [73,74] to dissipate the stress (see Figure 8).

#### 6.1.4.1. Recrystallized SiC (RSiC)

RSiC is a pure SiC ceramic material formed by heat treatment of shaped parts containing bimodal distributed SiC particles at a high temperature (>2200 °C) in protective gas atmosphere. The fine SiC particles sublimate and recondense on the contact points of the coarse particles,
resulting in autogenously bonding of the coarse grain framework and producing a porous material. The pore size is controlled through selection of the grain size of the coarse fraction, use of special additives and the heat treatment regime. RSiC was originally developed for use as kiln furniture (beams, shelves and capsules) for fast firing of porcelain.

The company IBIDEN began developing high-porosity RSiC for filtration applications in 1985 [75]; the first DPF honeycomb filters were reported in 1993 [76] and began being used standard in passenger cars in 2000 [77]. The pore size was 10 μm and the pore volume was 50% for regeneration with a fuel-borne additive. Now coarser and more porous variants are used for catalytic coatings [78].

The binder phase can be modified with fine particles or special additives to improve the material strength and toughness [79,80].

Almost at the same time in Denmark relatively coarse and thick-walled RSiC DPF honeycombs were developed with a relatively wide range of properties and geometries mainly for nonroad and retrofitting applications [81,82]. These honeycombs feature larger and circular segments. Special processes for plugging and the adhesive putty were also developed and employed.

Recrystallized SiC can also be produced by means of special microscopic surface structuring for better anchoring of catalytic coatings [83].

Another interesting material is a RSiC material in which, unlike in the above manufacturing process, silicon and a carbon additives are presintered to produce SiC, which then recrystallizes in a subsequent high-temperature heat treatment [84]. Boron and other additives were used to adjust the electrical resistance to regenerate the filter through direct electrical resistance heating [85]. The target application areas were stationary equipment and large engines on ships and locomotives, but development/fabrication was discontinued in 2003.

6.1.4.2. Reaction-Bonded SiC

Secondary bonding of SiC from the reaction of Si with C or bonding by pure silicon represents an alternative to RSiC. In 2000 a DPF material in which the SiC framework was bonded with silicon was introduced [86]. Because the thermomechanical properties (thermal expansion and conductivity) are similar for silicon and SiC (see Table 1) a porous Si–SiC produces similar properties to those of RSiC [87]. The silicon bonding yields lower thermal and chemical stabilities than pure SiC and the thermal conductivity of Si–SiC is lower than that of RSiC because of the conductivity barrier effect of the Si–SiC grain boundaries [88], but extensive applications have thus far demonstrated adequate performance data for standard use in passenger cars [89,90]. At approx. 1600 °C the sintering temperature is lower than for RSiC, but sintering must be done in a protective gas.

If Si3N4 and carbon additives are used the secondary SiC bonding of a SiC particle framework can be achieved at lower temperatures and pore parameters typical to DPFs (e.g. pore size of 12 μm and pore volume of 43%) can be obtained [91,92].

SiC DPFs can also be manufactured by reaction of a mixture of Si powder and Al and carbon additives, although reaction of the relatively coarse silicon powder results in coarse pores. Various starting powders can be used to adjust the pore size between 7 μm and 22 μm and obtain a comparatively high pore volume [93,94].

6.1.4.3. Clay-Bonded SiC Ceramics

Clay-bonded SiC ceramics can be sintered in air at temperatures of up to 1350 °C through addition (approx. 20%) of clay or synthetic SiO2 formers and a flux and are hence much less expensive than the materials requiring the expensive protective gas technology. During sintering the SiC surface oxidizes and a glassy silicate binder phase interspersed with mullite or other crystalline phases according to the additives used forms between the particles. There has been a great deal of testing of the corresponding commercial filter materials for hot gas filtration, particularly in test power plants with pressurized fluidized bed combustion at operating temperatures of up to 850 °C [95].

DPFs made from these materials were also developed [96] and used mainly for nonroad applications [97]. The materials have a low thermal conductivity and may have to be segmented depending on the loading conditions.

6.1.4.4. Liquid Phase-Sintered SiC (LPS-SiC)

The relatively new variant of SiC known as liquid phase-sintered silicon carbide (LPS–SiC) can be sintered using a small amount (approx. 4%) of special additives (e.g. alumina-yttria). This results in doped SiC–SiC bonds and thin grain boundary phases, e.g. composed of the very stable
ytria-alumina garnet (YAG) material; dense LPS–SiC grades are characterized by high strengths and toughness levels. With coarse SiC grains and special sintering techniques (<2000 °C, protective gas) a high-porosity filter material with a narrow pore size distribution can be manufactured [98]. A DPF made from this material is in commercial use for nonroad applications [99].

6.1.5. Silicon Nitride (Si₃N₄) Wall Flow Filters

Si₃N₄ can also potentially be used for DPFs. Its advantage of a low thermal expansion, which is slightly lower than that of SiC, is countered by the moderate thermal conductivity and the relatively high costs of the Si₃N₄ powder. This has resulted in the manufacture of lower-cost ceramics by nitriding of shaped silicon parts at around 1400 °C. This reaction-bonded silicon nitride (RBSN) is porous a priori. In 2003 RBSN DPFs of high porosity (60% at 9 μm pore size) were manufactured and used and were tested in the form of unsegmented honeycombs. They were reported to exhibit high thermal shock resistance and strength. Needle-like whiskers on the cell surfaces should result in loose packing of the detached soot particles and lessen the pressure drop during loading [100]. However, detachment of whiskers from the filter and contamination of the exhaust with the whiskers can lead to harmful secondary emissions as whiskers of certain fineness are classified as carcinogenic. There has been no known commercial standard use up to now.

6.1.6. Mullite (Al silicate) Wall Flow Filters

Mullite also has a low thermal expansion (similar to that of SiC) and high strength and corrosion resistance. It exhibits the disadvantage of a low thermal conductivity and the advantages of inexpensive raw materials (e.g. clay) and sintering in air. High-strength mullite ceramics with a needle-like crystalloite structure must be produced with a high content of glassy phase, which is detrimental to the high-temperature stability. A novel approach has been developed to manufacture DPFs by recrystallizing presintered honeycombs made from clay and alumina by catalytic aftertreatment in fluorine-containing gases [101]. This results in formation of needle-like mullite grains, which in turn form a porous, yet strong, network [102]. Although mullite has nearly the same thermal expansion as SiC, nonsegmented filters also yield good regeneration, pressure drop and durability results [103]. Commercialization, especially for light-duty and medium-duty commercial vehicles, was planned for Europe in 2008 and USA in 2009 [104,105].

6.2. Ceramic Fiber Filters

Fibrous filters exhibit less uniform and coarser pore structures than the above-mentioned filter materials made from sintered powders, but they can have very high pore volumes and accordingly high permeabilities. For this reason fibrous filters act much more like deep-bed filters and require thicker filters to reach high filtration rates. They also have relatively low strengths and stiffnesses. Fiber felts or materials woven from loose fibers are flexible and exhibit outstanding thermal shock resistance. Widely used and inexpensive ceramic fibers have high glass phase contents, making them problematic for use in DPFs due to the poor high-temperature properties. Pure alumina or SiC fibers with better high-temperature properties are prohibitively expensive as well as more brittle than glass-containing fibers and hence are difficult to weave or knit. Fiber fracture can also release harmful dust.

The company 3M processed its amorphous Al₂O₃–SiO₂–B₂O₃ NEXTEL-312 long fibers by winding on perforated stainless steel tubes that could be electrically heated to regenerate the filter. The company BUCK MASCHINENBAU GmbH developed knitted filter elements with catalytic coatings [106].

FLEETGUARD and 3M developed a filter material made from 3 μm-thick Al₂O₃ fibers connected by SiC deposited from the gas phase to form a rigid filter material. The raw material was processed into a wall flow filter design using paper technology [107].

All of these variants have different properties from those of the honeycomb filters described earlier. Positive aspects are the high thermal shock resistance, low weight and low backpressure. The frequent negative aspects of low filtration rates and limited durability preclude use of these filters in cars, only enabling them to be used in a few special cases.

In 2007 fiber-containing feedstock materials were extruded to produce catalyst substrates [108] and filters [109], the latter using presintered SiC or mullite fibers. This yielded a filter substrate with a porosity of 67%, a pore size of 15 μm and a correspondingly high permeability [110].

6.3. Open-cell Ceramic Foam Filters

The 1980s saw the first investigations on use of open-cell foam ceramics in DPFs [111,112]. The principle of filtration was demonstrated and diverse experiments were carried out to optimize filter geometry to solve the problem of conflicting objectives regarding filter volume, backpressure and filtration efficiency. Refractory ceramics used for molten metal filtration formed the materials basis (clay-bonded alumina, cordierite and similar materials). The results indicated slightly lower starting temperatures for regeneration than is the case in wall flow structures [113].

Mullite (aluminum silicate)-based foam ceramics have been investigated within the scope of EU DPF projects since the mid-90s [114]. The foams were developed for so-called diesel catalytic converters (DCCs) in connection
with the development of special catalysts (e.g., Cs-V compounds) and investigated with the objective of achieving continuous regeneration. Simulation of the pressure drop and thermal behavior of foam DPFs was advanced considerably in these projects and is now as successful as the general simulation of DPFs (especially wall flow filters) [115].

6.4. Metal DPFs

The development of sintered metal filters started with a concept [116] involving application of metal particles (Fe–Cr–Ni alloy) to a coarse mesh substrate and sintering. In this material, as in the wall flow ceramic filter material, the intershoces between the particles sintered together at the contact points form the uniform and easily controllable porosity of the filter. The pore size distributions (15–20 μm) and pore volumes (approx. 45%) correspond to those of typical SiC filters, enabling similar filtration efficiency and pressure drop performance to be achieved.

The material is fabricated as a quasi semi-finished product and then pleated and welded to form radial pocket filters. The filtration and regeneration principles are analogous to those of the variants with ceramic filters; depending on the application active or passive regeneration is used with or without catalytic support. Compared with the channel structures of the extruded ceramic filters, the pleated pocket filters exhibit better ash removal (cleaning) and hence the life or the service interval should be accordingly higher. Other advantages are the low Young’s modulus of the metal, which lessens the problems of thermally induced stress gradients, and the similar thermal expansion coefficient to that of the housing, which facilitates canning. In principle the metal structures can be coated with similar washcoats and catalysts as those of ceramic substrates. A similar approach was pursued for commercial vehicles, although instead of a woven material, more economical expanded metal and punched pocket filters welded in accordion style (i.e. not pleated) were used for the substrate [117]. The basic problem of sintered metal filters is that they are much more expensive than SiC filters due to the high price of metal powders and the complicated welding process needed to manufacture them. Hence these types of filters have not yet become established in OEM passenger cars and have only found limited use in commercial vehicle applications.

Deep-bed filters made solely from metal fiber felt (Fe–Cr–Al) with 85% porosity can be designed with direct electric heating for regeneration and are intended for non-road applications [119].

Open-cell metal foams are also used as DPFs. In principle they have the same cell morphologies as open-cell foam ceramics because they are also produced using polymer foams as starting materials. A system was developed for retrofitting city buses with a cascade of CRT filters with relatively coarse-celled stainless steel foams with a portion being catalytically coated [120]. A special high-temperature Cr–Ni foam can be produced from nickel foams prepared on an industrial scale and used as a wound filter or catalyst support [121,122].

7. MARKET SITUATION

DPFs were first standard used in passenger cars in 2000 in the Peugeot 607 HDI [77], then in the Citroen C5 and the Peugeot 307. A 200 cpsi RSiC wall flow filter with regeneration by a Ce-based fuel-borne catalyst and controlled reinjection was used. For this first filter generation a long-term filtration efficiency of >95% and a service interval of 80,000 km due to the above-mentioned ash problems were specified. A special service company was established to remove the ash from the filters using a special water vapor cleaning technique. The practicality and reliability were verified even in the extreme conditions of Paris taxi operation [123].

Through lowering of the catalyst concentration the ash content in the second generation could be lowered to such a point that the service interval could be extended to 150,000 km. With the introduction of special channel geometry in 2004 to increase the ash capacity the service interval was extended to approx. 250,000 km. In June 2003 shipment of the 500,000th DPF-equipped vehicle was announced; the one million mark was reached at the beginning of 2005 [124].

All remaining major carmakers followed with DPFs in Europe, creating temporary bottlenecks in filter supply. At the end of 2004 worldwide manufacturing capacity for SiC filters was about 1 M per year; by the end of 2006 it reached approx. 6–9 M (RSiC and SiSiC) filters/year thanks to new plants being opened in Poland and Hungary [125–128].

In early 2006 the first aluminum titanate filter as a standard passenger car feature was announced. The production capacity for cordierite filters was increased between 2004 and 2006, mainly to meet the demand for medium and heavy-duty vehicles [129,130]. A market leader forecast its turnover for DPFs to reach the same level as that of its substrate business for catalytic exhaust after-treatment in 2010.

A study of market trends from 2004 forecast an annual turnover of 1.47 B Euros for passenger car DPFs in 2012.
and a volume of 4.2–4.4 M per year in Western Europe [131], with the system price taken to be approx. 340 Euros (in 2006 approx. 450 Euros). In 2006 the price of the SiC ceramic filter made up approximately 30% of the cost of the overall system, i.e. between 130 Euros and 160 Euros depending on size and number of units. The oxidation catalyst accounts for a much higher share of the overall system costs due to the high price of the noble metal.

The actual developments greatly surpassed all predictions. A more recent study carried out at the end of 2009 [132] assumed 8.5 M DPFs per annum, albeit at a system price of approx. 520 Euros, corresponding to turnover of 4.4 B Euros. The year 2016 is expected to see a volume of 13 M units (system unit price: 480 Euros) and turnover running to 6.2 B Euros, approx. 2.45 B Euros of which would be attributable to the ceramic filter substrates.

A 2006 study covering passenger cars and vans, pickups, jeeps etc. [133] was carried out for the North American market (US, Canada and Mexico). The share of diesel engines was expected to increase from 3.6% (2005) to 9.8% in 2012, with about 50% of the vehicles considered being passenger cars. Starting in 2007 equipping of all the considered diesel vehicles (only those manufactured in the US) with DPFs was expected to be mandatory. Turnover was supposed to be US$2.05 B in 2012. The considerably higher unit price in comparison with that in the Western European study was associated with the larger vehicle sizes in the US. However, these predictions haven’t been met yet; in the US the diesel engine market has been stagnating over the last few years. A study from 2011 [134] for ‘light vehicles’ forecast an increase from about 160,000 units (1.3%) in 2009 to about 3% (545,000 units) in 2016, with all units assumed to be fitted with DPFs.

Approximately 3 M new commercial vehicles were registered in Europe in 2007. The lion’s share, 75%, was made up of commercial vehicles up to 3.5 T. The majority of the first standard applications for EURO IV relied on DeNOx systems (SCR) to reach the prescribed particle limits without DPFs. Only in 2012 will coupled (DPF + DeNOx) systems be introduced on a large scale.

In the US and Japan, however, other prescribed limits already made DPFs mandatory in 2007; these filters were made mainly from cordierite and to a lesser extent from SiC.

With the introduction of environmental zones in some cities in Western Europe in about 2006 the introduction of retrofitting systems for passenger cars also became a topic of discussion. Due to the technical difficulties posed by regeneration only open filter systems, usually the two above-mentioned metal filter types, were developed to the stage allowing them to be deemed marketable. Various studies were based on a volume of approx. 0.7–2 M vehicles in Germany alone. Because of (frequently politically motivated) delays and public discussions about the effectiveness of such filters the actual volume was only a fraction of the predicted numbers.

Greater durability and life of larger engines has transformed the topic of retrofitting lorries and nonroad systems. Commercial vehicles can be retrofitted with metal and ceramic DPFs [135,136]. There are currently approx. 3.2 M commercial vehicles in Germany. The share of retrofits depends on the legal situation as well as tax incentives and is hence difficult to estimate.

The oldest market for DPFs is in the retrofitting of nonroad vehicles. This can mainly be traced back to Swiss authorities who passed regulations for tunnel construction sites early on; assessment criteria and test procedures for complex approval testing (so-called VERT tests) were developed based on them. However, there are a very large number of different requirements and a correspondingly large number of systems and filters for this market, making it practically impossible for a comparative overview to be gained and predictions to be made. Test results and recommendations for filter systems are currently being updated and published [137]. In contrast to commercial vehicles, these machines are not surveyed statistically. The starting point for a market estimate is hence formed by the production data supplied by the manufacturers. The number of newly manufactured nonroad vehicles in Europe was estimated to be approx. 120,000 to 150,000 in 2007. Given an average system price of 3500 Euros the annual market volume could reach approx. 0.6 B Euros in 2012/2013.

8. OUTLOOK

The PM–NOx trade-off has driven the development of numerous technical alternatives for restricting emissions of harmful substances from diesel engines, especially through the use of DPFs, DeNOx systems, exhaust gas recirculation and internal engine alterations. However, trends in regulation of limits already point to use of DPFs, at least in combination with DeNOx systems, becoming obligatory in the future for nearly all diesel engines in industrialized nations.

The introduction of particulate number limits in EURO 6/VI can cause a redesign of some filter strategies, although most of the current wall flow filter systems reach the limit of $6 \times 10^{11}$ particles/km.

Ceramic DPFs are a great success from the ceramist’s point of view, comparable to the oxidation catalyst carrier business. The number and variety of ceramic DPFs have been increasing tremendously since their commercial introduction in cars. Wall flow filters made of silicon carbide ceramics are at an advantage in cars because of their robustness. Aluminum titanate, cordierite and mullite are of interest mainly because they can be produced more cheaply. Higher cell density, asymmetric cell design and catalyzed filters have been developed and introduced in the market.
The trend in materials is marked by strong diversification allowing optimization of properties, manufacturing technologies and costs, but also by competition and patent-related issues.

Development of the filter design and system is also focused on further optimization in relation to pressure drop, costs (filter, coating and additional fuel consumption), and the increase in ash tolerance and especially the combined DPF—DeNO\textsubscript{x} systems. Current filter system development activities also concentrate on lifetime guarantees, fast catalyst light-off and on-board diagnostics.

The extent to which the necessity of exhaust after-treatment and the associated costs will cancel out the advantages of the diesel engine in comparison with the Otto-cycle engine is an obvious question. However, expected limits for CO\textsubscript{2} emissions and also possibly the introduction of particle count-based limits for Otto-cycle engines will accelerate the trend toward diesel engines in the future. Thus, the market share for diesel engines should still continue to rise. Over the long term, however, at least for passenger cars, alternative drive technologies (hybrid drives, electric vehicles, fuel cells etc.) will have to be considered.

REFERENCES

[13] UN/ECE Reg 83(Suppl. 7).

[34] Shoji A. An improvement of diesel PM and NO\textsubscript{2} reduction system. Detroit: Diesel Engine-Efficiency and Emissions Research (DEER) Conference; August 2007.


Diesel Particulate Filters


[99] Rembor HJ, Rahn T. A new approach for a diesel particle filter material with liquid phase sintered silicon carbide and an innovative segmented geometry. SAE 2010-01-0532; 2010.


Chapter 9.1

Mechanical Properties of Ceramics

Robert Danzer,* Tanja Lube,* R. Morrell† and P. Supancic‡

* Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Peter-Tunner-Strasse 5, 8700 Leoben, Austria, † Materials Center, Institut fuer Struktur- und Funktionskeramik, (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Peter-Tunner-Strasse 5, 8700 Leoben, Austria, ‡ Materials Division, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

Chapter Outline

1. Important Properties 609
2. Elastic Properties 610
3. Plasticity, Compressive Strength, Yield Strength, and Hardness 613
4. Thermal Strains and Thermal Stresses 614
5. Thermal Diffusion 617
6. Toughness 619
7. Strength 622
8. Strength Degradation with Time 625
9. Final Remarks 628
References 629

1. IMPORTANT PROPERTIES

Ceramic materials behave—in general—as linear elastic, that is, there exists a linear (but tensor) relationship between the components of the stress tensor and the components of the strain tensor (Hooke’s law). For a one-dimensional bar, the relationship between the stress \( \sigma \) and the strain \( \varepsilon \) is \[ \sigma = E \varepsilon. \] (1)

The elastic constant \( E \) is Young’s modulus.

Although inelastic (plastic) deformations occur at very high stresses (i.e. compressive stresses in the order of some \( E/100 \)), plasticity generally does not have to be considered in the stress analysis of components, because the strength limit is exceeded before the yield stress is reached \[2,3\].

The yield strength of ceramic materials is strongly related to the hardness, which is an important property if wear becomes important.

Any successful application of ceramics needs a careful design against occurrence and propagation of cracks, which can be described in the framework of linear elastic fracture mechanics \[4,5\]. Failure is caused by the spontaneous extension of cracks, which can be described by the Griffith/Irwin criterion. It occurs if the stress intensity factor (SIF) \[ K = \sigma Y \sqrt{\pi a} \] reaches or exceeds the fracture toughness \( K_c \): \[ K \geq K_c. \] (3)

In Eqn (2), \( \sigma \) is the stress in the uncracked body, \( Y \) is a geometric factor, which describes the geometry of the crack, the specimen and the stress field and \( a \) is the crack length. The SIF describes the loading of the crack (more precisely the stress field at the crack tip), and the fracture toughness is the resistance of the material against crack extension. Note that the energy necessary to create 1 m\(^2\) of newly cracked area is \[ G_c = K_c^2/E, \] (4)

which is called the specific fracture energy.

In service, temperature changes often occur in components. This causes thermal strains and stresses, which may nucleate and extend cracks in the component \[6,7\]. Therefore, all properties having any influence on these strains and stresses are of highest relevance for the safe operation of ceramic components. If a material undergoes a temperature change \( \Delta T = T_0 - T \) (from the reference temperature \( T_0 \) to the temperature \( T \)), a thermal strain \[ \varepsilon_{th} = \alpha \cdot \Delta T \] (5)

occurs, where \( \alpha \) is the coefficient of thermal expansion (CTE). If thermal strains are mechanically constrained, thermal stresses come into existence, which (at the surface

---

1. For simplicity, a uniaxial stress state is assumed if not mentioned otherwise. For general stress fields, \( \sigma \) is a parameter, which characterizes the amplitude of the field.
Tensile strength is limited by the brittle extension of structural discontinuities, often termed ‘flaws’, which behave like cracks and which exist within the component and on its surface. From the fracture criterion (Eqns (2) and (3)), it follows that the tensile strength of a component is

\[ \sigma_t = \frac{K_c}{Y\sqrt{\pi a_c}}, \]  

where \( a_c \) is the size of the fracture origin (the flaw where fracture starts). Because the size of the fracture origin differs from component to component, the strength of individual components is also different. This makes a statistical description of strength necessary (e.g. using the so-called Weibull theory [8,9]). It also causes the well-known size effect of strength. The probability of finding a large flaw in a large component is higher than in a small component. Therefore, the mean strength of large components is generally lower than that of small components when loaded in the same way.

Of course, there also exist damaging mechanisms, which degrade the strength with time. Very important is the so-called subcritical crack growth (SCCG) [10], where a crack slowly grows, even under a constant load. SCCG always precedes brittle fracture and may cause a significant loss of strength with time. The reason is that crack growth is a thermally activated process that can be enhanced, especially in oxide materials, by the presence of water at the crack tip, which reduces the energy barrier to crack growth [11]. A second mechanism is caused by repeated (cyclic) loading (fatigue crack growth) and is thought to result from the local enhancement of stresses at the crack tip because crack faces do not completely separate or close. Fatigue effects are even possible high under compressive loading [12].

2. ELASTIC PROPERTIES

The linear elastic behavior of materials is described by Hooke’s law (Eqn (1)), which, in its general form, linearly relates the components of the stress tensor \( \sigma_{ij} \) with the components of the strain tensor \( \epsilon_{kl} \) [1]:

\[ \sigma_{ij} = C_{ijkl} \epsilon_{kl}, \]  

where \( C_{ijkl} \) is the tensor of the elastic constants [13]. It holds the Einstein convention, that is, a summation has to be done over the same indices.

To give an example, a stress in the z-direction \( (\sigma_{zz}) \) is—in general—related to nine strain components \( (\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}, \epsilon_{xy} = \epsilon_{yx}, \epsilon_{xz} = \epsilon_{zx}, \epsilon_{yz} = \epsilon_{zy}) \) and to nine elastic constants \( (C_{zzkl}) \). Symmetry can be used to reduce the number of independent constants. In cubic crystals, only three independent constants exist. For single
crystals, the tensor behavior described above has to be accounted for.

Many polycrystalline ceramic materials and glasses behave macroscopically isotropic. Then, the number of independent elastic constants is only two. Some important elastic constants are Young’s modulus \((E)\), which relates the stress with the strain in the stress direction \((\sigma_{ii} = E \cdot \epsilon_{ii})\), and the shear modulus \((G)\), which relates the shear stress with the shear strain \((\sigma_{ij} = G \cdot \epsilon_{ij}; i \neq j)\). Alternatively, Poisson’s ratio \((v)\) which is the ratio of the strain components perpendicular to the stress to the strain components in stress direction: \(v = -\epsilon_{ij}/\epsilon_{ii} = -\epsilon_{kk}/\epsilon_{ii}\)

can be used. Because in homogeneous materials only two constants are independent, there exists a relationship between these constants:

\[
G = \frac{E}{2(1 + v)} \quad (9)
\]

Some ceramic materials have a texture, and therefore, the tensorial elastic behavior has to be appropriately taken into account. For example, elongated grains may be oriented by slip or tape casting \([14,15]\). Other examples are piezoceramics, where the texture changes with the applied electrical field \([16,17]\). It should be noted that the effective symmetry of the polycrystalline ceramic may differ from the symmetry of the underlying single crystals.

The elastic properties of ceramics are directly related to the bonding between the atoms and to the placement of the atoms in the body \([16,17]\). Today, elastic constants can be determined from first principle quantum mechanical calculations using density functional theory. The interaction between neighboring atoms can be described by a potential \(P\) (Figure 1). The minimum of the potential corresponds to the equilibrium distance \((d_0)\) between the atoms (at zero Kelvin). The first derivative of the potential \((P)\) with the distance \((d)\) is the force and the second derivative the curvature, which—in the neighborhood of the equilibrium distance—is almost constant. This value corresponds to the stiffness of a spring and is called bond stiffness \((\frac{\partial^2 P}{\partial d^2})_{d=d_0} = S_0\) \([18]\).

Let us model the elastic behavior of a solid by neighboring atoms connected by springs (Einstein model) \([19]\). Then, stretching or compressing two neighboring atoms would require the force

\[
F = S_0(d - d_0). \quad (10)
\]

In the cross-section of a body, there are \(N = \left(\frac{1}{d_0}\right)^2\) atoms per square meter and \(N\) “springs” per square meter are bridging each cross-section. The force per square meter (i.e. the stress) needed to pull these “springs” apart is \(NF = \sigma = S_0(d - d_0)/d_0^2\). The elongation \((d - d_0)\) divided by the starting distance is the strain: \((d - d_0)/d_0 = \epsilon\). Therefore, we get \(\sigma = (S_0/d_0) \epsilon = E \epsilon\), and thus Young’s modulus is

\[
E = \frac{S_0}{d_0}. \quad (11)
\]

Note that a stronger binding between the atoms results in a deeper potential well, which has a larger bond stiffness (curvature) \(S_0\). Therefore, Young’s modulus increases with bond strength.

Of course, this model only applies if the application of a stress causes stretching or compressing of bonds. This generally happens if the atoms are surrounded by four or more neighbors. If the number of neighbors is lower, as in the case of oxygen in silica glass, elongations can be made by changing the bond angle \([20]\) and Eqn (11) may overestimate the modulus.

It is obvious that this simple model is only valid for very small strains \((d - d_0)/d_0\), that is, at distances where the potential can be modeled by a quadratic function \(P \approx S_0(d - d_0)^2\). At higher distances, the potential becomes asymmetric and nonlinear terms become important. Therefore, the linear elastic regime is restricted to small strains, that is, strains less than a few percent. Because the elastic strain at brittle fracture (which occurs at tensile loading) is typically a few per mille, this limit can hardly be reached in tensile loading.\(^5\) But in compression, strains of a few percent are possible, and nonlinear (plastic) strains can also occur in ceramic materials. This point will be addressed in the next section.

\(^{4}\) Note that a stress in the z-direction causes strains in the z-direction as well as in the x- and y-directions. Typical values of the Poisson ratio are between 0.2 and 0.3.

\(^{5}\) In some ceramic materials, nonelastic (plastic) deformations may also occur at relatively small strains due to the switching of twins or domains (e.g. in piezoelectric ceramic materials, lead zirconate titanate, PZT) or due to phase transformations (e.g. in partially stabilized zirconia PSZ).
In crystals, the bond stiffness as well as the equilibrium distance change with the direction. Therefore, the elastic modulus (the elastic constants) is not isotropic, and it depends on the crystal orientation.

In principle, the approximation of Eqn (11) is made for zero Kelvin, where the ground state near the energy minimum (Figure 1) is occupied. At higher temperatures, higher energy states are also occupied. Here, the symmetry of the potential curve becomes important and causes a slight increase of the distance $d_0$, which should cause a slight decrease of the modulus with increasing temperature. Such behavior is also observed experimentally [21].

For ceramic phases, which have mixed ionic/covalent bonding, the bond stiffness is relatively high, between 15 N/m and 200 N/m [18]. The distance between bonded atoms typically is $2 \times 10^{-10}$ m to $3 \times 10^{-10}$ m [18]. Therefore, Young’s modulus of ceramic phases is expected to be between 50 GPa and 1000 GPa.

In ceramic materials, the microstructure will also have some influence on the modulus. In polycrystalline ceramics, the net properties are an average over the orientation of the individual crystals. For multiphase materials, the average over the microstructural elements lies between the bonds given by the Voigt- (equal strains) and the Reuss- (equal stresses) rules of mixture. Porosity, which for some ceramic materials is an important microstructural constituent, has a significant influence on elastic modulus. The influence of porosity on the modulus of course depends on the shape and the distribution of the pores. For sintered ceramics, a simple estimate gives $E = E_0 (1 - b P)$, where $E_0$ is the Modulus of the pore free ceramic, $P$ is the porosity volume fraction, and $b$ is an empirical constant in the range of $b = 2$ [22].

There are many routes to measure elastic modulus [23–27]. Static bending of a thin bar can be used, but it requires precision in loading and measurement of very small displacements. It is difficult to achieve an accuracy of Young’s modulus to better than 5% [28]. Using strain gauges can achieve greater accuracy, and lateral strain measurement can be used to determine Poisson’s ratio. Ultrasonic wave velocity measurements can be used, and the moduli calculated from the density and the compression and shear wave velocities. This method is the best for coping with anisotropy to determine the tensor stiffness of a uniformly anisotropic material. Perhaps the most reliable method of determining the properties of isotropic materials is based on the natural vibration frequencies (eigen-frequencies) of a bar or a disc of very regular shape, which can be excited by either driving in resonance or impacting the test piece and determining the decaying vibration frequencies. These methods are readily extended to high-temperature measurement. An accuracy of better than 0.1% is achievable provided the shape is prepared to small dimensional tolerances. It should be noted that the dynamic methods of measuring modulus yield slightly higher values than in the quasistatic loading method, because the latter is essentially isothermal, and the others are adiabatic [29].

Some elastic properties are listed in Table 1. Covalently bonded materials have high bond strength and high bond stiffness. Therefore, their modulus is also high (Eqn (11)). Engineering ceramics have a very high Young’s modulus (150 GPa–1100 GPa), a little higher than metals and alloys (40 GPa–500 GPa) and much higher than engineering polymers (0.1 GPa–6 GPa). The stiffest material known is diamond. Our measurements on a slab of polycrystalline diamond give $E = 1080 \pm 10$ GPa and $v = 0.09 \pm 0.01$. Porous ceramic materials (rocks, pottery, etc.) have a modulus between 30 GPa and 200 GPa.

The elastic properties are design parameters for applications, which require a certain flexibility (e.g. springs) or stiffness (e.g. beams). In many cases, a combination of several material parameters can be defined to be a figure of merit, which—to optimize the performance in a particular application—should be as high as possible. Some very instructive examples on this topic can be found in the book “Materials selection in mechanical design” by Ashby [34]. Examples for figures of merit are $E^{1/2}/\rho$ for bent beams with minimum weight, $\sigma^2/E$ for springs with minimized volume

<table>
<thead>
<tr>
<th>Material</th>
<th>Young’s modulus GPa</th>
<th>Poisson ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-based</td>
<td>900–1100</td>
<td>0.08–0.12</td>
</tr>
<tr>
<td>Ceramics</td>
<td>400–600</td>
<td>0.17–0.20</td>
</tr>
<tr>
<td>Silicon carbides</td>
<td>300–400</td>
<td>0.8–0.9</td>
</tr>
<tr>
<td>Silicon nitrides</td>
<td>300–350</td>
<td>0.22–0.275</td>
</tr>
<tr>
<td>Alumina</td>
<td>250–400</td>
<td>0.21–0.27</td>
</tr>
<tr>
<td>Zirconia</td>
<td>150–250</td>
<td>0.2</td>
</tr>
<tr>
<td>3Y-TZP; 208</td>
<td>3Y-TZP; 0.32</td>
<td></td>
</tr>
<tr>
<td>Glasses</td>
<td>40–80</td>
<td>0.17–0.30</td>
</tr>
<tr>
<td>Steels</td>
<td>180–250</td>
<td>0.29–0.31</td>
</tr>
<tr>
<td>Aluminum alloys</td>
<td>60–90</td>
<td>0.33–0.35</td>
</tr>
<tr>
<td>Magnesium alloys</td>
<td>40–50</td>
<td>–</td>
</tr>
<tr>
<td>Nylons</td>
<td>2–5</td>
<td>–</td>
</tr>
<tr>
<td>Epoxy</td>
<td>1–3</td>
<td>0.33</td>
</tr>
<tr>
<td>High-density</td>
<td>0.4–0.9</td>
<td>–</td>
</tr>
<tr>
<td>Polyethylene</td>
<td>0.4–0.9</td>
<td>–</td>
</tr>
<tr>
<td>Low-density</td>
<td>0.1–0.3</td>
<td>–</td>
</tr>
<tr>
<td>Polyethylene</td>
<td>0.1–0.3</td>
<td>–</td>
</tr>
</tbody>
</table>
Mechanical Properties of Ceramics

In general, it holds that the compressive strength \( s_c \) is the mass density. Note that depending on the application, very high as well as very low values of Young’s modulus can be desirable.

### 3. PLASTICITY, COMPRESSION STRENGTH, YIELD STRENGTH, AND HARDNESS

Plastic deformation caused by the movement of dislocations in pure metals is very low, and strengthening mechanisms have to be tailored to give metals a usable strength [35]. Ceramic materials have a hybrid (mixed ionic/covalent) bond, so that binding orbitals of electrons are localized around the corresponding ion cores, that is, the electrons are immobile. This is the main reason why the movement and generation of dislocations in ceramics needs much more energy than in metals, where the valence electrons are mobile. Plastic deformation of ceramics is possible but only at very high stresses (or at relatively high temperatures), and furthermore often a result of there being a glassy second phase in the microstructure.

In tensile loading of ceramics, brittle fracture in general occurs before severe plastic deformation takes place and the (plastic) fracture strain is (almost) zero. But in compressive loading, brittle fracture is suppressed, and some plastic deformation of ceramic materials becomes possible [3]. The compressive strength is connected to the formation of glide bands and the pile-up of dislocations at grain boundaries, which may locally cause large tensile or shear stresses, and the generation of microcracks, which initiates the failure. Although other mechanisms may also be important, the yield strength \( \sigma_y \) is roughly equal to the compressive strength \( \sigma_c \) of most ceramics. Therefore, in general, it holds that \( \sigma_y \approx \sigma_c \) [36].

One example where plasticity is commonly present is in the testing of hardness. Hardness is a measure of the resistance of the material against plastic deformation. There exist numerous variants for hardness testing, but for ceramic materials, the Vickers or Knoop hardness test is generally used. Here, a diamond pyramid is pressed with the force \( F \) onto the surface of the material, which produces a square (kite-shaped in the case of Knoop indenters) indentation in the surface (experimental details will be discussed later). The so-called true hardness (or pressure hardness) is defined by

\[
H = \frac{F}{A}
\]

where \( A \) is the projected area of the impression. Around and under the indentation, the material is plastically deformed. There exists a relationship between the hardness and the yield strength [36]:

\[
H \approx 3 \sigma_y.
\]

In principle, all strengthening mechanisms known from metals also apply for ceramics. Therefore, they can be used to increase the hardness of ceramics.

There are some other reasons for inelastic deformations of ceramics than the movement of dislocations. A prominent example is the family of piezoceramic perovskites, where a cubic high-temperature phase transforms to a tetragonal and/or a rhombohedral low-temperature phase (which is ferroelastic). Let us restrict the discussion to the tetragonal phase. Compared with the cubic phase, the tetragonal crystals are elongated in the direction of one of the three cubic orientations. During cooling from the sintering temperature, tetragonal crystals of all three possible orientations are built in the cubic mother crystals. If a stress is applied, inelastic (plastic) strains can arise by switching the orientation of tetragonal subcrystals, corresponding to a movement of domain walls. It is interesting to note that because the tetragonal crystals have a dipole moment, a switching of the crystal orientation can also be caused by the application of an electric field (i.e., ferroelectric effect). This causes the high piezoelectric effect (length change due to the application of an electric field) of this class of materials [37].

Hardness tests for ceramics have been standardized, based on experimental experience. There are two principal issues of concern: one is an apparent indentation force dependence of hardness, especially at low force levels, and the other is the damage (cracking, spallation) associated with indentation, which makes the measurement of indentation size unreliable or impossible (Figure 2). For Vickers hardness tests for ceramics have been standardized, based on experimental experience. There are two principal issues of concern: one is an apparent indentation force dependence of hardness, especially at low force levels, and the other is the damage (cracking, spallation) associated with indentation, which makes the measurement of indentation size unreliable or impossible (Figure 2).

**FIGURE 2** HV5 indentation in alumina. Typical problems arising with hardness testing of ceramics become apparent: radial cracks emanating from the indentation corners, secondary cracks, lateral cracking (indicated by the halo around the indent), unclear corners. For color version of this figure, the reader is referred to the online version of this book.
indentation, 1 kgf indentation force is preferred [38–41] to minimize damage around indentations while maximizing the size of indentation for measurement accuracy. The hardness value is conventionally computed as force divided by the actual area of the indentation surface, which is related to the diagonal size \( d \) through \( HV = 1.8455 \ F/d^2 \) with no units (kgf/mm\(^2\) implied), or in gigapascals. For Knoop indentation, the test force is recommended to be 1 kgf or 2 kgf [38,40,42], and the hardness value is computed as \( HK = 14.22 \ F/d^2 \) with, in this case, \( d \) being the long diagonal length, and with similar units. Less frequently, hardness from other types of tests may also be reported, such as Rockwell tests [40], or depth sensing hardness (e.g. Martens hardness).

Hardness values of ceramic materials in comparison to those of other materials are listed in Table 2. Compared to polymers and even to metals, ceramic materials are very hard. The hardest material ever tested is diamond.

Hardness values of ceramic materials in comparison to those of other materials are listed in Table 2. Compared to polymers and even to metals, ceramic materials are very hard. The hardest material ever tested is diamond.

<table>
<thead>
<tr>
<th>Material</th>
<th>Yield Strength (MPa)</th>
<th>True Hardness (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-based materials</td>
<td>35,000</td>
<td>20,000–70,000</td>
</tr>
<tr>
<td>Cemented carbides</td>
<td>1000–3000</td>
<td>3000–20,000</td>
</tr>
<tr>
<td>Silicon carbides</td>
<td>5000–9000</td>
<td>15,000–20,000</td>
</tr>
<tr>
<td>Silicon nitrides</td>
<td>4000–7000</td>
<td>12,000–21,000</td>
</tr>
<tr>
<td>Alumina</td>
<td>3000–5000</td>
<td>9000–15,000</td>
</tr>
<tr>
<td>Zirconia</td>
<td>1000–5000</td>
<td>3000–15,000</td>
</tr>
<tr>
<td>Glasses</td>
<td>600–3000</td>
<td>1800–5000</td>
</tr>
<tr>
<td>Mild Steels</td>
<td>300–600</td>
<td>900–1800</td>
</tr>
<tr>
<td>High-speed steels</td>
<td>500–1000</td>
<td>1500–3000</td>
</tr>
<tr>
<td>7075 Aluminum alloys</td>
<td>300–700</td>
<td>900–2100</td>
</tr>
<tr>
<td>Nylons</td>
<td>60–120</td>
<td>180–360</td>
</tr>
<tr>
<td>Epoxies</td>
<td>40–90</td>
<td>120–270</td>
</tr>
<tr>
<td>High-density polyethylene</td>
<td>15–50</td>
<td>45–150</td>
</tr>
<tr>
<td>Low-density polyethylene</td>
<td>8–20</td>
<td>24–60</td>
</tr>
</tbody>
</table>

Hardness is a key property in all applications related to wear. The wear mechanisms of ceramic materials can be classified—with increasing wear rate—in wear caused by (i) atomic adhesion, (ii) chemical reactions, (iii) grooving and (iv) material fragmentation [44]. A high hardness especially reduces the grooving mechanism, sometimes tested via a ‘scratch hardness’ test. Because ceramic materials are—in general—also chemically very stable (which restricts adhesion and chemical reactions), they are an ideal solution for wear protection either as a bulk product or as a coating on a softer substrate. Prominent ceramic materials used in wear protection are silicon nitride ceramics, alumina ceramics, silicon carbides, cemented carbides, and diamond-based materials. Typical applications, where wear reduction is of the highest interest are knives for industrial cutting, indexable inserts for metal cutting, ceramic tiles for wear protection of paper-making machines, or yarn guides in looms.

Ceramic materials have also been used for many years for knife edges and pivots, for example, in mechanical watches, where the high yield strength causes an excellent load bearing capacity. The corresponding figure of merit is \( \sigma_y^2/E^2 \). Ceramic materials are used for face seals (\( \sigma_y/E \)) and bearings, where the high yield strength of the material is used. Such seals, for example, based on silicon carbides, can also be used at very high temperatures (up to 1400 °C).

### 4. THERMAL STRAINS AND THERMAL STRESSES

#### 4.1. Thermal Strains

Generally, a change of temperature induces a dimensional change (i.e. expansion or shrinkage) of solids [6] generally described by a displacement or strain field. In crystals, the thermal strains \( \varepsilon_{th} \) are described by a (symmetric) second-order tensor,

\[
\varepsilon_{th,ij} = \alpha_{ij} \Delta T, \tag{14}
\]

and therefore, the shape of an initially spherically shaped (unconstrained) solid will change to an ellipsoid due to a temperature change. Note that a change of the coordinate system into the principal axes leads to a set of three fundamental heat expansion coefficients \( \alpha_{ij} \) with respect to the principal axes in the solid. For isotropic materials, this will result in a spherical tensor, where the essential parameter of the heat expansion is given by one single heat expansion coefficient \( \alpha \)

\[
\alpha_{ij} = \alpha \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{15}
\]

The trace of \( \alpha_{ij} \) gives the volume (bulk) heat expansion coefficient [13]. Care has to be taken on the definition of the coefficient of linear thermal expansion (CTE), because two different ones are used. The technical (or mean) CTE is defined as the relative change in length \( \Delta l = l - l_0/l_0 \) (\( l \): length at temperature \( T \); \( l_0 \): initial length at the reference temperature \( T_0 \)) resulting from a temperature change:

\[
\alpha_{tech}(T, T_0) = \frac{1}{l_0} \frac{\Delta l}{(T - T_0)} \tag{16}
\]
This CTE is suitable for engineering purposes, because it is the mean CTE in a given temperature interval (i.e. secant coefficient). Additionally, this quantity is the natural result of thermal expansion measurements, which are performed on bar-shaped specimens by instruments such as dilatometers, recording the length change with respect to the temperature change.

The linear (or differential) CTE, sometimes termed expansivity, is defined by the relative length change due to an infinitesimal temperature increase [45,46]:

\[ \alpha_{\text{diff}}(T) = \lim_{T_0 \to T} \frac{\Delta l}{l_0} \frac{l(T) - l_0}{T - T_0} = \frac{1}{l} \frac{dl}{dT}. \]  

(17)

Using the differential CTE the thermal strain within linear thermoelasticity is

\[ \varepsilon_{\text{th}}(T, T_0) = \int_{T_0}^{T} \alpha(T') \cdot dT'. \]  

(18)

Figure 3 shows a typical shape of the thermal strain with the temperature and the difference between the differential and the technical CTEs.

In an analysis of thermal strains and stresses, the accurate description of the temperature dependence of the CTE can be of high relevance and a simple approach, which ignores this dependence, can be misleading. In that respect, a useful approximation is presented in the following.

Let us restrict our analysis to homogeneous and isotropic solids. Remember that the length change \( \Delta l(T, T_0) = l(T) - l_0 \) is caused by the temperature change \( \Delta T = T - T_0 \). In most cases, the length change can nicely be fitted by a simple second-order polynomial:

\[ \frac{\Delta l(T, T_0)}{l_0} = a_{\text{th}} \cdot \Delta T(T, T_0) + b_{\text{th}} \cdot (\Delta T(T, T_0))^2. \]  

(19)

Then, the technical CTE simply is

\[ \alpha_{\text{tech}}(T, T_0) = \varepsilon_{\text{th}} / \Delta T(T, T_0) = a_{\text{th}} + b_{\text{th}} \cdot (\Delta T(T, T_0))^2. \]  

(20)

and for the differential CTE \( (T \to T_0) \) we get

\[ \alpha_{\text{diff}}(T) = d\varepsilon_{\text{th}} / dT = a_{\text{th}} + 2b_{\text{th}} \cdot \Delta T(T, T_0). \]  

(21)

Of course, the coefficients for the power series and the technical CTE depend on the reference temperature, but the differential CTE does not. In this approximation, the CTE linearly increases with the temperature difference, and the slope of the differential CTE is twice that of the slope of the technical CTE.

Similarly to Young’s modulus, the CTE is strongly related to the shape of the binding potential between the neighboring atoms. At temperatures above zero Kelvin, higher energy states above the ground state are occupied. Because the potential is asymmetric (nonparabolic) and wider at higher energy states (compared with the ground state), the mean position is shifted to a larger distance between the atoms (see the dashed line in Figure 1). For atoms with four or more neighbors, most of this increase in distance is translated into a length increase of the solid body. If atoms in a crystal are lower coordinated, the unit cell is not perfectly filled and contains a large fraction of empty space. Then, bond angles can be changed, atoms can be shifted into the empty spaces, and the CTE can be smaller than expected from the binding potential asymmetry.

Of course, the shape of the potential and the arrangement of the atoms in the crystal depend on the direction in the crystal. Therefore, the CTE of noncubic crystals is anisotropic. But in polycrystalline ceramic materials without overall texture and in glasses, an average over all direction occurs, and the CTE appears macroscopically isotropic.

The influence of microstructure on the CTE is small, and simple rules of mixture can, in general, be used to account for the influence of several microstructural constituents. If the grains in polycrystalline ceramics have some texture, this has to be considered to describe the anisotropy of heat expansion correctly, and dependent on the orientation, the CTE has three different values.

The measurement techniques for thermal elongations and CTEs using dilatometry are well established since a long time [47–50]. A good description can be found in

---

6. In some cases, a quadratic fit is not sufficient, and a higher order (e.g. third order) or other nonlinear ansatz function is necessary.
Although the CTEs of stiff materials may be low, they still generate significant thermal stresses when a component is subjected to temperature gradients and are thus highly relevant for thermal stress analysis. Therefore, the measurements generally have to be performed with very high precision. To give an example, for some designs, a precision in CTE of ±0.1 ppm/K is necessary. If the temperature range of interest is ΔT = 300 K and the specimen length is l₀ = 25 mm (if specimens are cut out of components, they can even be shorter), this corresponds to a length change of the specimen during the experiment as small as 0.75 μm. This is of the order of magnitude of the roughness of machined surfaces. Therefore, a key problem of high precision dilatometer measurements is the proper preparation of the specimens, which has to be done with highest precision. The parallel ends of the specimens should be very smooth (polished) and must be exactly perpendicular to the specimen’s axis.

Some CTE values of engineering materials are listed in Table 3. It can be recognized that ceramic materials and glasses have small CTEs, smaller than those of most metals. For some zirconia ceramics and alloys and even much smaller than those of polymers, the CTE is around two orders of magnitude higher. The parallel ends of the specimens have to be done with highest precision. This is of the order of magnitude of the roughness of machined surfaces. Therefore, a key problem of high precision dilatometer measurements is the proper preparation of the specimens, which has to be done with highest precision. The parallel ends of the specimens should be very smooth (polished) and must be exactly perpendicular to the specimen’s axis.

**Table 3** CTE and the Thermal Stress Parameter of Some Materials at Room Temperature [2,30,32]

<table>
<thead>
<tr>
<th>Material</th>
<th>Differential CTE 10⁻⁶/K</th>
<th>TSF MPa/K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-based</td>
<td>1–2</td>
<td>1–1.5</td>
</tr>
<tr>
<td>materials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cemented carbides</td>
<td>5–6</td>
<td>3.5–4.5</td>
</tr>
<tr>
<td>Silicon carbides</td>
<td>4–5</td>
<td>1.5–3</td>
</tr>
<tr>
<td>Silicon nitrides</td>
<td>2.5–4</td>
<td>0.8–1.2</td>
</tr>
<tr>
<td>Alumina</td>
<td>7–9</td>
<td>2–3</td>
</tr>
<tr>
<td>Zirconia</td>
<td>6–12</td>
<td>1.5–3</td>
</tr>
<tr>
<td>Glasses</td>
<td>3–5</td>
<td>0.2–0.6</td>
</tr>
<tr>
<td>Steels</td>
<td>12–18</td>
<td>3–4</td>
</tr>
<tr>
<td>Aluminum alloys</td>
<td>20–30</td>
<td>1.8–3</td>
</tr>
<tr>
<td>Copper alloys</td>
<td>20–30</td>
<td>2–3</td>
</tr>
<tr>
<td>Nylons</td>
<td>80–150</td>
<td>0.3–0.5</td>
</tr>
<tr>
<td>Epoxy</td>
<td>60–100</td>
<td>0.1–0.3</td>
</tr>
<tr>
<td>High-density polyethylene</td>
<td>130–200</td>
<td>0.07–0.15</td>
</tr>
<tr>
<td>Low-density polyethylene</td>
<td>150–250</td>
<td>0.03–0.07</td>
</tr>
</tbody>
</table>

Some steels, the CTE is (almost) equal, which makes a joining with a reasonably small thermal strain mismatch possible.

### 4.2. Thermal Stresses

Thermal stresses in a body come into existence, if thermal strains are constrained. They are proportional to the thermal strains, to the appropriate elastic constant E’, and to a factor S_th, which is between zero and one and which depends on the constraint [7]. If—to give an example—the ends of a one-dimensional bar are free to move, a temperature increase causes an elongation and a thermal strain but no thermal stress (there is no constraint) in the bar. If the ends are clamped, a temperature increase causes no elongation but a thermal strain and a thermal stress (there is full constraint). In general, the thermal stress is

\[
\sigma_{th} = \epsilon_{th}(T, T_0) \cdot E' \cdot S_{th} = \alpha_{tech}(T, T_0) \cdot E' \cdot \Delta T \cdot S_{th}
\]

For a uniaxial stress state, as it occurs in the one-dimensional bar discussed above, it holds that E’ = E. For a plane stress state, as occurs during quenching of components on their surfaces, it holds that E’ = E/(1 − v). For the bar with free ends, there is no constraint (S_th = 0), and for the bar with clamped ends, the constraint is complete (S_th = 1). The S_{th,0} is the thermal stress, if the strains are completely constraint (for S_th = 1). From Eqn (22), it follows that thermal stresses scale with the material parameter \( \alpha_{tech} E/(1 − v) = f_{\sigma,th} \), which we call thermal stress factor (TSF).

It is interesting to note that for dense (non-porous) metallic and ceramic materials, in which atoms are highly coordinated (K ≥ 4), the values of the TSF are in a narrow range [18]:

\[
f_{\sigma,th} = \alpha_{tech} \frac{E}{1 − v} \approx 2 − 3 \text{ MPa K}^{-1}. \tag{23}
\]

This rule of thumb can also be related to the shape of the binding potential. There is a tendency that the CTEs of strongly bonded materials are low and vice versa. As stated in the paragraph on the elastic constants, a stronger binding needs a deeper potential well, which causes a larger Young’s modulus. A deep well is also narrower and less asymmetric. Therefore, the CTE is smaller, and the product of CTE and modulus is almost constant. Some data of the TSF are listed in Table 3. In solids, which contain atoms with a small coordination number, strains between atoms can be accommodated by changing of bond angles and the TFS may become smaller than defined in Eqn (23). Important examples are diamond materials and silicon nitride, which also have an excellent thermoshock resistance, and glasses. A similar effect can
result from microcracks, which are developed in some materials during cooling down from the consolidation temperature. This may happen, if the individual grains have largely anisotropic expansion/contraction characteristics or in multiphase materials. In this case, opening and closing of crack borders accommodate the strains and reduce the constraint. Although their tensile strengths are generally low, these materials may have an extremely high thermal shock resistance. Prominent examples are aluminum titanate and some plasma sprayed ceramics [52, 53].

Let us discuss the consequences of the rule of thumb (Eqn (23)) in more detail. It tells us that for most ceramic and metallic materials, very small temperature changes can cause very high mechanical stresses. Temperature differences of only 100 °C can cause a stress of about 300 MPa, which is about the yield strength of simple structural steel. Temperature differences of about 500 °C can cause stresses even beyond the strength of high speed steels. Creating inhomogeneous temperature distributions in a constrained body is probably the easiest way to generate very large tensile stresses. Brittle failure of ceramics occurs if the stress exceeds the tensile strength \( \sigma_f \):

\[
\sigma_f \geq \sigma_t. \tag{24}
\]

Using Eqs (22) and (24), we can determine the critical temperature, \( \Delta T_c \) where failing of ceramic components (under plain stress loading) will occur under full constraint [54]:

\[
R = \Delta T_c = \frac{\sigma_f}{f_{\sigma,th}} = \frac{\sigma_f(1-v)}{\alpha E}. \tag{25}
\]

Many ceramics have a low tensile strength. For example, for many electro ceramics, the strength is about 100 MPa. Then, the critical temperature can be as low as 30 °C. It should be noted that, even if the difference in the TSF between different materials is relatively small, it can be of the highest relevance for the mechanical reliability of the components.

Temperature changes can occur during the production of the ceramic component or in service. Typical reasons for temperature changes during production are heat treatments, soldering of components, or the application of layers on ceramic components. Sintering temperatures are— in general— higher or even much higher than 1000 °C. Therefore, during cooling down from sintering temperature, thermal expansion mismatch may cause significant thermal stresses, which may even destroy the component.

In operation, for example, in heat engines, temperature changes of several hundred degrees Celsius are also possible. Then, very high thermal stresses can occur, which have to be considered in the design of the component.

In multimaterial components, thermal stresses may also occur if the materials have a different CTE (which is the common case), such as in rigid joints. This CTE mismatch causes—to give a common example—the bending of a bimetal strip with temperature. In ceramic multilayers, different CTEs can be used to produce layers with compressive stresses, which can act as crack stoppers [55, 56]. In modern multilayer ceramic systems such as those used in electronic systems—the CTE mismatch between the ceramic and the metallic electrodes can cause severe thermal stress. In principle, the stresses scale as follows:

\[
\sigma_{th} = \Delta \alpha \cdot E^* \Delta T. \tag{26}
\]

where \( \Delta \alpha \) is the difference between the CTEs of the materials and \( E^* \) is a “mean” elastic modulus of the two materials. The CTE mismatch between silicon nitride ceramics and steel is around 10 ppm/K, and a temperature change of 100 °C can therefore cause stresses of \( E^*/1000 \), which can be a few hundred MPa. If in a multimaterial component copper or aluminum alloys are in combination with ceramic materials (as in the case of electroceramic components), the CTE mismatch is between 20 and 30 ppm/K. This can be the reason for crack formation in the ceramic after sintering or after some thermal cycling. It may even occur as a result of the temperature differences between night and day [57].

Thermal strains are also of great relevance, if components are joined together. Changing the temperature from that at which the joint is stress-free will cause significant mismatch strains (Eqn (26)) to develop, leading to fracture if these are excessive. An example is shrink fitting, in which initial dimensional differences have to be very carefully calculated to avoid overstressing the ceramic. The guiding rule is always to place the ceramic into compression.

5. THERMAL DIFFUSION

Thermal strains and stresses in monolithic components are triggered by temperature differences developed in the components, which are determined by the rates of heat fluxes and heat sinks or sources (i.e., energy transformations of heat). The transfer of heat within a material is mainly determined by thermal diffusion and the corresponding change in the temperature field \( T(\vec{r}, t) \). It can be described with the well-known equation of heat conduction, based on Fourier’s law [6]. In a simplified approach (i.e., neglecting phase transformations, anisotropy), this equation can be expressed as

\[
\frac{dT(\vec{r}, t)}{dt} = a \nabla^2 T(\vec{r}, t). \tag{27}
\]

---

7. For the definition of the tensile strength, see the following sections.
where \( \mathbf{r} \) is the position vector, \( t \) is the time, \( \nabla^2 \) is the Laplace operator, and \( \alpha \) is the thermal diffusivity, which is taken here as constant.

The heat transport in ceramics is dominated by the movement of phonons [58]. The thermal diffusivity is proportional to the heat capacity, the phonon velocity, and the mean free path of the phonons. In the classical limit, the heat capacity is three times the Boltzmann constant \( (c_v = 3k) \), but quantum mechanical arguments show that this is only true at temperatures well above the Debye temperature (which is several hundred degrees Celsius for most ceramic materials). At low temperatures, the quantum states are not completely filled (“they freeze out”) and they cannot consume energy. Therefore, at lower temperatures, the heat capacity is much lower than described in the classical limit and even reaches zero at zero Kelvin. For details, see Ref. [45]. The phonon velocity is equal to the speed of sound, which is \( v_s = \sqrt{E/\rho} \). This property only weakly depends on temperature.

The mean free path of phonons in a perfect crystal is restricted by the phonon–phonon scattering, which increases with increasing nonparabolicity (asymmetry) of the binding potential. Therefore, the phonon–phonon scattering is low if the binding partners have the same or a similar mass, but it is high if they have a very different mass. In diamond, where C-atoms are bonded to C-atoms, the nonparabolicity is small, and the mean free path is large. Diamond is the material with the highest thermal diffusivity known. In zirconia, where the mass of the binding partners is very different, the nonparabolicity is large and the mean free path is small (about one lattice vector). The phonon–phonon scattering largely increases with temperature, and the mean free path decreases. In summary, the thermal diffusivity has a maximum (which is at temperatures much lower than zero degree Celsius) and then decreases (for several orders of magnitude in some cases). Therefore, a proper analysis of temperature fields and of their evolution with time needs the knowledge of the thermal diffusivity in the temperature range of interest.

The microstructure has a strong influence on the thermal diffusivity. Alloying increases the phonon–phonon scattering and drastically reduces the thermal diffusivity. The arrangement of the constituents is of great importance in multiphase materials. If heat conducting channels exist, they dominate the diffusivity. If the heat conducting phases are enclosed by insulating layers, the insulating phase dominates the conduction rate. This may happen in liquid-phase sintered ceramics, in which glassy films exist at the grain boundaries.

Thermal diffusivity is most readily measured by the so-called ‘laser flash’ method [59–61]. A parallel-faced disc (thickness, \( t \)) is subjected to a laser flash on one face, and the time taken for the temperature of the other face to increase to half its maximum \((\tau_{1/2})\) is determined. Thermal diffusivity \( \alpha \) (units m\(^2\)/s) is then computed from the expression: \( \alpha = 0.138 \frac{r^2}{\tau_{1/2}} \). In practice, depending on the equipment, a number of corrections may have to be made for factors such as finite pulse time effects and radiation losses. Compared with more traditional ways of measuring thermal conductivity under steady-state temperature gradients, thermal diffusivity measurements employ much smaller temperature differentials and thus tend to be more representative of true material behavior. The method is also readily extended to very high temperatures.

Thermal conductivity \( \lambda \) can be computed from thermal diffusivity by multiplying the latter by density \( \rho \) and specific heat \( C_p \): \( \lambda = \alpha \rho C_p \). Example data are shown in Table 4.

High values of thermal diffusivity or thermal conductivity implicate that localized hot or cold spots are quickly removed and the temperature is homogenized faster than in materials with low values of these properties. Therefore, thermally induced internal strains are reduced and the propensity to cracking also. High thermal conductivity

<table>
<thead>
<tr>
<th>Material</th>
<th>Thermal diffusivity ( 10^{-6} \text{ m}^2/\text{s} )</th>
<th>Thermal conductivity ( \text{W/m K} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-based materials</td>
<td>400–1300</td>
<td>600–2000</td>
</tr>
<tr>
<td>Cemented carbides</td>
<td>20–30</td>
<td>60–80</td>
</tr>
<tr>
<td>Silicon carbides</td>
<td>30–100</td>
<td>100–200</td>
</tr>
<tr>
<td>Silicon nitrides</td>
<td>4–12</td>
<td>7–40</td>
</tr>
<tr>
<td>Alumina</td>
<td>6–11</td>
<td>15–30</td>
</tr>
<tr>
<td>Zirconia</td>
<td>0.7–1.2</td>
<td>1–2</td>
</tr>
<tr>
<td>Aluminum nitride</td>
<td>30–80</td>
<td>80–200</td>
</tr>
<tr>
<td>Glasses</td>
<td>0.6–0.8</td>
<td>0.7–1.3</td>
</tr>
<tr>
<td>Steels</td>
<td>3–15</td>
<td>11–55</td>
</tr>
<tr>
<td>Aluminum alloys</td>
<td>30–90</td>
<td>75–230</td>
</tr>
<tr>
<td>Copper alloys</td>
<td>50–120</td>
<td>160–390</td>
</tr>
<tr>
<td>Nylons</td>
<td>0.09–0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Epoxies</td>
<td>0.85–2.5</td>
<td>0.2–0.5</td>
</tr>
<tr>
<td>Polyethylene</td>
<td>0.15–0.3</td>
<td>0.42</td>
</tr>
</tbody>
</table>
is therefore of advantage in most thermal shock situations.\textsuperscript{8}

\section*{6. TOUGHNESS}

Toughness of ceramic materials can be defined in the framework of linear elastic fracture mechanics\textsuperscript{[4,5]}, where it is assumed that the material can be approximated by a homogeneous and isotropic elastic continuum, which contains a crack. Via the crack borders, no loads can be transferred. Then, the local stress field $\sigma_{ij}$ around the crack tip depends on the SIF $K = \sigma Y \sqrt{\pi a}$ (Eqn (2)):

$$\sigma_{ij}(r, \theta) = \frac{K}{\sqrt{2\pi r}} f_{ij}(\theta),$$

where $r$ is the distance of the volume element from the crack tip, where the stress is analyzed, and $\theta$ is the angle between the crack plane and the line from the tip to the volume element (Figure 4). The tensor $f_{ij}(\theta)$ is dimensionless and only depends on the angle, $\theta$. Note that the stress field scales with $K = \sigma Y \sqrt{\pi a}$, that is, a large stress applied in a body containing a small crack causes the same crack tip field as a small stress on a large crack. Different crack opening modes, that is, opening, sliding and tearing crack tip field as a small stress on a large crack. Different applied in a body containing a small crack causes the same result to the SIF: $G = K^2/\pi$. It relates to the SIF: $G = K^2/\pi$. The critical strain energy release rate is the energy needed to produce 1 m$^2$ of new fractured area (specific fracture energy). It holds (Eqn (4)) that $G_c = K_c^2/\pi$. It can be expected that this energy is the surface energy, which has to be invested to generate the new surface ($G_{c,0} = 2\gamma$; the factor two arises because two new fracture surfaces are created in each new fracture; typical surface energies are in the order of magnitude of 1 J/m$^2$), but it can easily be shown that the specific fracture energy of technical materials is much higher. As will be discussed in the following, there exist also other contributions to the specific fracture energy, which may become important.

Two main processes are known to contribute to the specific fracture energy: process zone mechanisms and crack bridging mechanisms. Process zones are built ahead of the crack tip, if the material shows—at least a little—nonlinear behavior. Equating the yield strength ($\sigma_y$) with the stress given in Eqn (28) (for simplicity, we neglect the tensor behavior of the stress field and set $f_{ij}(\theta) = 1$) and if the SIF reaches the fracture toughness ($K = K_c$), we get an estimate for the diameter of the inelastically deformed zone $d_{PI}$ (process zone):

$$d_{PI} \approx \frac{1}{\pi} \left( \frac{K_c}{\sigma_y} \right)^2.$$  

$K_c$ is the fracture toughness without the action of the mechanisms increasing the fracture energy (i.e. it is the fracture toughness arising from the surface energy). This is shown in Figure 5a. Examples for reasons of inelastic strains ($\epsilon_{in}$) can be plasticity (in this case, the process zone is called plastic zone), the switching of domains (examples are piezoceramics) or phase transformations (examples are partially stabilized zirconia, PSZ and zirconia toughened alumina, ZTA).

If a crack propagates, the process zone moves with the crack tip through the body, leaving behind a wake of nonelastically deformed material at the crack borders (see Figure 5b). The energy dissipated per deformed volume (assuming “ideal plastic” behavior) is $\approx \sigma_y \epsilon_{in}$. If the crack extension ($d\alpha$) is large compared with the zone diameter and if we assume a straight through crack in a plate of thickness $t$, the volume of the deformed material can be estimated to be $d_{PI} d\alpha = d_{PI} \partial A$, where $\partial A$ is the newly cracked area. Multiplying the dissipated energy density with
with the deformed volume we get for the dissipated energy per newly cracked area,

$$\partial G_{PZ} = d_{PZ} \sigma_{PZ} \varepsilon_{in}. \quad \text{(30)}$$

If we consider, for example, the relationships of typical steel (where inelastic deformation is plastic deformation), typical values are $d_{PZ} = 10^{-3} \text{m}$, $\sigma_{PZ} = 500 \text{ MPa}$ and $\varepsilon_{in} = 5 \times 10^{-2}$. The specific fracture energy is $2.5 \times 10^3 \text{J/m}^2$. If we set, for the elastic modulus $E = 200 \text{ GPa}$, and if the other contributions to the fracture energy can be neglected, the fracture toughness is $K_c = \sqrt{EG_c} \approx 70 \text{MPa}\sqrt{\text{m}}$. This is a typical value of mild steels [34].

In the case of ceramics, the yield strength is very high (see Section 3), and the diameter of the plastic zone is very small. To give an example, for $\sigma_y \approx 3000 \text{ MPa}$ and $K_c \approx 3 \text{ MPa}\sqrt{\text{m}}$, the diameter of the plastic zone diameter is only one-third of a micron, much too small to be recognized by a light microscope. The contribution of plasticity to the specific fracture energy is then negligible.

In the case of transformation toughened ceramics, inelastic strains are produced by the phase transformation of metastable particles. To improve strength, the volume of the particles after transformation has to be larger than the volume before transformation. A prominent example is the transformation of tetragonal into monoclinic zirconia, which is associated with a volume increase of the unit cell of about 5%. The transformation is triggered by the tensile stresses ahead of the crack tip, that is, it occurs only in the so-called “transformation zone.” The transformation (yield) stress ($\sigma_T$) can be adjusted by the particle size and the process zone diameter can reach several tens to more than hundred micrometers. Then, the transformation zone can be even seen with the naked eye [63–65].

Transformation-toughened ceramics have a relatively high toughness. For $d_{PZ} \approx 5 \times 10^{-5} \text{m}$, $\sigma_{PZ} = \sigma_T \approx 1000 \text{ MPa}$, and $\varepsilon_{in} = 1 \times 10^{-2}$ (if, e.g., $20 \text{ vol.}\%$ of particles transform and the volume increase of each particle is $\approx 5\%$, the resulting inelastic strain of the ceramic is $\approx 1\%$), the specific fracture energy is $500 \text{ J/m}^2$. With $E = 200 \text{ GPa}$, we get for the fracture toughness $K_c = \sqrt{EG_c} \approx 10 \text{ MPa}\sqrt{\text{m}}$.

A third kind of toughening mechanism related to process zones is the so-called microcrack toughening, which may occur in ceramics being composed of two or more phases, for example, in ZTA [66], SiC–TiB$_2$ [67], and B$_4$C–TiB$_2$ [68]. During cooling down from the sintering temperature, the difference between CTEs of different phases causes a strain mismatch and thermal stresses. At sintering temperature, the atomic mobility is very high and internal stresses can quickly be relaxed by diffusion; therefore, the body is assumed to be stress free. But at lower temperatures, the atomic mobility is reduced and diffusion can no longer relax internal stresses. These stresses between the phases vary at the scale of the grains. They can be added to the external stress field. If the sum of external and internal stresses is large enough, microcracks can come into existence (at a grain boundary, in the matrix or in particles), which relax the strain mismatch. The cracking is therefore related to some inelastic strains of the ceramic (which correspond to the mismatch strains between the grains) and the “yield strength” is the stress, where cracking occurs. Again, the contribution to the specific fracture energy can be estimated using Eqn (30).

In the case of bridging mechanisms, load can be transferred across the crack borders. Bridging forces act behind the crack tip (Figure 6). Typical examples for crack bridges are fibers, which connect the borders of a crack, but bridges also exist in any material that has a rough fracture surface (which causes a geometric interlocking of the surfaces). Prominent examples are silicon nitride ceramics, which have deliberately elongated grains (the aspect ratio can be up to 1:10), but bridging is also very relevant for ceramics with coarse equiaxed grains (e.g., alumina) [69–72]. The dissipated energy can be estimated to be

$$\partial G_{Fr} \approx \int_0^{u_{\text{max}}} p(u) \cdot du, \quad \text{(31)}$$

where $p(u)$ is the stress in a “typical bridge,” which depends on the crack opening $u$ and $u_{\text{max}}$ is the opening, which causes the failing of the bridge. The bridges cause a relieve of stresses at the crack tip (a reduction of the SIF), which can be described as an increase of the fracture...
toughness (in a similar way as shown for process zone mechanisms). The bridging stresses may become very high but in monolithic ceramics, $u_{\text{max}}$ is in general very small. To give a simple example, if the bridging stress increases linearly with the crack opening and if it reaches 100 MPa at $u_{\text{max}} = 1 \mu m$, the integral Eqn (31) gives 100 J/m$^2$, and the toughness of such a material would be around 4.5 MPa m$^{-1}$ (for a material with $E = 200$ GPa). In fiber-reinforced materials, the critical crack opening may be much larger (e.g. hundred times and even more), and very high contributions to the fracture energy may result from bridging mechanisms [69].

In general, several toughening mechanisms act simultaneously, and the specific fracture energy has several additive contributions [73,74]:

\[
G_c = 2\gamma + \sum_i \partial G_{PZ} + \sum_i \partial G_{Br} + \ldots \quad (32)
\]

There exists a strong influence of the microstructure on the fracture toughness, but this influence depends on the analyzed mechanism and a short summary is in danger of oversimplifying this aspect. To give some ideas, in the case of process zone mechanisms, the increase of specific fracture energy is $\partial G_{PZ} \approx \partial \sigma_{PZ} \varepsilon_{in}$ (Eqn (30)), which—by the use of Eqn (29)—becomes $\partial G_{PZ} \approx \varepsilon_{in}/\sigma_{PZ}$. The property to be optimized is therefore $\varepsilon_{in}/\sigma_{PZ}$. In the case of toughening by plasticity, many influences on the yield strength are known (see the standard textbooks on physical metallurgy) [35]; some are solid solution hardening, precipitates, duplex structures, grain size, and so on. Of course, the hardening also has an influence on the plastic strain. In the case of transformation-toughened ceramics, the particle size has an influence on the strength where the transformation occurs (the transformation stress decreases with increasing particle size up to a critical size, where the transformation still happens without any stress). The toughening is maximum if the transforming particles are a little smaller than the critical size, but there is no toughening at all, if they are larger [75,76]. In the case of bridging mechanisms, coarse microstructures in general cause higher toughening than fine microstructures (Eqn (31)), because the maximum possible crack opening ($u_{\text{max}}$) scales with the size of the relevant microstructural feature [77,78].

As is evident from Figures 5 and 6, toughening mechanisms become only active if a crack propagates. This leads to an increase of the fracture toughness $K_c$ with crack extension of typically several times the size of the process zone from a lower starting value to a higher saturation value. The crack-length-dependent fracture toughness is also referred to as the R-curve.

Toughness measurements are based on the Griffith/Irwin criterion (Eqn (3)). A well-defined crack is made into a specimen, which is then loaded. The loading causes an SIF (Eqn (2)) that can be increased by increasing the load until it reaches the fracture toughness. Then, the specimen fails. If the fracture stress, the geometry and size of the fracture initiating crack and the geometric factor (Eqn (2)) are known, the critical SIF (the fracture toughness) can be determined.

The production of a well-defined starter crack is the most relevant problem in the fracture toughness measurement of ceramics. In the single-edge precracked beam (SEPB) method [79–81], a sharp planar crack is ‘popped in’ to a beam specimen from a row of indentations using a stiff jig system. This can be a difficult task for the inexperienced. A modified version is the single edge V-notch method (SEVNB) [82,83] in which a straight through notch is machined into a bending specimen using diamond paste on a razor blade to reach a sharp tip radius of $\leq 3 \mu m$, approaching a sharp crack. An investigation showed [84] that the SEVNB method is applicable if the grain size of the material is larger than the notch tip radius.

In the case of the surface-crack-in-flexure (SCF) method [81,85], a sharp crack is introduced by indenting a bending specimen with a Knoop indenter. To obtain a residual-stress free starting crack, a surface layer of the specimen, which contains the plastically deformed material, has to be removed (e.g. by grinding). Then, the specimen can be fractured in bending, and the size of the indentation crack can be determined fractographically.

\[\text{FIGURE 6} \quad \text{Bridging stresses can be transferred across the crack borders: (a) schematic sketch and (b) crack bridges in a silicon nitride material.}\]
Another standardized method, the chevron notch beam (CNB) method \cite{81,86}, requires a chevron notch to be machined into bending specimens. During loading the chevron notch, a sharp crack pops into the specimen, which acts as a starter crack for the fracture toughness measurement. In a European round robin concerning the practical applicability of this method, it turned out that the method is relatively expensive and difficult to perform properly \cite{87}. In particular, a stiff testing machine is needed to obtain controlled fracture, a requirement of the analysis.

If only a limited amount of material is available, the indentation fracture (IF) method \cite{88} is very commonly used. However, there exists a myriad of proposed evaluation formulas, all relying on empirical ‘calibrations’ that are often erroneous correlations with fast fracture toughness data from other methods. The computed values are usually termed the ‘IF resistance’ and are calculated from the length of the cracks emanating from the corners of Vickers indentations (Figure 2). Normally, no two formulas give the same result for a given material. Because the measurement is performed under a condition of crack arrest (instead of the situation of the moving crack becoming critical as in all other methods), the measured quantity may be somehow related to but is never equal to fast fracture toughness \cite{89}. Measurement of crack length is also subjective. The method should never be used for material comparison or specification purposes.

In summary, to the authors’ point of view, the SEVN method is a testing procedure that is easy to perform and is applicable to all types of ceramics having grain diameters of several micrometers or larger. The SCF method may be a little more precise, but the experimental procedure is more complex and time consuming, and does not always work with coarser-grained materials. Both methods employ effective crack lengths of <100 µm and are therefore short-crack methods. The SEPB and CNB methods employ longer cracks, and thus, results may additionally be affected by R-curve behavior.

Generally speaking, in terms of strength, the greater the fracture toughness, the greater is the tolerance of the material to the presence of small defects. The materials may also be more difficult to machine because this is generally a chipping process involving the propagation of small cracks. Similarly, resistance to abrasive wear may be increased by higher toughness. Some values of fracture toughness are given in Table 5.

<table>
<thead>
<tr>
<th>Material</th>
<th>Fracture Toughness (MPa√m)</th>
<th>Tensile strength (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-based materials</td>
<td>8–12</td>
<td>8000–10,000</td>
</tr>
<tr>
<td>Cemented carbides</td>
<td>7–20</td>
<td>1700–3000</td>
</tr>
<tr>
<td>Silicon carbides</td>
<td>2.5–4.0</td>
<td>350–650</td>
</tr>
<tr>
<td>Silicon nitrides</td>
<td>4.0–8.0</td>
<td>650–1000</td>
</tr>
<tr>
<td>Alumina</td>
<td>3.5–4.0</td>
<td>350–500</td>
</tr>
<tr>
<td>Zirconia, Y-TZP</td>
<td>4.5–5.0</td>
<td>200–1400</td>
</tr>
<tr>
<td>Glasses</td>
<td>0.6–0.7</td>
<td>20–150</td>
</tr>
<tr>
<td>Steels</td>
<td>&gt;50</td>
<td>300–2000</td>
</tr>
<tr>
<td>Aluminim alloys</td>
<td>22–35</td>
<td>60–550</td>
</tr>
<tr>
<td>Copper alloys</td>
<td>30–90</td>
<td>100–550</td>
</tr>
<tr>
<td>Nylons</td>
<td>2–5</td>
<td>90–160</td>
</tr>
<tr>
<td>Epoxies</td>
<td>0.4–2</td>
<td>45–90</td>
</tr>
<tr>
<td>Polyethylene</td>
<td>1.5</td>
<td>20–45</td>
</tr>
</tbody>
</table>

7. STRENGTH

In this chapter, we will focus on tensile strength as defined in Eqn (7). The tensile strength is limited by the existence of flaws, which are stress-concentrating inhomogeneities in the specimens. They are described as cracks of length \(a\) and which are distributed in the specimen or on its surface. The critical flaw, \(a_c\), is the flaw, which is loaded with the highest SIF, \(K = \sigma Y / \pi a^{1/2}\). Brittle fracture occurs, if the SIF of the critical flaw reaches or exceeds the fracture toughness of the material: \(K \geq K_c\) (Eqn (3)). The tensile strength is the stress (Eqn (7)) necessary to fulfill this condition.

In the daily routine, brittle fracture is the most important factor limiting the successful use of ceramics. The main reasons for unexpected failures are hidden or unexpected tensile stresses, which have not been accounted for in the design. The important reasons for such stresses are temperature changes or temperature gradients and contacts between the ceramic component and other objects.

An important consequence of Eqn (7) is that the tensile strength of ceramic specimens (components) is not given

9. Because—in general—the tensile stress (\(\sigma\)) in the uncracked specimen and the geometric factor (\(Y\)) can be different from position to position (the geometric factor also depends on the orientation of the crack in relation to the tensile stress components), the critical flaw must not be the crack with the largest size.

10. For simplicity and without loss of generality, we restrict the following discussion to a homogeneous and uniaxial stress field (as occurs in a tensile specimen) and to a homogeneous and isotropic flaw population in a homogeneous and isotropic elastic continuum.
by a simple number but by a distribution function. There is always some risk that a specimen fails at a very low stress (if it contains a very large flaw). It is important to understand that the distribution of strength data is caused by the size distribution of flaws. There is a direct correlation between the size of the critical flaw, the fracture toughness, and the strength

\[ a_c = \frac{1}{\pi} \left( \frac{K_c}{\sigma_f} \right), \]  

(33)

which follows from Eqn (7) and is a direct consequence of the Griffith/Irwin criterion (Eqn (3)).

A number of articles \[8,9,90-92\] have been written in the past to develop a theory of brittle fracture, which accounts for the scatter of the strength. It has been shown that for specimens containing noninteracting flaws (i.e. the flaws should be sparsely distributed), which can be described as cracks and which fail due to the weakest link criterion, the probability of failure is given by the simple relationship \[93\] as follows:

\[ F(\sigma, S) = 1 - \exp\left( -\langle N_c, S \rangle \right), \]  

(34)

where \(\langle N_c, S \rangle\) is the mean number of critical flaws. It can be shown that, using a failure criterion of the Griffith/Irwin type (Eqn (3)), \(\langle N_c, S \rangle\) depends on the applied stress, \(\sigma\) and the size of the specimen \(S\). The mean is taken over a large sample of identical specimens. At very high stresses, it is possible that a specimen contains “on average” several critical flaws (their size is small at high stresses, see Eqn (34)), and the frequency for small flaws will in general be much higher than for large flaws. But there remains a (small) probability to find no critical flaw in one specimen, and the probability of failure is therefore always a little smaller than 1.\[11\] At a very low number of critical flaws \((\langle N_c, S \rangle) \ll 1\), the probability of failure becomes simply equal to the expectation number of critical flaws in the specimen: \(F \approx \langle N_c, S(\sigma) \rangle\), that is, a series expansion of Eqn (34). It should be noted that, for the design of components, only low failure probabilities are relevant.

To evaluate \(\langle N_c, S \rangle\) the size distribution of flaws has to be known. In general, it is expected that a most frequent flaw size exists and that the frequency of flaws decreases right of the maximum with an inverse power law. This is shown in Figure 7.

In many theories, it is assumed, that the right side of the distribution can be approximated by an inverse power law

\[ g(a) \approx g_0 \left( \frac{a}{a_0} \right)^{-r}, \]  

(35)

where \(g_0\) and \(a_0\) are scaling parameters, which can be freely selected \[90,93\]. The parameter \(r\) describes the decrease of flaw number with increasing flaw size and therefore the width of the relevant size distributions of the flaws for a given specimen. Note that for volume flaws and for surface flaws, different flaw populations may exist, for volume flaws, the dimension of \(g_0\) is \(\text{m}^{-4}\) (per length and per volume) and for surface flaws \(\text{m}^{-3}\) (per length and per surface). In the following, the discussion will be restricted to volume flaws.

We get the density of critical flaws by integrating the frequency density \(g(a)\) over the flaw sizes. The integration goes from the length \(a_c\) to infinity, because all cracks with such a length are critical. Then, the mean number of critical flaws per specimen is obtained by integrating the density over the specimen volume:

\[ \langle N_c(\sigma, V) \rangle = \int_{a_c}^{\infty} g(a) \cdot da \cdot dV = V \int_{a_c}^{\infty} g(a) \cdot da. \]  

(36)

Remember that we restricted the analysis to a constant uniaxial and homogeneous tensile stress state and a solid with an isotropic, constant flaw distribution, where the integration over the volume is trivial. In the case of more general stress states, solutions can be found in the literature \[93\]. Using Eqns (33) and (35), we get

\[ \langle N_c(\sigma, V) \rangle = \frac{V}{V_0} \frac{\sigma^{2(r-1)}}{\sigma_0}. \]  

(37)

Inserting in Eqn (34) gives the well-known Weibull distribution, which describes the probability of failure in dependence of the applied stress and the specimen’s volume:

\[ F(\sigma, S) = 1 - \exp\left[ -\frac{V}{V_0} \left( \frac{\sigma}{\sigma_0} \right)^m \right]. \]  

(38)

11. Of course, at very high stresses, also other failure criterions will become important. For example, it can be expected that failure occurs, if the yield strength is reached. Such types of failure are not considered here.
This function has been derived >70 years ago by the Swedish engineer Weibull on the basis of empirical arguments [8,9]. Note that the Weibull parameters are related to the shape of the flaw population. The Weibull distribution is defined by the shape parameter $m$ and a scale parameter $\sigma_0 V_0^{1/m}$. The normalizing parameter $V_0$ can be freely selected, but the characteristic strength $\sigma_0$ depends on the definition of $V_0$. It holds that $m = 2(r - 1)$ and $V_0 \sigma_0^m = (m/2\sigma_0 a_0) \cdot (K_c/\pi a_0 Y)^m$.

These equations demonstrate that the shape of the Weibull distribution depends on the shape of the frequency distribution density of the flaws in the material. The Weibull distribution describes the experimentally observed behavior of ceramics: the strength scatters and the mean strength decreases with the specimen size (Figure 8). Note that that ‘size’ refers to the volume of the specimen if volume defects cause failure and to the surface, if surface defects are responsible for fracture.

The preceding analysis is specific and related to many model assumptions, but the results are much more general and can be extended to many other cases, for example, to surface or edge flaw distribution or to inhomogeneous and multiaxial stress fields [94–97]. In the latter case, the volume in Eqn (38) has to be replaced by an effective volume and the stress by an equivalent stress. This is described in an other chapter of this book.

The Weibull distribution is a special case out of a more general class of distribution functions for flaws whose size frequency distribution decreases with increasing flaw size corresponding to a power law (Eqn (35)). Although this occurs in many cases, it needs not to be true in any case. Therefore, the reader should bear in mind that the Weibull distribution is only a model description, which is true in many but not in all cases.

The microstructure has an influence on the strength (Eqn (7)) via the fracture toughness and the flaw population. The first aspect has been discussed in the last section. The characteristic strength can be increased by reducing the frequency of large flaws; the scatter can be reduced by a narrower flaw size distribution. In fact, both have successfully been achieved in the last 30 years to improve strength and reliability of modern ceramic materials.

Large flaws that act as fracture origins can be identified on the fracture surface of specimens. An example is shown in Figure 9. The fracture origin is an agglomerate produced by spray drying of powders, which has not been completely destroyed during the uniaxial pressing of the component. Such type of defects could be avoided by changing the spray drying conditions of the powders. This resulted in an increase of strength of nearly 20% [98].

In fact, the characteristic strength of commercial ceramic materials could be increased significantly over the last 30 years and the scatter of data could be reduced. Thirty years ago, the Weibull modulus of a state-of-the-art ceramic was between 5 and 10, and nowadays, it is between 15 and 25. This causes an enormous increase in the reliability of components. To give a simple example, we compare two samples of ceramic specimens, having Weibull moduli of $m = 5$ and $m = 25$. The characteristic strength of both samples is 500 MPa. If the specimens are loaded with a stress of 250 MPa, 3% of the specimens fail in the first, but only $3 \cdot 10^{-6}$% in the second sample. In other words, the risk of failure at half of the characteristic strength is reduced by a factor $10^6$ by increasing the Weibull modulus from 5 to 25.

As shown earlier, the strength of brittle materials is related to the defect population(s) present in the specimen or component (cf., Eqn (7)). Due to the distribution of defect sizes, it exhibits an intrinsic scatter. Following a statistical approach, a strength distribution describes adequately the strength of brittle materials, which is typically defined by a Weibull distribution and the corresponding Weibull parameters (i.e. shape and scale parameter).

![FIGURE 8](image1) Characteristic strength versus (effective) surface for an alumina ceramic. Test results on bending specimens (▲) as well as on biaxially tested discs (○) and rectangular plate specimens (■) are shown [99].

![FIGURE 9](image2) Fracture origin in a standard flexure specimen. Sickle-shaped pores indicate the location in a powder agglomerate that was not completely destroyed during pressing.
The basic goal of strength measurements is to determine the previously unknown strength distribution parameters of a material as accurately as possible. The measurement of a sample containing a finite number of specimens reflects this intrinsic scatter, and due to statistical reasons, the result of a such strength measurements can only deliver the distribution parameters within a certain confidence range according to a given confidence level (i.e. confidence intervals for both Weibull parameters). Consequently, meaningful results of strength measurements are not unique numbers for the distribution parameter but certain confidence intervals, whose widths depend on the confidence level and the sample size.

Strength tests for ceramics should be so simple to perform that a sufficient number of tests can be easily accomplished to obtain a sound basis for the statistical evaluation. The standards suggest a minimum of 30 specimens per condition as a tradeoff between experimental effort and a sensible small confidence bandwidth [100–103].

Tensile tests are in principle possible but require expensive specimens and sophisticated testing equipment to provide true uniaxial stressing, that is, without excessive superimposed bending [104].

The strength of ceramics is thus preferentially measured in bending. Bending tests can be performed under either uniaxial [105–109] or biaxial loading conditions [110]. Uniaxial bending tests are usually performed on long, thin (length over thickness ratio $> 10$) prismatic beams or cylinders. For such specimens, simple expressions from the linear-elastic theory can be used for the calculation of the strength from the fracture force and the geometry. The possible measurement error is well investigated [111,112] and is in the range of 2–10% for specimens with a length $> 10$ mm. The lower limit of this error budget can be achieved when tests are performed with utmost care. The upper limit may be too much to determine high Weibull moduli above $m > 20$. The experimental scatter then masks the inherent strength scatter [113].

Biaxial strength tests may have advantages over uniaxial bending tests if, for instance, specimen machining can be omitted because components already have a suitable geometry. Furthermore, biaxial stresses are often highly relevant for applications, as in the case of (cyclic) thermal loading. Most of the suggested biaxial bend tests, that is, ball-on-ring, ring-on-ring, or punch-on-ring tests suffer from the discrepancy between the actual loading situation and the idealized descriptions, which are assumed to derive the evaluation formulas. The maximal stresses occurring in the specimen strongly depend on the contact situation (at the load application points). Analytical solutions are usually only approximations, which may differ significantly from the real situation [114]. An exception is the ball-on-three-balls test, which is evaluated using a numerical fit on parametric finite element method (FEM) calculations [114].

(A tool for the evaluation of this test is available on http://www.isfk.at/de/960/). This test is also extremely tolerant against imprecisions of specimens or testing equipment. With this test, a measurement error around $\pm 1\%$ can even be achieved with specimens with diameters of approximately 2 mm and as thin as 500 $\mu$m [115].

For design purposes, not only the magnitude of a materials’ strength but also its scattering property play a significant role. Using, for example, the Weibull distribution as the basis, it is possible to calculate a tolerable load level to achieve a certain short-time reliability or vice versa. For such calculations, it is indispensable to use representative values for the strength distribution, that is, those that are caused by the same defect population that causes fracture during operation [116]. In the standards for bending tests, requirements for the quality of the tensile surface of the bend bars are given to test the failure strength (distribution) as a material property and not as a consequence of a certain machining procedure. The so-measured strength distribution may be the relevant one for failure during operation, but it can also be an irrelevant one, if, for example, other parts of the components that are in the as-sintered state and certainly have a different defect population are under sufficient high stress. Another issue that complicates reliability assessment is the fact that strength is usually tested on different (effective) volumes (or surfaces) than the components have. The volume dependence of strength may be another reason why critical defect populations are different in specimens and components [117].

8. STRENGTH DEGRADATION WITH TIME

In the last section, we focused on fast brittle fracture, that is, on failures in which the crack velocity almost reaches the speed of the sound. Fracture could therefore be described as instantaneous. But in practice, brittle fracture has also been observed to occur some time after load application. This can arise for several reasons (e.g. slow plastic deformation (i.e. creep), corrosion, cyclic fatigue, etc.). In this section, we will focus on the so-called slow or subcritical crack growth (SCCG), which is the most relevant time-dependent damage mechanism in ceramics. It not only occurs at low and intermediate temperatures under the action of a constant load but also under varying loads and causes a progressive decrease of remaining strength with time.

The first direct observations of SCCG were made in glass [118]. It was recognized that SCCG is thermally activated and also strongly depends on the environment. Tests in a humid atmosphere show (up to some orders of magnitudes) larger crack growth rates than in vacuum. Similar observations have also been made in the presence of other polar molecules [11]. The crack growth rate also strongly depends on the SIF ($K_I$; Eqn (2)). Very slow growth rates
(<10^{-13} \text{ m/s}) have been reported for SIF values significantly lower than those giving fast fracture [62], but if the SIF is only a little smaller than the fracture toughness, the crack growth rate reaches almost the speed of sound (i.e. >10^3 \text{ m/s} for most ceramic materials). So the observed SCGG rates vary over the enormous range of 17 orders of magnitude. This is an extremely wide range for a parameter (for comparison, the age of the world is about 10^{17} \text{ s}).

Several crack growth mechanisms have been proposed to occur in the SCGG range, but a clear understanding is still missing. Therefore, the statements in the next paragraphs are the current state of understanding. To describe SCGG in vacuum, thermally activated breaking of bonds at the crack tip has been proposed [119,120]. At high crack propagation rates, this is also the dominating mechanism. In polycrystalline ceramics, this often leads to transgranular fracture. In chemically inert environments or in vacuum, thermally activated bond breaking may even be responsible for SCGG at very low and intermediate growth rates.

At very slow crack extension rates, there exists a strong influence of the environment. Many ceramics contain (glassy) silicate grain boundary phases. Here, water molecules can hydrolyze silicate bonds at the crack tip, and this process can be many orders of magnitude faster than the thermally activated breaking of the silicate bonds. The chemical reaction at the crack tip and the subsequent breaking of the hydrolyzed bonds is the rate-controlling factor. At intermediate growth rates, the transport of water molecules at the crack tip may become rate limiting. In polycrystalline ceramics, this SCGG mechanism promotes intergranular crack propagation.

A consequence of SCGG is that the lifetime of components could lie somewhere between some hours and some tens of years. This lifetime range is caused by SCGG rates in the environmentally influenced low crack growth rate regime, that is, at crack growth rates between 10^{-13} and 10^{-5} \text{ m/s}. In this regime, the crack growth rate (v) is often empirically described by the so-called Paris type crack growth law:

\[ v = v_0 \left( \frac{K}{K_c} \right)^n, \]  

with the (temperature- and environment-dependent) material parameters \(v_0\) and the SCGG exponent (Paris exponent) \(n\).

For the following discussion, it is important to realize that the SCGG exponent in general has a very high value. Some SCGG data for a commercial alumina ceramic [121] and a commercial silicon nitride [122] are shown in Figure 10. At room temperature, the SCGG exponent is \(n \approx 33\) for alumina and \(n \approx 50\) for silicon nitride, respectively. These data are typical for materials with glassy silicate grain boundary phases. For ceramics that do not have a glassy grain boundary phase, values around \(n \approx 200\) are reported, for example, for silicon carbide ceramics [123]. For the silicon nitride ceramic, the crack growth rate increases significantly if the temperature reaches or exceeds the softening temperature of the glassy grain boundary phase, and the exponent decreases markedly.

Very little is known about the influence of the microstructure on the SCGG behavior of ceramics apart from the fact that silicate grain boundary phases can accelerate the crack growth rate significantly. If the strong dependence of SCGG on stress, or more precisely, on the SIF, is considered, it seems obvious that local internal stresses may have a tremendous influence on the growth rate. For the same reason, it is obvious that the size of the flaw that extends under the action of SCGG also has a significant influence on the growth rate. This point will be discussed in more detail in the paragraphs on testing procedures. In any case, reducing the size of the internal flaws through improved

![FIGURE 10](image-url) Crack growth rate versus SIF for (a) an alumina replotted from Ref. [121] and (b) a silicon nitride ceramic, replotted from Ref. [122]. Paris laws are indicated by straight lines.

12. It should be emphasized that this is only a model that is broadly matched by experiment, but there is no physical reason why \(n\) should be a constant over the entire crack growth velocity range.
processing is one of the best strategies to increase the lifetime of components under stress. Another strategy is to avoid grain boundary phases.

Crack growth data obtained on test pieces can be used to predict the lifetime of components. Under the action of a stress \( \sigma(t) \), which may be time dependent, a crack grows from its initial length \( a_i \) to the critical final (Griffith) length \( a_c = a_f \), at which point failure occurs almost immediately.

The time to failure (lifetime) can be determined in the following way: the crack growth rate is the differential of the crack length with time \( v = da/dt \). Then, the differential of the time is \( dt = (1/v) \cdot da \). By separation of variables and using Eqns (2) and (39) we get:

\[
\int_{t_0}^{t_f} \sigma(t)^n \, dt = \frac{1}{v_0} \int_{a_i}^{a_f} \left( \frac{K_c}{Y\sqrt{\pi a}} \right)^n \, da.
\]

(40)

If we assume that small cracks already exist in the specimen (or component), we can set \( t_i = 0 \). For such small cracks, which are commonly present in most ceramic materials, the geometric factor \( Y \) is a constant, and the integration can be made analytically. For \( n > 2 \), it holds that

\[
\int_{0}^{t_f} \sigma(t)^n \, dt = \frac{2}{n-2} \frac{a_i}{v_0} \left( \frac{K_c}{Y\sqrt{\pi a_i}} \right)^n \left( 1 - \left( \frac{a_i}{a_c} \right)^{(n-2)/2} \right).
\]

(41)

The right-hand side of Eqn (41) is a constant value. Therefore, the stress dependence of lifetime results from the left-hand side, so the lifetime strongly depends on the applied stress and its history. It should be noted that if significant crack growth occurs, the second term in the square bracket can be neglected because \( n >> 1 \). This shows that the most prominent material properties influencing the lifetime are the initial crack length \( a_i \) and the Paris exponent \( n \).

Measurements of SCCG have been made using specimens with artificial cracks [124]. The crack advance after progressive time intervals can be determined using microscopic techniques. If a crack propagation distance of, for example, 10 \( \mu \)m is recorded after one day, the determined crack velocity is \( v \approx 10^{-5} \text{ m/s}^2 \approx 10^{-10} \text{ m/s} \). This demonstrates that measuring very slow crack growth rates on artificial (long) cracks needs very long testing times (at least three months) under difficult conditions and is therefore not practicable. In addition, the tests usually required long cracks that can be reliably measured, but because components generally do not have long cracks in them while still serviceable, it is questionable whether such macrocrack growth data are equivalent to microcrack growth that is the usual source of delayed failure in components. Instead, to simulate the latter, measurements are normally based indirectly on the growth of natural flaws. Two different testing procedures are frequently used: specimens are either loaded at a constant stress and the time to failure is recorded or they are loaded with a linearly increasing stress (constant stress rate) until failure occurs. These situations will now be discussed in more detail.

Experiments under constant load \( (\sigma(t) = \sigma_{\text{stat}}) \) are made in most cases in bending. Very simple testing equipment can be used, for example, the stress can be applied by simply loading the specimens with dead weights. However, gaining the data for a proper \( v-K \)-curve may still require three months or even more.

For constant load testing, the solution of Eqn (41) is straightforward, giving

\[
t_f = \frac{2}{n-2} \frac{a_i}{v_0} \left( \frac{K_c}{Y\sqrt{\pi a_i \sigma_{\text{stat}}}} \right)^n \left( 1 - \left( \frac{a_i}{a_c} \right)^{(n-2)/2} \right) = \frac{2}{n-2} \frac{a_i}{v_i} \left( 1 - \frac{a_i}{a_c} \frac{v_i}{v_f} \right).
\]

(42)

The indices \( i \) and \( c \) refer to the initial and final (Griffith) crack length, respectively, \( v_f \) is the final crack velocity.

In experimental practice, the scatter of the defect sizes has to be taken into account. For example, at a fixed stress level, a distribution of failure times can be determined, which reflects the distribution of initial defect sizes, \( a_i \). This distribution is a Weibull distribution. If the Weibull distribution of the strength (which should be determined in such a way that SCCG has only negligible influence, i.e. it should be the inert strength distribution) is known, the Paris parameters can be determined by comparison of both distributions. For example, the modulus of the distribution of lifetimes is \( m' = m/(n-2) \). The lifetime distribution is very wide, because typical values of \( m \approx 10-20 \) and of \( n \approx 30 \); therefore, \( m' \approx 1/3-2/3 \). In the case of ceramics without glassy grain boundary phases, the SCCG exponent is much larger, the modulus of the distribution of lifetimes is smaller, and the distribution of times to failure even wider. Therefore, some early failures can arise if the experimental design is not properly made. It follows from Eqn (42) that the time to failure is proportional to \( t_f \propto a_i^{-(n-2)/2} \cdot \sigma_{\text{stat}}^{-n} \). If the high value of the SCCG-exponent is considered, it is immediately clear that a reduction of the stress level is a very effective measure to increase the lifetime of a component and to reduce the probability of early failure.

An alternative approach to determining the SCCG exponent is to test the subsets of test specimens at different constant stress levels. It can be expected that for every subset at the 50% percentile value of the lifetime, the initial
crack \((a_i)\) had almost the same length, and the SCCG-exponent can be determined from the relationship:

\[ t_f \propto \sigma_{\text{stat}}^{-n}. \]

As a result of the high value of the SCCG exponent, the development of the crack length with time under constant load may appear a little surprising. This is shown in Figure 11 for three hypothetical materials having different SCCG exponents. The increase in crack size is very slow at the beginning of the experiment. The crack needs a large fraction of its lifetime to grow just a few percent in length. Then, it quickly accelerates, and brittle failure occurs. This explains why failure due to SCCG occurs quite unexpectedly. Recognized cracks seem to be stable for a long time, and changes in crack length are so small that they can hardly be determined. But after some “incubation time,” almost instantaneous brittle fracture occurs.

The second frequently used type of test for the indirect determination of \(v-K\) curves is the constant stress rate test: 

\[ \sigma = \dot{\sigma} \cdot t \quad [125]. \]

In this case too, tests are in general conducted in bending. For load application, a universal testing machine is necessary, which can precisely be regulated even at slow loading rates. The left-hand side of Eqn (41) can be integrated analytically. After some rearrangement, we get

\[ \sigma_i^{n+1} = \frac{2(n+1)}{n-2} \frac{a_i}{\nu_0} \sigma_i^n \dot{\sigma} \left[ 1 - \left( \frac{\sigma_f}{\sigma_i} \right)^{n-2} \right], \quad (43) \]

where \(\sigma_i = K_c/(\sqrt{\pi a_i})\) is the inert strength (i.e., the strength of the specimen without any SCCG) and \(\sigma_f\) is the determined fracture strength of the test specimen. Of course, \(\sigma_f \leq \sigma_i\). Again, if any significant SCCG occurs, the second term in the square bracket is very small compared with unity and can be neglected. In this region, the strength depends on the stress rate according to \(\sigma_f \propto \dot{\sigma}^{1/(n+1)}\) (see also Figure 12; region \(\dot{\sigma} < 10^5 \text{ MPa/s}\)). At higher stress rates, the test duration is too short for significant SCCG to occur, and the strength almost equals the inert strength. Data as shown in Figure 12 can be used to derive the \(v-K\) curve of the material.

FIGURE 11 Crack length versus time for materials having a different SCCG exponents \(n\) (using typical data for alumina, silicon nitride and silicon carbide, respectively, and setting \(\sigma_{\text{stat}}/\sigma_i = 0.95\) in relative units. The cracks need a large fraction of their lifetime to grow the first few percent. They then accelerate, and fracture occurs very quickly. This behavior is strongly dependent on the SCCG exponent.

FIGURE 12 (Mean) strength versus stress rate for constant stress rate tests performed on an alumina ceramic [126].

Compared with tests with constant load, constant stress rate tests offer some advantages. In particular, the time of individual tests can be estimated much easier. But in this type of test too, very slow crack growth rates can only be assessed with very time consuming and expensive experiments. The determination of a \(v-K\) curve (in the technically relevant parameter regime) may need about 3 months.

Due to the expensive and long-lasting experiments necessary to measure a \(v-K\) curve, data on SCCG behavior are very rare and determined in most cases at room temperature and in laboratory air. For these conditions, a typical SCCG exponent for materials having silicate grain boundary phases is \(n = 30-50\). Recent measurements on low temperature co-fired ceramics (LTCCs), which are nowadays used for integrated circuit boards, and which can be used at elevated temperatures and under severe mechanical conditions (vibrations), gave similar values [127,128]. The same happens for the SCCG exponent of glasses. In SiC ceramics, which have no glassy grain boundary phases, an SCCG exponent of about \(n \approx 200\) has been determined using this technique [123].

The technical significance of SCCG is the fact that the strength degrades with time. A desired reliability cannot be guaranteed for infinite time, and the reliability decreases with time. These aspects, which are highly relevant for a successful use of ceramics, have to be considered in the design [129]. Allowable stress levels must be modified to account for SCCG, and the lower the value of the exponent, the bigger the required reduction in allowable stress.

9. FINAL REMARKS

The mechanical properties discussed so far are used to analyze the response of ceramic components to stress and to predict their reliability and lifetime. But fractographic evidence tells us that there are three, often hidden reasons...
for additional causes of stresses in components, which are often ignored in stress analyses and which are responsible for unexpected failures: internal residual stresses, imposed thermal stresses, and stresses arising in contact situations. We will conclude this chapter with some comments on the effects of these and on the need to guarantee properties in critical components.

Generally distributed internal residual stresses can arise inside components as a consequence of either thermal expansion mismatches from one area to another, or because some regions of the component are cooled much faster than others during the final stages of the sintering or densification cycle. They are more prevalent in larger monolithic items than in smaller ones, and are best removed by the use of appropriate annealing conditions. In some cases, they are deliberately introduced, such as in multilayered materials, but they have a strong influence on strength and reliability. In addition, machining can introduce surface residual compressive stresses. Often, these are beneficial in reducing the ability of small surface cracks to propagate, but they strongly depend on the manner in which the machining is undertaken. For predictive purposes, it is crucial to match the machining conditions of specimens for experimental determinations of strength with those that are employed in manufacturing components.

Thermal stresses can also arise from temperature differences between one part of the component and another. Unlike in metal component designs where the consequences of such stresses are often ignored, in ceramic components, they are of much more importance, adding to imposed mechanically derived stresses. Often, the cooler parts of components are subjected to tensile stress, while the hotter parts are in balancing compression, assuming a positive expansion coefficient. Stresses are clearly reduced in lower thermal expansion materials, which tend to drive material choice for thermal applications.

A special case of thermal stress arises when temperatures are changed very rapidly, particularly during rapid cooling. Then, the surface of the ceramic is subjected to high tensile stress, which if too large is relieved by surface crazing. Again, such effects can be mitigated by using materials possessing low thermal expansion characteristics.

The third type of additional stresses are those due to mechanical contact. Any form of load transfer through localized contacts produces highly localized stress concentrations. Usually associated with such contacts are high levels of local tensile stress, which can by themselves introduce strength-reducing cracks. It should also be noted that machined ceramics have rough surfaces, so proper modeling of contacts is difficult. Often, some means of soft load spreading through metal or polymer layers is needed. Designing "benign" load transfer is essential to minimize such effects.

A final issue that needs discussion is the means that can be taken to ensure that any individual component has the desired strength properties. The only practical solution to this problem is to undertake mechanical proof testing, that is, to apply a stress level to the component, which will ensure that it will survive in the application. There are two strategies for deciding what stress to use; this can be done by analysis of subcritical growth and of probabilistic risks of failure under the stress distribution applied (using the approaches described in the preceding sections) or this can be done in a pragmatic manner with the intention of removing the weak end of the distribution of component strengths. The former requires detailed relevant input data concerning strength statistics and SCCG and is seldom undertaken. The latter approach, combined with design stress minimization, is more commonly employed and successfully reduces the risks of failure in service. By assuming that proof testing will be employed, an appropriate maximum design stress level can be estimated for the ceramic by a simple estimate of Weibull modulus and the SCCG parameter. In this way, the advantageous properties of ceramics, such as hardness, dimensional stability, wear resistance, electrical insulation, can be exploited while minimizing the risks of brittle fracture.

REFERENCES

Mechanical Properties of Ceramics


Cranmer DC, Richerson DW, editors. Mechanical testing methodology for ceramic design and reliability. New York: Marcel Dekker; 1998. 434.


Baratta FI, Matthews WT, Quinn GD. Errors associated with flexure testing of brittle materials. Watertown: U.S. Army Materials Technology Laboratory; 1987, 43.


Testing and Evaluation of Mechanical Properties

Tatsuki Ohji
National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

1. INTRODUCTION

Because of their brittle nature, ceramics show totally different fracture behavior from that of metals and polymers. Ceramic engineers should properly understand this unique behavior when designing ceramic components for safe and reliable use. Fracture behavior of ceramics can be clarified through proper mechanical tests suitable to materials and their applications; so mechanical tests comprise a significant part of ceramic technology. Over the years, substantial progress has been made in our understanding of mechanical properties of ceramics as well as test methods needed to determine those properties.

This chapter addresses mechanical properties of advanced ceramics; particular interest is given to fast fracture strength and time-dependent failure properties such as slow crack growth, fatigue, and creep. The testing and evaluation procedures are reviewed for establishing the reliability specific to each of these properties, together with some related research topics. Numerous international and domestic standards related to these procedures have been issued in recent years, which are also listed in this chapter. More detailed information on the testing and evaluation procedures is available in these standards or other literature [1]. Fracture toughness, which is one of the most important parameters for expressing brittle fracture of ceramics, is not discussed here in detail however, since it is dealt with in the Chapter, “Fracture Mechanics Measurements.”

2. FRACTURE STRENGTH

Fracture strength is one of the most common and widely used mechanical properties for structural ceramics because it is the most simple and important material reliability parameter. Ceramic component design demands critical investigation of fracture strength and its distribution due to the nature of probabilistic fracture of ceramics.

A number of methodologies and techniques have been developed for determining ceramics fracture strength [2–32]. Most of these techniques equate fracture strength to maximum stress (tensile or compressive) at fracture. Consequently, in order for a particular load and specimen geometry to be useful for determination of fracture strength, stress distribution must be well established. A complicating factor in the determination of fracture strength is that ceramic materials strength is quite sensitive to size, shape, and surface finish. Such sensitivity is largely responsible for wide variation in strength values reported for a given material. Viable test methodology must therefore account for these effects. This section describes the commonly used methods, typically divided into four types: flexural tests, tensile tests, compressive tests, and biaxial tests, followed by the related standards.

2.1. Testing Methods

2.1.1. Tensile Tests

Uniaxial tensile tests have been widely used for many years in order to obtain information on mechanical strength...
properties of industrial materials. However, in the field of ceramics, flexural tests are commonly used instead of tensile tests. This is because stress concentration occurring during tests cannot be relaxed by local plastic deformation in ceramics; hence, even slight eccentric loading in tensile tests causes parasitic bending stress. Susceptibility of stress concentration also causes improper fractures inside the gauge section. Especially in high-temperature measurements, the degree of difficulty increases because material and structure of the testing fixture are restricted. On the other hand, ceramic elasticity is usually maintained until fracture occurs, enabling researchers to determine strength in flexural tests by elementary stress elastic beam equations. In addition, flexural tests are superior to tensile tests due to the simple specimen geometry, low cost of specimen preparation, and adaptability to high-temperature measurements. Nevertheless, tensile tests have the following advantages. First, fracture strength and its distribution can be determined from the stress field within large effective volumes of the material. This is most useful for reliable design of large mechanical components. Second, fractures in tensile tests more frequently originate from internal natural fracture defects than those in flexural ones, resulting in more intrinsic strength. Third, even in the high-temperature region where the material shows nonelastic behavior and validity of stress calculation based on the elastic body in flexural tests is questionable, tensile tests can determine stress precisely and simply. Also, in such a region, tensile tests are most advantageous for determining stress—strain curves, and creep curves, because deformation amounts can be detected properly with an appropriate extensometer. Examples of nonlinear stress—strain curves of silicon nitride at high temperatures in tension are shown in Figure 1.

As already stated, there are two major barriers to ceramic tensile tests. One is eccentric loading which results in unexpected parasitic bending strain. The other is stress concentration inside the gauge section. In general cases, the latter barrier can be overcome by employing the long gauge section, which is distant from the loading points with moderate change of cross section (which is due to St. Vennant’s effect). For the first barrier, there are several sources of eccentricity including off-center loading, end bending moment, end torque moment, and an imperfect or curved specimen. Difficulty in solving the first barrier exists in these sources being involved with it in the complex. Sources include an improperly aligned load train, improper specimen/grip interface, imperfect (e.g. curved or nonsymmetric) specimen geometry, and so on. Examples of successful tensile testing ceramics are available in Refs. [2—15].

The geometry of tensile test specimen depends on the purpose of the tensile test itself. When a tensile test is performed for knowing the tensile strength of an as—fabricated component, its dimension limits the dimensions of the specimen. Gripping devices also restrict the design of the test specimen geometry. The range of possible tensile test specimen geometries is illustrated in ISO 15490, “Fine ceramics (advanced ceramics, advanced technical ceramics) — Test method for tensile strength of monolithic ceramics at room temperature,” and ASTM C 1273, “Standard test method for tensile strength of monolithic advanced ceramics at ambient temperatures.”

**FIGURE 1** Stress—displacement curves showing effect of strain rate for a hot-pressed silicon nitride at 1260 °C.
2.1.2. Compressive Tests

Although the compressive strength of ceramic materials is generally much higher than tensile strength, the dominant loading mode in many actual applications is compression. Reliable component design in these applications requires precise compressive strength measurements. Compressive strength measurements are also necessary for development of fundamental insights into micromechanical and statistical aspects of compressive failure.

Compressive failure is thought to involve coalescence of damage in the form of microcracks and microvoids which propagate by alternating growth and arrest mechanisms [16]. For ceramic materials tested at ambient temperatures, this damage is thought to be generated by localized microplasticity arising from twinning and slip [17]. Failure occurs through structural collapse when density of damage in a zone reaches a critical level. Because this size appears to depend only upon intrinsic material characteristics, compressive strength is not generally affected by specimen size.

For uniaxial testing, compressive stress is given simply by the applied load divided by the cross-sectional area. However, high stresses required for failure under compressive loading have created a number of problems in test fixture design and specimen geometries required for successful compression strength measurement. Particularly, problems related to improper alignment and load block stress concentrations can lead to the generation of tensile stresses in the specimen sufficient to cause failure. Therefore, it should be noted that measured compressive strength tends to underestimate true values. Examples for successful measurements of compressive strength in ceramics are available in References [16–21].

In recent years, compressive tests have been most frequently used for measuring strength of porous ceramics, which are now expected to be adopted for a wide variety of industrial applications from filtration, absorption, catalysts, and catalyst supports to lightweight structural components [21–24]. It has been known that the strength widely scatters, by one to two orders of magnitude, even for the same porosity, depending on the porous structure, etc. [24].

2.1.3. Flexural Tests

Due to advantages including simple specimen geometry, low-cost specimen fabrication, and adaptability to high-temperature measurements, flexural tests are most widely used for measuring ceramic materials strength [25–41]. Compared to other testing methods such as tension and compression, specimens and testing fixtures for flexural tests are much simpler and, consequently, less expensive. Flexural tests, however, are disadvantageous by vulnerability of strength to surface machining, small effective volume of specimen, and other reasons. When resultant flexural strength data are used for component design, one must assume that strength limiting flaw distribution (pores, surface flaws, agglomerates, inclusions, etc.) is equivalent to other methods. Therefore, surface machining processes for test specimens are extremely important; when machining processing causes severe surface or subsurface damage, it becomes difficult to obtain intrinsic strength. This tendency increases with higher material strength. Such materials have limited fracture origin size and require damage-less surface treatment.

Two kinds of flexure loading are commonly used: three-point loading and four-point loading (Figure 2). In four-point flexure, pure flexure moment without shear stress is generated within an inner span and beam theory can be applied more properly for strength calculation than in three-point flexure. Also, four-point flexure has larger effective volume or effective surface than three-point flexure for identical specimen dimensions. Still, three-point loading has simpler geometry and simplifies testing fixture design. This is particularly important when adapted to high-temperature measurements or with limited available specimen size. The fracture strength, \( \sigma_f \), is equated to the maximum tensile stress generated at the failure load, \( P \), as follows:

\[
\sigma_f = \frac{1.5P(L_o - L_i)}{bh^2}
\]

where \( L_o \) and \( L_i \) are the outer and inner load spans, respectively (\( L_i \) is zero for three-point loading), and \( b \) and \( h \) are the width and height of the rectangular specimen.

![FIGURE 2 Three-point and four-point loading geometries.](image)
In recent years, application of ceramic materials has been expanded widely and flexural strength measurements have been reported for various types of ceramics including lead zirconate titanate (PZT) ceramics [33–35], armor ceramics [36], ultra-high-temperature ceramics [37], and dental ceramics [38–40]. Flexure tests have been also used for determining strength of ceramic tubes specifically for circumstances when flaws located at the tube’s outer diameter are the strength limiters (e.g., heat exchangers) [41].

2.1.4. Biaxial Tests

Strength measurements are usually performed in uniaxial loading. Most ceramic components are subjected to biaxial (or multiaxial) loading in actual practice. In addition, biaxial loading conditions are much more severe in a stress state than uniaxial conditions, leading to more reliable component design. One of the most frequently used biaxial loading techniques is biaxial flexure using thin plates as specimens. Biaxial flexure is free from the specimen edge effects which substantially affect strength measurements in three- or four-point flexural tests. This technique also can reflect flaws of any direction. In biaxial stress fields, a larger number of flaws are subjected to tensile stress than in uniaxial stress fields, resulting in more conservative strength evaluation. In addition, simplified loading geometry eases specimen preparation and lowers its cost.

Several types of loading methods have been employed to evaluate biaxial flexural strength: ball-on-ring, piston-on-ring, and ring-on-ring. [42–57]. The ring can be substituted for balls placed at the circumference of the same radius since applied stresses are independent of angles between and the number of supports for this loading geometry. The ball-on-three-ball and the piston-on-three-ball formations are therefore very frequently used instead, as shown in Figure 3. One advantage of the ball-on-three-ball is its tolerance of rough specimen surfaces, simplifying surface machining. The fracture strength, \( \sigma_f \), can be calculated from the failure load, \( P \), as follows:

\[
\sigma_f = \frac{3P(1 + \nu)}{4\pi t^2} \left[ 1 + 2 \ln \frac{a}{b} + \left( 1 - \frac{a}{b} \right) \left( 1 - \frac{b^2}{2a^2} \right) \right] \tag{2}
\]

where \( a \) is the radius of the support ring, \( R \) and \( t \) are the specimen radius and thickness, respectively, and \( \nu \) is the Poisson’s ratio. \( b \) is the radius of the piston for the piston-on-ring, that of the contact area of the loading ball for the ball-on-ring, or that of the loading ring for the ring-on-ring. Similar to other strength measurements of ceramics, biaxial strength of ceramics is affected by the effective volume or area of the stressed portion of the specimen [49].

Thin plates and minute components are now used in many technical applications including semiconductor devices, information storage media, and biomedical devices, and biaxial flexural tests have been used for the strength evaluation of ceramic components in such applications [50,51]. In recent years, this test method also has been employed most frequently to determine strength of dental materials [52–56].

2.2. Related Standards

In these past few decades, considerable efforts have been made to standardize fracture strength measurements for both room and high temperatures. Applicable general standards for strength measurement tests of advanced ceramics are as follows (some additional standards for specific uses not included):

- ISO (International Organization for Standardization), TC206 “Fine ceramics (advanced ceramics, advanced technical ceramics)”
- ISO 13124: Test method for interfacial bond strength of ceramic materials
- ISO 14704: Test method for flexural strength of monolithic ceramics at room temperature
- ISO 15490: Test method for tensile strength of monolithic ceramics at room temperature
- ISO 15733: Test method for tensile stress–strain behavior of continuous, fiber-reinforced composites at room temperature
- ISO 17565: Test method for flexural strength of monolithic ceramics at elevated temperature
- ISO 20501: Weibull statistics for strength data
- ISO 20504: Test method for compressive behavior of continuous-fiber-reinforced composites at room temperature
- ISO 20505: Determination of the interlaminar shear strength of continuous-fiber-reinforced composites at ambient temperature by the compression of double-notched test pieces and by the Iosipescu test
- ISO 20506: Determination of the in-plane shear strength of continuous-fiber-reinforced composites at ambient temperature by the Iosipescu test

ASTM International (formerly known as the American Society for Testing and Materials), C28 “Advanced Ceramics”

- C1161: Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperature
- C1211: Standard Test Method for Flexural Strength of Advanced Ceramics at Elevated Temperatures
- C1239: Standard Practice for Reporting Uniaxial Strength Data and Estimating Weibull Distribution Parameters for Advanced Ceramics
- C1273: Standard Test Method for Tensile Strength of Monolithic Advanced Ceramics at Ambient Temperatures
- C1275: Standard Test Method for Monotonic Tensile Behavior of Continuous Fiber Reinforced Advanced Ceramics with Solid Rectangular Cross Section Test Specimens at Ambient Temperature
- C1292: Standard Test Method for Shear Strength of Continuous Fiber Reinforced Advanced Ceramics at Ambient Temperatures
- C1322: Standard Practice for Fractography and Characterization of Fracture Origins in Advanced Ceramics
- C1323: Standard Test Method for Ultimate Strength of Advanced Ceramics with Diametral Compressed C-Ring Specimens at Ambient Temperature
- C1358: Standard Test Method for Monotonic Compressive Strength Testing of Continuous Fiber Reinforced Advanced Ceramics with Solid Rectangular Cross Section Test Specimens at Ambient Temperatures
- C1359: Standard Test Method for Monotonic Tensile Strength Testing of Continuous Fiber Reinforced Advanced Ceramics With Solid Rectangular Cross Section Test Specimens at Elevated Temperatures
- C1366: Standard Test Method for Tensile Strength of Monolithic Advanced Ceramics at Elevated Temperatures
- C1424: Standard Test Method for Monotonic Compressive Strength of Advanced Ceramics at Ambient Temperature
- C1425: Standard Test Method for Interlaminar Shear Strength of 1 D and 2 D Continuous Fiber Reinforced Advanced Ceramics at Elevated Temperatures
- C1468: Standard Test Method for Transthickness Tensile Strength of Continuous Fiber Reinforced Advanced Ceramics at Ambient Temperature
- C1469: Standard Test Method for Shear Strength of Joints of Advanced Ceramics at Ambient Temperature
- C1499: Standard Test Method for Monotonic Equibiaxial Flexural Strength of Advanced Ceramics at Ambient Temperature
- C1557: Standard Test Method for Tensile Strength and Young’s Modulus of Fibers
- C1674: Standard Test Method for Flexural Strength of Advanced Ceramics with Engineered Porosity (Honeycomb Cellular Channels) at Ambient Temperatures
- C1683: Standard Practice for Size Scaling of Tensile Strengths Using Weibull Statistics for Advanced Ceramics
- C1684: Standard Test Method for Flexural Strength of Advanced Ceramics at Ambient Temperature Cylindrical Rod Strength

CEN (European Committee for Standardization), TC184 “Advanced Technical Ceramics”

- EN 658-1: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Part 1: Determination of tensile properties
- EN 658-2: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Part 2: Determination of compression properties
- EN 658-3: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Part 3: Determination of flexural strength
- EN 658-4: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Part 4: Determination of interlaminar shear strength by compression loading of notched test specimens
- EN 658-5: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Part 5: Determination of interlaminar shear strength by short span bend test (three points)
- EN 820-1: Advanced technical ceramics — Methods of testing monolithic ceramics — Thermomechanical properties — Part 1: Determination of flexural strength at elevated temperatures
- EN 843-1: Advanced technical ceramics — Mechanical properties of monolithic ceramics at room temperature — Part 1: Determination of flexural strength
- EN 843-5: Advanced technical ceramics — Mechanical properties of monolithic ceramics at room temperature — Part 5: Statistical analysis
3. FATIGUE AND SLOW CRACK GROWTH

Fatigue, or fracture after loading of some duration at a stress level less than that required for immediate fracture, is caused by several mechanisms. (While the term fatigue in the metallurgy community indicates solely cyclic fatigue, cyclic mechanical loading of a specimen, or a component, fatigue for ceramic and glass includes static and dynamic fatigue as well.) One well-known mechanism is environmentally enhanced crack growth, very frequently referred to as slow crack growth (SCG), which occurs in nearly all structural ceramics [58]. This is crack growth enhanced by some chemical reaction, which substantially weakens the atomic bond at the crack tip even at stresses well below intrinsic strength.

The SCG behavior of ceramics and glass is generally expressed by a relationship between crack velocity, v, and an applied stress intensity factor, K, to show the so-called v−K diagram. These v−K diagrams of ceramics and glasses typically contain stages I, II, and III with threshold stress intensity below which SCG does not occur (Kth), as shown in Figure 4. Crack velocity at each stage of the diagram is governed by a different mechanism; and the diagram varies substantially depending on materials and environments. Some cases show no threshold stress intensity while in other cases stages II and III are negligible. Once the v−K diagram is established for a particular material and condition, fracture mechanics can predict its lifetime. Since stage
I crack propagation occupies a large part of total lifetime, this stage is essential for critical lifetime prediction.

A typical example of $v-K$ diagram exists in the classic work by Wiederhorn [59]. They measured crack growth of soda lime silica glass in inert nitrogen environments with different humidities and obtained $v-K$ diagrams which all consist of stages I, II, and III. It is generally considered that stages I, II, and III correspond to impurity controlled activity, impurity diffusion limited, and bulk environment controlled regions, respectively. In the above crack growth study, Wiederhorn considered that moisture activity controls crack velocity at stage I, while stage II is limited by moisture diffusivity. Thus, moisture concentration is decisive and insensitive to $K_i$, leading to the so-called plateau region. Here, $K_i$ is the applied stress intensity factor in mode I.

Great effort has been made to understand fundamental mechanisms determining the relationship between SCG and stress intensity factor. Although a number of mathematical expressions to describe this relationship have been proposed, the following forms are used presently:

$$v = v_0 (K_i/K_{IC})^N$$  \hspace{1cm} (3)

$$v = v_0 \exp(K_i/K_{IC})^B$$  \hspace{1cm} (4)

Here, $v_0$, $N$, and $B$ are constant, and $K_{IC}$ is the critical applied stress intensity factor, or fracture toughness, of the material. Neither expression has obtained general acceptance as a “fundamental” form. Eqn (3) is most commonly employed in component design due to its ease in mathematical manipulation; log ($v$) is linear in log ($K_i$), with $N$ slope. Parameter $N$ is recognized as an indicator of susceptibility to SCG; the larger the $N$ value, the smaller the SCG. However, it is frequently observed that obtained $N$ values do not agree between two different measurement techniques or two different specimen sizes for identical materials and environments [60]. This indicates that the power law form does not always adequately express crack velocity at small stress intensity factors. This shortcoming significantly limits potential of extrapolating the smaller velocity range.

Several techniques have been proposed for measuring a $v-K$ diagram. Many researchers have used direct measurement of propagating cracks in a fracture mechanics test specimen. [59—61]. An advantage of this method is that a $v-K$ diagram can be drawn easily from one specimen or one test. The fracture mechanics test specimen usually has a long artificial crack for which the $K$ value can be determined easily. However, such long artificial cracks do not always behave similarly to short natural cracks. Ceramic components are usually inspected through nondestructive tests before service so that they do not contain intolerant long cracks. Thus, it is short cracks that govern the lifetime of ceramic components in service. Another way to establish a $v-K$ diagram is to estimate it indirectly from strength measurement results of static or dynamic fatigue tests [62—68]. Static fatigue tests involve static loading of tensile stresses lower than the fast fracture strength until failure occurs: this shows the relationship between time to failure and stress. In dynamic fatigue tests, tensile stress is monotonically increased at various loading rates until there is a failure to obtain a relationship between loading rate and strength. Both obtained relations lead to a $v-K$ relationship using fracture mechanics. One advantage of these methods is that observed crack propagation behavior can be known from short flaws. This has practical importance, as mentioned above. Short cracks can occur in ceramic components as pores, inclusions, and machining surface flaws.

When $K_{IC}$ depends on crack length $a$ (or $R$-curve behavior), SCG behavior differences between short and long cracks become clear. As shown schematically in Figure 5, short crack extension starts at considerably lower stress intensities than long cracks for materials with $R$-curves; and plateau fracture toughness is not attained before catastrophic failure [67,68]. Unlike tests where long cracks are forced to initiate from a notch tip in a controlled fashion, short cracks originate from local natural flaws that are frequently located at the surface. The short crack with the shallowest $R$-curve slope becomes unstable first (the fracture stress at this instability corresponds to flexure stress in this instance). This indicates that under equivalent loading conditions, specimens or components cannot take advantage of the full $R$-curve potential. In addition, crack resistance measurements generated from long cracks might
therefore overestimate ceramic component toughness in real, load-bearing applications.

The relationship between material microstructure and its properties has been addressed; also, the $v-K$ relationship is often characterized by information on interactions between the moving crack and the microstructure [69,70]. Importance of fractographic analysis on specimens after the test has been recognized with strength measuring techniques because this can show whether cracks propagate trans- or intergranularly, thereby clarifying effects of grain boundary phases on microstructural crack propagation. For specimens tested at high temperatures, fractography can determine whether strength degradation arises from SCG or other mechanisms such as creep [71,72]. Furthermore, degree of moving crack velocity can be estimated in some cases from fractography, as cracks grow trans- or intergranularly depending on their velocity [66]. Finally, direct observation of indentation cracks has been proposed recently for investigating both the $v-K$ relationships and fundamental aspects of flaw shape evolution [73,74].

This section describes fatigue and slow crack growth tests including: (1) static and dynamic fatigue for short (or natural) cracks, (2) cyclic fatigue for long cracks, (3) static and dynamic fatigue for long cracks, and (4) cyclic fatigue for long cracks, followed by the related standards. In addition, we will discuss “Cyclic Fatigue of Toughened Ceramics” for a related research topic.

3.1. Testing Methods

3.1.1. Static and Dynamic Fatigue Tests Using Short Cracks

Fatigue tests for specimens containing short natural flaws can provide data closest to the real service application. The static test involves applying a mode I, plane strain static stress in an appropriate environment and measuring the time to failure. Conducting measurements at different applied stresses, the $v-K$ curve can be established even though natural flaw sizes are unknown. The methods are essentially identical to simple strength measurement methods. Principally, the fatigue parameter $N$ can be measured without any assumptions about flaw size or shape. However, this method is very time consuming to obtain even one data point and a number of specimens are needed for each condition to account satisfactorily for statistical variability in strength. Dynamic fatigue generates principally the same results, but testing time is shortened through testing at several constant stress rates. Any specimen configuration which is generally used for any strength test can be used for both fatigue tests. Examples of successful measurements on static and dynamic fatigue of ceramics are available for short artificial cracks [71–77] and natural cracks [78–83].

3.1.2. Cyclic Fatigue Tests Using Short Cracks

Structural ceramic components in service very often are subjected to cyclic mechanical loading. Cyclic fatigue performance of structural ceramics determines the life of components in many cases, including gas turbine rotors or blades and diesel engine valves. Therefore, characterizing and properly understanding ceramic cyclic fatigue is crucial for confident use of components subjected to cyclic loading. Many reports explain cyclic fatigue of ceramics using short or natural flaws [84–94].

Several mechanisms for cyclic fatigue of ceramics have been proposed. Most of them are related strongly to toughening mechanisms of ceramics at room temperature. They include degradation of transformation toughening zone at the crack tip (such as in some zirconia); damage to bridging reinforcements (grains, whiskers, platelets, and brittle and/or ductile particles) over the bridged interface; cyclic loading-induced accumulation of microcracks around the crack tip; crack tip blunting/resharpening; and relaxation of any mechanically beneficial residual stress due to microcracking, as shown in Figure 6. It should be noted that these degradation mechanisms are not active in static or dynamic fatigue.

In some cases, at elevated temperatures where softening of the grain boundary glassy phase can occur, other mechanisms become active. These include oxidation-induced deformation and viscoelastic effects [84,85].

Lead zirconate titanate (PZT) ceramics have potential for increased use in advanced industrial applications such as piezoelectric actuators and ultrasonic vibrators because of their quick response, compactness, and good power efficiency. When they are subjected to a powerful alternating electric field, cyclic stresses are introduced into these piezoelectric ceramics which can lead to fracture.
and/or microcracking. Moreover, degradation of piezoelectric properties is a serious practical concern, as this is a major cause of reduced device lifetime. Therefore, in recent years, many studies have been devoted to clarifying cyclic fatigue behavior and mechanical reliability of this kind of material [95–110].

Ceramic materials have been also extensively applied in prosthodontics in recent years, since all-ceramic crowns are superior to traditional porcelain-fused-to-metal crowns in esthetics, wear resistance, and chemical inertness. In such applications, the materials are susceptible to cyclic fatigue that substantially degrades their mechanical reliability and structural durability. Therefore, considerable efforts have been made to investigate cyclic fatigue behavior of various ceramics that can be used for these applications [55,111–115].

3.1.3. Static and Dynamic Fatigue Tests Using Long Cracks

Crack propagation testing using long cracks (i.e., fracture mechanics specimens) is fundamentally different from crack propagation testing with short or natural flaws (i.e., strength measurement specimens) for several reasons. Crack length being propagated may be on the order of millimeters, while crack lengths dealt with in short fatigue crack propagation were on the order of micrometers or tens or hundreds of micrometers. “Long” cracks are not easily influenced by microstructural residual stresses while “short” cracks are.

By using fracture mechanics specimens (i.e., specimens which are often used for measuring fracture toughness), one can measure directly the crack velocity \( v \) as a function of applied stress intensity \( K \) [59–61]. However, test specimens and the experimental facility are more complicated; it is probably difficult to operate at high temperature or in an extremely corrosive environment. Specimens may be loaded in load control or in displacement control and relaxation techniques may also be used. Figure 7 shows examples of frequently used long-crack test geometries. Test specimens must have a relatively large sharp crack which will propagate during testing. Ideally, the crack tip should be atomically sharp. However, a notch produced by a diamond saw has been used frequently, assuming the sharp crack is produced from tip of the notch.

Several studies have been conducted to investigate slow-crack growth behavior and mechanical properties of inert bioceramics, such as alumina- and yttria-stabilized zirconia, which are used as components for orthopedic and dental applications [116–119].

3.1.4. Cyclic Fatigue Tests Using Long Cracks

Also, there is substantial difference in cyclic loading between crack propagation with long cracks and that with

![Schematic classification of primary microstructural toughening mechanisms, which are strongly involved with property degradation during cyclic loading.](image)

![Frequently used long-crack test geometries: (a) compact tension (CT); (b) single-edge notched beam (SENB); (c) chevron notch (CN); (d) double torsion (DT); (e) double cantilever beam (DCB). Grayed area is crack.](image)
short or natural flaws. The so-called S-N (stress-number of cycles) curves represent results of fatigue crack propagation testing of short cracks, while crack propagation per cycle as a function of stress intensity range or maximum stress intensity represents results of crack propagation testing with long cracks. However, toughening mechanisms, which are related to enhanced crack propagation in cyclic fatigue with short cracks, can be similarly related to that in cyclic fatigue with long cracks. There are examples of successful measurements on cyclic fatigue of ceramics with long cracks [120–129].

Similar to the \(v-K\) relationships described above, crack propagation rates or crack growth extension per unit cycle \((da/dN)\) for cyclic fatigue testing with long cracks are also typically represented by a power-law relationship. The commonly known Paris law [130] relates \(da/dN\) to the stress intensity range \((\Delta K = K_{\text{Imax}} - K_{\text{Imin}}\) with \(K_{\text{Imax}}\) and \(K_{\text{Imin}}\) being maximum and minimum applied stress intensities, respectively) as

\[
da/dN = C(\Delta K)^m\tag{5}
\]

where \(C\) is a constant and \(m\) is a cyclic fatigue exponent. Crack propagation rate \((da/dt)\) or \(v\) is related to \(da/dN\) as

\[
da/dt = v = (da/dN)f\tag{6}
\]

where \(f\) is the cyclic frequency. For determining the cyclic fatigue exponent, \(m\), the crack growth increment per cycle \(da/dN\), or the crack propagation rate \(da/dt\), as a function of stress intensity range, \(\Delta K\). While typical \(m\) values of most metals range between 2 and 4, those of ceramics are as high as 50 or larger in mid-growth rate regime [120,122].

**Figure 8** shows comparison of growth rate data for some structural ceramics and metal alloys.

### 3.2. Related Standards

Applicable general standards for fatigue, slow crack growth, and fracture toughness tests of advanced ceramics are as follows (some additional standards for specific uses not included). It should be noted that most standards for simple strength measurements can be applied to fatigue tests with natural cracks, by changing loading rates, modes, etc.

- ISO (International Organization for Standardization), TC206 “Fine ceramics (advanced ceramics, advanced technical ceramics)”
  - ISO 15732: Fine ceramics (advanced ceramics, advanced technical ceramics) — Test method for fracture toughness of monolithic ceramics at room temperature by single-edge precracked beam (SEPB) method
  - ISO 18756: Fine ceramics (advanced ceramics, advanced technical ceramics) — Determination of fracture toughness of monolithic ceramics at room temperature by the surface crack in flexure (SCF) method
  - ISO 22214: Fine ceramics (advanced ceramics, advanced technical ceramics) — Test method for cyclic bending fatigue of monolithic ceramics at room temperature
  - ISO 23146: Fine ceramics (advanced ceramics, advanced technical ceramics) — Test methods for

![Figure 8](image_url)
fracture toughness of monolithic ceramics — Single-edge V-notch beam (SEVNB) method
- ISO 24370: Fine ceramics (advanced ceramics, advanced technical ceramics) — Test method for fracture toughness of monolithic ceramics at room temperature by chevron-notched beam (CNB) method
- ASTM International (formerly known as the American Society for Testing and Materials), C28 “Advanced Ceramics”
  - C1360: Standard Practice for Constant Amplitude, Axial, Tension—Tension Cyclic Fatigue of Continuous Fiber-Reinforced Advanced Ceramics at Ambient Temperatures
  - C1361: Standard Practice for Constant Amplitude, Axial, Tension—Tension Cyclic Fatigue of Advanced Ceramics at Ambient Temperatures
  - C1368: Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics by Constant Stress Rate Strength Testing at Ambient Temperature
  - C1421: Standard Test Methods for Determination of Fracture Toughness of Advanced Ceramics at Ambient Temperature
  - C1465: Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics by Constant Stress Rate Flexural Testing at Elevated Temperatures
  - C1576: Standard Test Method for Determination of Slow Crack Growth Parameters of Advanced Ceramics by Constant Stress Flexural Testing (Stress Rupture) at Ambient Temperature
- CEN (European Committee for Standardization), TC184 “Advanced Technical Ceramics”
  - EN 843-3: Advanced technical ceramics — Mechanical properties of monolithic ceramics at room temperature — Part 3: Determination of subcritical crack growth parameters from constant stressing rate flexural strength tests
  - EN 13234: Advanced technical ceramics — Mechanical properties of ceramic composites at ambient temperature — Determination of fatigue properties at constant amplitude
  - EN 15156: Advanced technical ceramics — Mechanical properties of ceramic composites at room temperature — Determination of fatigue properties at constant amplitude
  - EN 15157: Advanced technical ceramics — Mechanical properties of ceramic composites at high temperature under inert atmosphere — Determination of fatigue properties at constant amplitude
  - EN ISO 15732: Fine ceramics (advanced ceramics, advanced technical ceramics) — Test method for fracture toughness of monolithic ceramics at room temperature by single-edge precracked beam (SEPB) method
  - EN ISO 18756: Fine ceramics (advanced ceramics, advanced technical ceramics) — Determination of fracture toughness of monolithic ceramics at room temperature by the surface crack in flexure (SCF) method
- JISC (Japanese Industrial Standards Committee), Divisional Council on Ceramics
  - JIS R 1607: Testing methods for fracture toughness of fine ceramics at room temperature
  - JIS R 1617: Testing method for fracture toughness of fine ceramics at elevated temperature
  - JIS R 1621: Testing method for bending fatigue of fine ceramics at room temperature
  - JIS R 1632: Test methods for static bending fatigue of fine ceramics
  - JIS R 1658: Testing method for bending fatigue of fine ceramics at elevated temperature
  - JIS R 1668: Testing method for fracture toughness of porous fine ceramics
  - JIS R 1677: Testing method for bending fatigue of porous fine ceramics at room temperature
  - TS R 0002: Testing method for crack growth resistance (R-curve) of fine ceramics

3.3. Research Topic: Cyclic Fatigue of Toughened Ceramics

Most ceramics subcritical crack growth rates under cyclic loads exceed corresponding growth rates under sustained, quasi-static loads by many orders of magnitude at equivalent stress intensity levels. This contrast clearly indicates that mechanical properties of ceramics are degraded by
cyclic fatigue. A number of tough ceramics show crack shielding effects which substantially influence cyclic fatigue behaviors. Mechanisms for cyclic fatigue crack growth in ceramic materials are generally based on suppression of crack-tip shielding in the crack wake under cyclic loading.

Gilbert et al. [129,131] characterized cyclic fatigue crack propagation behavior of “in situ toughened” monolithic silicon nitride, in light of fatigue crack growth models for ceramics toughened by grain-bridging mechanisms. They used a model based on diminished crack-tip shielding in the crack wake under cyclic loads due to frictional wear degradation of the grain-bridging zone [120], as shown in Figure 9. The notion of cyclic crack growth promoted by diminished shielding was consistent with measured (long crack) growth rates, fractography, in situ crack profile analyses, and measurements of back face strain compliance. Growth rates depended on maximum applied stress intensity, $K_{\text{max}}$, much more than on the applied stress intensity range, $\Delta K$. Fatigue thresholds exhibited a marked dependence on the load ratio, $R = K_{\text{min}}/K_{\text{max}}$. These effects were inconsistent with traditional models of fatigue crack closure. In particular, when characterized in terms of $K_{\text{max}}$, growth rates below about $10^{-9}$ m/cycle exhibit an inverse dependence on load ratio. This observation was consistent with the grain-bridging phenomenon. Specifically, with increasing $R$, sliding distance between the grain bridges decreased, leading to less frictional wear and less degradation in shielding per loading cycle.

The in situ toughening has been adapted also into silicon carbide recently. Similar crack growth resistance curves (R-curve) and cyclic fatigue behaviors have been found in such silicon carbide materials [132–137]. Gilbert et al. [132] investigated R-curve behavior in two silicon carbide (SiC) ceramics with sharply contrasting fracture properties, using small (<400 µm) surface cracks and long (>3 mm) through-thickness cracks. The first, an in situ toughened material designated ABC-SiC, fails by intergranular fracture, whereas the second, a commercial SiC (Hexoloy SA), fails by transgranular cleavage. The former material, hot-pressed with aluminum, boron, and carbon additives, contained a network of plate-shaped grains [133]. An amorphous grain boundary film of ~1-nm thickness promoted debonding and crack deflection. Resultant grain bridging generated R-curve toughening. In contrast, no evidence of crack-tip shielding is observed in the latter SiC, Hexoloy SA. Although the Hexoloy SA fails catastrophically at <3 MPa.m$^{1/2}$, ABC-SiC exhibits much improved

![FIGURE 9](image)

**FIGURE 9** Schematic illustration of the model for cyclic fatigue crack propagation involving the cyclic loading induced degradation in crack-tip shielding by (a) grain bridging in the crack wake. The crack-opening displacement, $2u(x)$, is indicated at position $x$ in the bridging zone. The grain-bridging stress, $p(u)$, is shown rapidly rising from the initial residual compressive stress, $\sigma_R$, as the grain debonds, followed by frictional sliding which gradually decreases as the grain is extracted from the matrix under (b) monotonic and (c) cyclic loading conditions [120]. (Reprinted with permission of Elsevier. All rights reserved)
flaw tolerance with significant rising $R$-curve behavior and a steady-state fracture toughness of ~9 MPa.m$^{1/2}$ after crack extension of ~600 µm. Cyclic fatigue properties of such in situ toughened silicon carbide were also examined using similar long (>3 mm) and small (<200 µm) cracks [134]. Crack extension was monitored optically under cyclic load. Small cracks showed high growth rates with negative dependence on the far-field driving force. This behavior was attributable to both indent-induced residual stresses and relative size of cracks compared with bridging zone lengths.

At elevated temperatures also, bridging traction developed behind a crack tip influence crack propagation substantially under cyclic loading conditions in tough monolithic ceramics. Assuming a temperature range where grain boundary phases were soft enough that bridging can occur due to a viscous layer in the boundary, McNaney et al. [138] developed a viscoelastic model to incorporate: microstructural details (grain shape, grain boundary thickness, and glass viscosity) along with effects of external variables (loading rate and temperature). In the model, bridging forces associated with shear resistance of the grain boundary phase were transmitted across the crack surfaces. This phenomenon is applicable to a wide range of ceramic materials at elevated temperatures. Chen et al. [139] examined fatigue crack growth at elevated temperatures of 25–1300°C under cyclic loading for in situ toughened, monolithic silicon carbide, along with roles of temperature, load ratio, cyclic frequency, and loading mode (static and cyclic). It was found that at equivalent stress intensity levels, crack velocities under cyclic loads were significantly faster than under static loads. Fatigue thresholds were decreased with increasing temperature up to 1200°C. At 1200°C, they detected no effect of frequency in the range of 3–1000 Hz. Indeed, fractography and crack path sectioning revealed a fracture mode at 1200–1300°C that was essentially identical to that at room temperature, i.e., predominant intergranular cracking with evidence of grain bridging in the crack wake. This excellent crack growth resistance was attributed to a process of grain boundary microstructural evolution at elevated temperatures, specifically involving crystallization of the amorphous grain boundary films/phases.

Excellent cyclic fatigue resistance, high strength, and high fracture toughness in the tough silicon carbide were not compromised significantly at elevated temperatures up to 1500°C [140]. Particularly, fatigue crack growth properties up to 1300°C were essentially identical to those at room temperature. Mechanistically, the cyclic fatigue crack advance mechanism was similar up to 1300°C, involving mutual competition between intergranular cracking ahead of the crack tip and interlocking grain bridging in the crack wake. Moreover, good creep resistance was observed below 1400°C, with little evidence of cavitation. Such unusually good high-temperature properties of silicon carbide were attributed to crystallization of the grain boundary amorphous phase which can occur either in situ, due to prolonged thermal exposure associated with high-temperature fatigue and creep tests, or by prior heat treatment.

As stated above, cyclic fatigue growth rates in tough ceramics with rising $R$-curve behavior depend very strongly on the maximum applied stress intensity, $K_{\text{max}}$, and weakly on the stress intensity range, $\Delta K$. Jacobs and Chen [141] identified mechanical equilibrium between shielding accumulation by crack growth. They further identified shielding degradation by frictional wear of sliding interfaces for steady-state cyclic fatigue in a gas pressure sintered silicon nitride. This equilibrium gave rise to a rate law for cyclic fatigue. They found that accumulation and degradation processes were the origin of strong $K_{\text{max}}$ dependence and weak $\Delta K$ dependence, respectively.

Chen and Engineer [142] also studied a model for fatigue crack propagation based on sliding wear of bridging grains of polycrystalline ceramics such as tough silicon nitride. Taking into account damage development and crack-tip energy balance, they developed equilibrium and compatibility equations in the crack wake under monotonic and cyclic loading/unloading conditions. Fatigue was controlled by frictional resistance to reverse sliding, suggesting a mechanics similar to an elastic–plastic one of a path-dependent hardening material. The unloading crack-opening displacement is more strongly dependent on $K_{\text{max}}$ than on $\Delta K$; such displacement caused wear on bridging grains. Incremental crack growth brought in new bridging grains with a shielding effect on the crack-tip stress field. Such an effect strongly depended on $K_{\text{max}}$, but was independent of $\Delta K$. In the steady state, with balanced shielding accumulation and shielding degradation, the fatigue crack growth rate had the form $da/dN = A(K_{\text{max}})^b(\Delta K)^c$, where $A$, $b$, and $c$ are material-dependent parameters. Fatigue was predicted to have a very high $b$, modest $c$, a higher fatigue resistance for tougher ceramics, and a stronger $K_{\text{max}}$ dependence for less tough ceramics. These predictions agreed with experimental observations.

For grain-bridging ceramics, the fatigue threshold substantially depends on crack size during bridging zone development. Based on experimental studies for two different grain-bridging ceramics (Al$_2$O$_3$ and Si$_3$N$_4$), Kruzic et al. [143] demonstrated that the effect of crack size on the threshold for fatigue crack growth may be expressed in terms of a fatigue threshold $R$-curve, which is analogous to the fracture toughness $R$-curve and may be used for predictions of the endurance strength under cyclic loading conditions. They succeeded in expressing both the predicted fatigue threshold $R$-curve and the fracture toughness $R$-curve for the same material in one simple plot. Figure 10 shows an
example of such a plot (Al2O3). Furthermore, it has been indicated that, using measured bridging stress profiles, the fatigue threshold $R$-curve for short cracks may be predicted based solely on results from many millimeter long, through-thickness, fatigue cracks. Direct comparisons of predictions to experimentally measured fatigue crack growth thresholds in Al2O3 of coarse grains over a range of crack sizes demonstrated the accuracy of those methods.

4. CREEP AND CREEP RUPTURE

Advanced ceramics have great potential for use as high-temperature materials in devices since their heat resistance is far superior to that of currently available single crystal super alloys. When ceramic materials are used for such applications, the delayed fracture properties should be assessed thoroughly to establish reliability. High-temperature durability of materials is very often limited by stress-enhanced cumulative damage, which is very strongly related to time-dependent deformation, or creep [143–232].

Creep is a phenomenon of solid materials to slowly deform permanently when the materials are subjected to loading for long-term, and becomes more severe at higher temperatures. The rate of the deformation, creep strain rate, is a function of the material properties, exposure time, exposure temperature, and the applied structural load. Creep deformation is typically expressed by creep curve at a specific temperature, where the vertical axis is deformation and the horizontal one is time, as shown in Figure 11. It typically consists of three creep stages: primary creep (transient creep), secondary creep (steady-state creep), and tertiary creep (accelerated creep). The “creep strain rate” generally refers to the rate in the secondary creep, and creep mechanism is most often discussed in terms of stress dependence of this creep strain rate. In creep of metals, primary creep is due to work hardening, and tertiary one due to necking phenomena. Since such phenomena hardly occur in ceramics, creep of ceramics often lacks these two stages. The creep strain rate is generally expressed by the following equation:

$$\frac{de}{dt} = CD\sigma^n/d^b/kT$$

where $e$ is the creep strain, $C$ is a constant, $D$ is the diffusion co-efficient, $\sigma$ is the applied stress, $d$ is the grain size of the material, $k$ is Boltzmann’s constant, $T$ is the absolute temperature, and $n$ and $b$ are exponents dependent on the creep mechanism. In actual creep tests, steady state is sometimes not observed, and the minimum creep strain rate is used for it in such a case.

Test methods required for evaluating creep behavior are classified into three areas: flexure, tension, and compression. While all three methods involve uniaxial stress states, only flexure has a nonuniform stress state. This nonuniformity complicates creep behavior evaluation, particularly when the ceramic deforms differently in compression than in tension. Several methods have been proposed to extract uniformly stressed creep behavior from flexural tests. Still, to add sufficient reliability, such methods require at least one test in either tension or compression in addition to the flexure test. Thus, flexure tests are simple, and inexpensive; however, they should accompany complicated analysis.
As ceramic components used in advanced engineering application are complicated, it is essential to collect reliable data on creep behavior to predict component durability. To do this, creep tests conducted in uniaxial tension and compression are superior to flexure tests. Particularly, tensile creep tests have become routine during the past two decades. Principal tensile test advantages include precise determination of applied stress and strain in the uniform stress field. This section first describes these three creep test methods and the related standards, and reviews a relevant research topic, “Creep of Silicon Nitride”.

4.1. Testing Methods

4.1.1. Tensile Creep Test

There have been several testing methods for investigating ceramic creep behaviors including flexural tests, tensile tests, and compressive tests. Among them, the tensile test gives the most advantages for obtaining creep strain curves of materials because deformation factors such as dislocation, diffusion, and cavitation occur uniformly and throughout a gauge length area of a specimen. Therefore, correspondence between time and deformation can be properly and easily obtained. Nevertheless, tensile tests are not so widely employed in ceramic creep measurements compared to the flexural test. This is because a stress concentration produced during the test cannot be attenuated by local plastic deformation. Hence, even slight eccentric loading causes specimen fracture around the grips and a large parasitic bending stress. Especially when there is some limit in the ceramic billet size, the specimen is required to be small and then gripped in a hot zone of a furnace (hot grip). Successful testing methods for tensile creep of ceramics are available in Refs. [173–205]. Specimen design and testing assembly used for long-term (~10,000 h) tensile creep tests [178] are shown in Figure 11. In this testing method, the specimen has a cylindrical gauge length with a diameter of 2.8 mm and a length of 25 mm and two target arms at the end of the gauge section. The elongation of the distance between them is measured by an optical extensometer.

The normal longitudinal stress in the gauge section is given simply by the applied load divided by the cross-sectional area of the gauge section. The normal longitudinal strain, \( \varepsilon \), is calculated from the elongation of the gauge section, \( \delta \):

\[
\varepsilon = \frac{\delta}{l_0}
\]

where \( l_0 \) is the initial length of the gauge section.

4.1.2. Flexural Creep Test

Flexural creep tests have the advantage of low specimen preparation cost, easy experimental design, and minimum specimen material and preparation. Therefore, many creep studies have employed flexure tests so far [206–215]. There are, however, critical problems in creep testing ceramics in flexure. They include nonuniform stress distribution over the specimen cross section, chemically and mechanically predominant surface phenomena, indirect determination of stress and strain, and small volume (surface) under maximum stress. Particularly, assumption of linear elastic behavior in flexural bars very often results in erroneous measurement results.

Similar to the fracture strength measurements, there are two typical flexural testing methods: three- and four-point bending. Because of a uniformly applied moment in the inner load span and larger volume (surface) under the maximum stress, four-point bending is generally employed for flexural creep tests of ceramics. The maximum tensile stress in four-point bending with a rectangular specimen is given by an equation similar to Eqn (1), and the maximum creep strain, \( \varepsilon_{\text{max}} \), is calculated from the deflection of the specimen at the center of the inner load span, \( \delta_c \):

\[
\varepsilon_{\text{max}} = 4h\delta_c/L_i^2
\]

where \( L_i \) is the inner load spans, and \( h \) is the height of the specimen.

4.1.3. Compressive Creep Test

Compressive creep testing, shown in Figure 12, is an old and very commonly used method for studying ceramic material creep deformation. In principle, uniform stress distribution is obtained in compressive testing, so material behavior can be studied in a pure stress state. For a variety of reasons explained later, creep in compression can differ substantially from creep in tension. Therefore, to completely understand material creep behavior, it is important to characterize creep responses to both tension and compression. Examples of compressive creep tests of ceramics are available in Refs. [18,216–232]. Compression creep using a C-ring specimen, which is particularly useful for investigating creep behavior of high-temperature ceramic tubes, has been also reported [232].
Compression test advantages include low cost of specimen preparation (compared to tensile ones), small size of specimens (allowing use of limited small batches), and uniform stress distribution (precise determination of stress–strain relationship). However, specimen alignment is very often more difficult than with flexure or tension. In addition, cavitation and microcracks effects, which greatly impact creep lifetime, cannot be evaluated in compression.

4.2. Related Standards
Applicable general standards for creep tests of advanced ceramics are as follows.

ISO (International Organization for Standardization), TC206 “Fine ceramics (advanced ceramics, advanced technical ceramics)”
- ISO 22215: Fine ceramics (advanced ceramics, advanced technical ceramics) – Test method for tensile creep of monolithic ceramics
- ASTM International (formerly known as the American Society for Testing and Materials), C28 “Advanced Ceramics”
  - C1291: Standard Test Method for Elevated Temperature Tensile Creep Strain, Creep Strain Rate, and Creep Time to Failure for Advanced Monolithic Ceramics
- CEN (European Committee for Standardization), TC184 “Advanced Technical Ceramics”
  - EN 820-4 Advanced technical ceramics – Thermo-mechanical properties of monolithic ceramics – Part 4: Determination of flexural creep deformation at elevated temperatures
  - EN 13235: Advanced technical ceramics – Mechanical properties of ceramic composites at high temperature under inert atmosphere – Determination of creep behavior
  - EN 15365: Advanced technical ceramics – Mechanical properties of ceramic fibers at high temperature in a nonreactive environment – Determination of creep behavior by the hot end method
  - CEN/TS 15658: Advanced technical ceramics – Mechanical properties of ceramic fibers at high temperature under nonreactive environment – Determination of creep behavior by the hot end method
- JISC (Japanese Industrial Standards Committee), Divisional Council on Ceramics
  - JIS R 1612: Testing method for bending creep of fine ceramics
  - JIS R 1631: Test method for tensile creep of fine ceramics

4.3. Research Topic: Creep of Silicon Nitride
Silicon nitride, one of the most attractive candidate materials for high-temperature components, generally exhibits creep deformation at high temperatures above 1300°C. It is known that dislocation motion contributes very little to silicon nitride creep below 1700°C [233]. Also, diffusional creep involving transfer of vacancy from interfaces under tension to those under compression is unlikely to be a rate-controlling process in silicon nitride containing considerable amounts of glassy phase at the interfaces. The dominant creep mechanism for Si3N4 containing a grain boundary phase is grain boundary sliding (GBS). This GBS occurs by viscous flow of the grain boundary phase (which acts as a lubricant between Si3N4 grains) or by solution–reprecipitation of Si and N (which occur by diffusion through the grain boundary phase). In addition, cavitation at either two-grain or multiple-grain junctions is generated within Si3N4 containing a grain boundary phase; void nucleation and growth are controlled by the same mechanism that determines deformation. Thus, creep behavior of silicon nitride is largely dependent on characteristics and behavior of the glassy phase. These might include composition (viscosity), state of dispersion, dimensions, devitrification, and outward diffusion of the cation enhanced by oxidation. Glassy phase characteristics are affected by even a small addition of impurity [234,235].

There have been systematic studies of the effect of small additions of fluorine or chlorine impurity on creep behavior of silicon nitrides.

Several studies on tensile creep of silicon nitride suggest that cavitation contributes substantially to the creep strain [192–194]. Luecke et al. [192] reported that, during tensile creep of an HIP-ed silicon nitride, the volume fraction of cavities increases linearly with strain (Figure 13); these cavities produce nearly all of the measured strain. Transmission electron microscopy (TEM) revealed cavities formed during creep in silicate glass pockets of multigrain junctions. They investigated cavity size distribution using small-angle X-ray scattering technique and found that cavity addition, rather than cavity growth, dominates the cavitation process. It was suggested that during creep, while silicon nitride grains remain rigid, cavitation in multigrain junctions allows the silicate to flow from cavities to surrounding silicate pockets, leading to microstructure dilatation and material deformation. Grain boundary sliding accommodates this expansion and results in specimen elongation. Luecke et al. [236] also proposed
a model to accurately describe tensile creep of silicon nitride, which is controlled by rate of formation and growth of cavities in the second phase.

Uniaxial creep tests, such as tension and compression, are advantageous since the specimen is subject to uniform stress such that obtained results can be analyzed simply and precisely (if tests are conducted ideally). One principal difference between tensile and compressive creep tests is the degree of cavity formation; tensile creep is much more affected by cavitation. Luecke et al. [237] have investigated silicon nitride creep in both tension and compression at 1430°C at stresses between 40 and 300 MPa. The material crept up to 100 times faster in tension than in compression. Stress dependence of the tensile strain rate was not characterized by a single stress exponent rising from about two at low stresses to about five at higher stresses. While cavitation contributed significantly to the tensile strain, the material exhibited a stress exponent of unity, and cavitation was almost completely suppressed in compression. Asymmetry in creep behavior between tension and compression is due to dilatation of the material. Yoon et al. [238] compared tensile and compressive creep behavior of silicon nitride at 1350°C and 1400°C in more detail. Again, over the entire investigated stress range, the tension creep rate increased linearly with stress at low stresses and exponentially at high stresses, but the compression creep rate increased linearly with stress. While the cavity volume fraction increased linearly with increased tensile creep strain, the volume fraction of cavities in compression was ~18% of that in tension at 1.8% axial strain and approached zero at strains <1%. In compression, where cavitation is suppressed, deformation occurs by grain boundary sliding mechanism accommodated by solution–precipitation [225–228].

In these decades, efforts to strengthen the grain boundary phase for improving heat resistance of silicon nitride have made significant progress [195–200]. For example, some grades of silicon nitrides, SN 281 and 282 (Kyocera Corp. Kyoto, Japan) [239–241], showed excellent creep resistance even at temperatures above 1400°C [195–197], because of highly refractory phase of grain boundaries. There are two regions in a delayed fracture mechanism map of silicon nitride at high temperatures: slow crack growth failure and creep damage rupture. The former is a fracture occurring when a crack grows subcritically from a preexisting flaw and reaches critical size. This is predominant in the high-stress, short-term life region. The latter is due to formation of a macrocrack with the critical size by cavity nucleation and coalescence, which prevails in the low-stress, long-term life region. Creep life for structural ceramics is conventionally expressed as

$$t_f = C_L \sigma^{-N}$$  \hspace{1cm} (10)

where $t_f$ is the time to failure, $\sigma$ is the applied stress, and $C_L$ and $N$ are constants. Exponent $N$ determines stress dependency of creep life and it is often referred to as a fatigue exponent. In Figure 14, creep lives of SN 282 are shown as functions of stress and temperature in comparison with those of conventional silicon nitrides, SN YA (hot-pressed silicon nitride doped with 5 wt% yttria and 3 wt% alumina) [195]. Plots indicate some changes in slopes around 200–250 MPa. Fractography revealed that fracture occurred from slow crack growth in the high stress range for SN 282, as well as the SN YA; the estimated fatigue exponent was ten or higher. In marked contrast, creep damage was obvious for the SN YA, and the fatigue exponent was two to three in the low-stress range. Occurrence of creep damage rupture substantially limits high-temperature structural reliability of silicon nitride. In the SN 282, however, creep rupture hardly occurred at stresses below 200 MPa at tested temperatures. While creep curves
of conventional silicon nitrides consist of three regimes, transient, steady state, and accelerated creep regimes, these materials showed solely transient creep \[195−197\]. Although high-resolution TEM study revealed that a glassy phase existed at some grain boundaries, thickness was very narrow—typically below 1 nm (Figure 15). Glassy phase was also identified at some boundaries between silicon nitride matrix and secondary phase. In XRD analysis, there were strong peaks of Lu$_2$Si$_2$O$_7$ aside from α and β silicon nitrides (or Lu$_4$Si$_2$N$_2$O$_7$) \[195−197\]. Excellent creep resistance was attributed to highly refractory phase of grain boundaries and triple junctions such as Lu$_2$Si$_2$O$_7$.

REFERENCES


Chapter | 9.2 Testing and Evaluation of Mechanical Properties


Chapter 9.3

Microstructural Control and Mechanical Properties

Tatsuki Ohji
National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidani, Moriyama, Nagoya, Aichi 463-8560, Japan

Chapter Outline

1. Introduction 657
2. Grain Morphology Control 657
3. Fibrous Grain Alignment 659
4. Grain Boundary Phase Control 663
5. Porous Structure Control 666

References 671

1. INTRODUCTION

As the properties of materials are closely related to their microstructures, the realization of a material with certain property can be achieved through controlling them. The microstructures of ceramic materials, however, consist of many types of elements including particles, grains, pores, defects, fibers, layers, and interfaces. These structural elements can be classified by size into four scale levels: (1) atomic-molecular scale; (2) nanoscale; (3) microscale; and (4) macroscale, as shown in Figure 1 [1]. Furthermore, these elements have their own features including nature, morphology, distribution, orientation, and others. Therefore, there are a vast amount of factors to control the structure of materials.

Many properties of materials are affected by the microstructure. Among them, however, the mechanical properties of ceramics have been tremendously improved through well-organized control of the microstructure elements during the past few decades. This chapter describes such research efforts, taking an example of silicon nitride. Silicon nitride is one of the most important engineering ceramics, because of its high strength, high fracture toughness, good heat/corrosion resistance, good wear resistance, etc. Further substantial improvements have been observed in mechanical properties of silicon nitride through the microstructure control during these decades. For example, in the field of ceramics, it has been recognized that a homogeneous and fine-grained microstructure yields materials with high strength but low fracture toughness, while materials with a coarse microstructure give us high toughness but low strength. However, it has become possible to realize both high strength and high fracture toughness in silicon nitride through controlling the size, morphology, and alignment of the grains simultaneously. At the same time, controlling the chemical phases of grain boundaries gives rise to further significant improvements in the fracture toughness, heat resistance, thermal conductivity, etc. Tremendous progress has also been made in understanding atomic behaviors in the grain boundaries. In addition, by controlling the morphology and orientation of pores as well as grains, even porous silicon nitride that is stronger and tougher than dense one has been obtained.

2. GRAIN MORPHOLOGY CONTROL

Generally, high fracture toughness of ceramics can be obtained by enhancing crack-wake toughening mechanisms including grain bridging, frictional translation, rotation of partially separated grains, and frictional grain pullout. These toughening mechanisms operate after a crack propagates from an initial flaw, resulting in a pronounced rising fracture resistance with increasing crack length, namely $R$-curve. Several researchers reported that coarsened microstructure of β-fibrous (or elongated) grains exhibited such a strong R-curve behaviors [2–5]. For example, Li and Yamanis [2] demonstrated that silicon nitrides containing large (>1 μm diameter) fibrous grains exhibit plateau toughness values for long cracks which are close to 10 MPa m$^{1/2}$ while those of submicrometre
grain-sized materials are about 3 MPa m$^{1/2}$. Bridging of
cracks by intact fibrous grains and grain pullout was
observed in the region behind the crack tips, just as with
whisker-reinforced ceramics. The plateau fracture tough-
ness of silicon nitrides increased with the diameter of the
larger fibrous grains, which is consistent with the prediction
for frictional bridging and pullout processes [6,7].

These large fibrous grains are hexagonal prisms of
β-silicon nitride, which are anisotropically grown with the
c-axis growth rates larger than those normal to the prism
faces. Such grain growth can be controlled by the ratio of
α- to β-phase in the starting powders, sintering additives,
and sintering conditions including temperatures, pressure,
and hold time [8–11]. Adding some amount of β-phase
powders into α-phase ones is often used to enhance the
formation of larger fibrous β-grains in a matrix of finer
β-grains.

However, the fracture strength of the self-reinforced
silicon nitrides decreases with the size of the larger fibrous
grains while the plateau (or long crack) toughness increases
[4,12–15]. This phenomenon has been frequently observed
in ceramics materials as already stated. It has been gener-
ally considered that the fracture strength would decrease
with increasing the grain size. This antagonistic relation
between fracture toughness and strength suggests that high
value of the plateau toughness is derived from long cracks
which are much larger than inherent flaw size and is not
necessarily relevant with catastrophic fracture strength. In
fact, $R$-curve measurements for the self-reinforced silicon
nitrides indicate that coarsening overall microstructure of
both the finer matrix grains and the reinforcing grains may
be beneficial to the plateau toughness but only yield
a slowly rising $R$-curve. Thus, more judicious control of
microstructure is essentially required to enhance the overall
$R$-curve response including both the rate of rise with initial
crack extension and the plateau toughness.

The use of fibrous β-seed crystals [15–19] is a more
refined approach to enhance the formation of the large
fibrous grains in fine-grained matrix and to precisely
control the size, morphology, and fraction of these grains.
The seed crystals are usually prepared by heating a powder
mixture of α-Si$_3$N$_4$ and oxides as sintering additives. By
selecting types of Si$_3$N$_4$ starting powders and oxides,
different sizes and shapes of seed crystals can be obtained
[19], as shown in Figure 2. Hirao and co-workers [16] have

![FIGURE 1 Structural elements of ceramics. (Reprinted from Ref. [1] with permission of Elsevier. All rights reserved)](image1)

![FIGURE 2 Fibrous crystals of silicon nitride seeds. (Courtesy of Yoshiaki Inagaki et al.)](image2)
succeeded in realizing compatibly high fracture toughness and high fracture strength, by using seed crystals of a relative low aspect ratio (a mean diameter of 0.96 μm and a mean length of 3.84 μm). The seed crystals were dispersed into well milled methanol slurry of α-Si₃N₄, 5 wt% Y₂O₃, and 2 wt% Al₂O₃ powders. The powder mixture was sintered at 1850 °C in a relatively low nitrogen gas pressure sintering of 0.9 MPa, and bimodal distribution of silicon nitride grains was obtained owing to epitaxial grain growth with the faster rate of c-axis from the dispersed seed crystals, as shown in Figure 3. The fracture toughness (determined by single-edge precracked-beam (SEPB) method) was improved from ~6 MPa m¹/² at no seed addition to ~9 MPa m¹/² at 5 vol.%, while the fracture strength determined by four-point flexure tests remained at high level of about 1 GPa (see Figure 4 (a) and (b)). The crack path observation for seeded materials revealed substantial deflection associated with the larger fibrous grains grown from the seed crystals, while non-seeded ones showed almost flat crack propagation. The compatibility of high fracture toughness and high strength is attributable to well controlled size, fraction, and distribution of the larger fibrous grains.

3. FIBROUS GRAIN ALIGNMENT

3.1. Coarse Grain Control via Seeding and Tape-Casting Techniques

By incorporating a controlled amount of fibrous β-Si₃N₄ single-crystal seed crystals into silicon nitrides, the seeding method not only provides a novel approach to improve the fracture toughness with retaining high strength [16] but also serves as an effective method to control the orientation as well as the size, content, and morphology of the fibrous β-silicon nitride grains [20,21]. In the whisker-reinforced ceramic composites, it has been well known that the orientation of whiskers plays an important role to govern their mechanical properties. The whisker orientation has been controlled by many techniques including tape-casting, slip-casting, hot-pressing, extrusion, etc. The fracture toughness measured for the crack perpendicular to the whisker orientation is higher than that of the composites with random distribution of whiskers.

By combining the seeding and tape-casting techniques, Hirao and co-workers [20] have tried to control the orientation as well as size and morphology of the fibrous grains in silicon nitrides. They added the seed crystals (a mean diameter of 1.3 μm and a mean length of 5.4 μm) into the slurry of α-Si₃N₄, 5 wt% Y₂O₃, and 2 wt% Al₂O₃ powder mixture. The slurry was tape-cast to form green sheets by using a doctor-blade equipment (the thickness of the sheets was adjusted to range from 130 to 150 μm), and subsequently the green sheets were punched in rectangular shape and stacked together aligned in the casting direction. After cold isostatic press (CIP) at 500 MPa and calcination at 600 °C, the compact was sintered at 1850 °C for 6 h under a nitrogen pressure of 1 MPa. The schematic illustration of the fabrication procedure is shown in Figure 5.

Figure 6 shows the microstructure of the seeded (2 vol.%) and tape-cast silicon nitride; the coarse fibrous grains with diameter of a few micrometers and length of about 20 μm, which were embedded in matrix of fine grains, tended to be aligned parallel to the casting direction. The polished surface parallel to the casting direction exhibited that the fibrous grains were somewhat randomly aligned.
oriented, but always stayed within the casting plane. It is very likely that the seed crystals aligned themselves during the casting process, since fibrous grains were epitaxially grown from the seed crystals. However, it should be noted that seed crystals are not completely aligned during this process; many are substantially tilted. When the stress was applied in the directions parallel to the grain alignment, high fracture toughness of $>11$ MPa m$^{1/2}$ (determined by SEPB methods) as well as high fracture strength of $>1.1$ GPa (measured by the four-point flexure tests) were obtained at 5 vol.% dispersion of the seed crystals (see Figure 4 (C)). Another feature of this silicon nitride is a narrow strength distribution. The Weibull modulus of the distribution is $>45$, which is substantially high compared with those for conventional silicon nitrides, typically $10-25$. Thus, the seeded and tape-cast silicon nitride shows synergistic improvement in all of the fracture strength, fracture toughness, and strength stability.

Imamura et al. [22] used three different-sized seed crystals to investigate the effects of the seed size on the microstructure and mechanical properties of the seeded and tape-cast silicon nitrides, and found that small seed crystals (a mean diameter of $~0.5 \mu m$ and a mean length of $~2 \mu m$) lead to highly anisotropic and fine microstructure with a monomodal grain-diameter distribution and improved mechanical properties. The fracture strength determined by the four-point flexure tests was $~1.4$ GPa and the fracture toughness determined by SEPB methods was $12$ MPa.m$^{1/2}$ in stress application parallel to the grain alignment (see Figure 4 (d)). Teshima et al. [23] employed extrusion process, in place of tape-casting, and succeeded in unidirectional alignment of fibrous seed crystals due to the high shear stress exerted during the extrusion process. Small matrix grains were almost consumed by selective grain growth into larger grains grown from seed particles, and consequently, large elongated grains with high aspect ratio were dominant in the microstructure of this material. Such microstructure development is related to the fact that the unidirectional alignment of seeds along a precise direction inhibits mutual impingement of elongated grains during growth along their length direction. Both high strength (four-point flexure tests) of $~1.4$ GPa and high fracture toughness (SEPB methods) of $~14$ MPa m$^{1/2}$ could be
achieved by testing in the direction parallel to grain alignment (see Figure 4(e)).

In the silicon nitrides with aligned fibrous grains, a large number of fibrous grains are involved with the crack-wake toughening mechanism when a crack propagates perpendicularly to the grain alignment, and then the toughening works more quickly in a short crack extension. Ohji et al. [24] investigated the fracture resistance curves ($R$-curves) of the seeded and tape-cast silicon nitrides, by measurements of initial (as indented) and fracture instability (indentation-strength measurement) crack lengths by Vickers indentation. The results are compared with that of a conventional self-reinforced silicon nitride as shown in Figure 7. The fracture resistance of the self-reinforced material increases continuously as a crack propagates and reaches 11 MPa m$^{1/2}$ at about 500 μm, similarly to other measurements for self-reinforced silicon nitrides [25,26]. The seeded and tape-cast material, however, exhibited high fracture resistance above 10 MPa m$^{1/2}$ from the beginning of the measured crack length range and remained almost constant in the following crack extension, when a crack propagated in the direction normal to the fiber axis (transverse crack propagation). Generally the fibrous grains, which span the crack wakes behind the crack tip, induce two crack-wake toughening mechanisms: elastic grain bridging and frictional grain pullout. The fibrous grain alignment gives at least a couple of benefits to these toughening mechanisms: one is that a greater number of the fibrous grains are involved with the toughening and the other is that the toughening works effectively because the fibrous grains stand normal to the crack plane. If the fibrous grains are inclined to the crack plane, additional bending is applied to the grains and the grains are readily fractured during crack opening, reducing substantially the toughening effects [27]. These benefits give rise to the above unique $R$-curve behavior of the seeded and tape-cast material in the transverse crack propagation. It is generally known that the crack bridging (or elastic bridging) is responsible for short crack toughening while the pullout is for the long crack one. It is suggested that the former toughening occurs very effectively in the seeded and tape-cast material due to the above benefits from the fibrous grain alignment.

The rising $R$-curve behavior generally observed in the conventional self-reinforced material has little contribution on catastrophic fracture, since long crack extension is required to obtain a substantial increase of fracture resistance. On the contrary, the steep rise of fracture resistance in the seeded and tape-cast material should be related to an increase in the catastrophic failure strength. In other words, high strength and high fracture toughness are compatible as revealed by Hirao et al. [20]. Another benefit brought by this steep $R$-curve is that the strength distribution can be narrower resulting in a higher Weibull modulus. It is known that if the flaws involved in strength determination exhibit rising $R$-curve behavior, the Weibull modulus substantially increases [28–30]. This leads to the high Weibull modulus of $>45$ as above stated.

3.2. Fine Grain Control via Forging Techniques

Superplastic forging and superplastic sinter-forging techniques are approaches to align relatively fine fibrous grains in the process of deformation. In the former the deformation stress is externally applied to a sintered body, while in the latter it is done during sintering. Schematic illustrations of these techniques are shown in Figure 8. In either approach, the fibrous grains are aligned along the tensile direction, or the direction perpendicular to the compression.

Superplastic deformation of silicon nitride occurs due to the grain boundary sliding when the grain size is small even if the grain shape is fibrous. In the case of fibrous grains, they are aligned along the tensile direction in the tensile deformation [31]. Using this phenomenon, Kondo et al. [32,33] realized superplastic plane-strain compressive deformation on silicon nitrides of fine and fibrous grains and investigated their strength and fracture toughness in comparison to those of the original material. Silicon nitride
prepared by gas-pressure sintering with additives of $Y_2O_3$ and $Al_2O_3$ was plane-strain deformed superplastically in a graphite channel die, with a pressure of 49 kN for 3 h at 1750 °C. This deformation resulted in 50% height reduction. The X-ray diffraction analysis confirmed that both the original and deformed specimens consisted of $\beta$-silicon nitride grains without a trace of $\alpha$-phase. The microstructure of the deformed specimen is shown in Figure 9 (a); the fibrous grains were aligned perpendicularly to the pressing direction. The fracture strength determined by three-point flexure tests for the original specimen was ~1.1 GPa, while it increased to ~1.7 GPa for the deformed specimen when the stress was applied along the extruding direction. The fracture toughness, measured by a single-edge-V-notched-beam (SEVNB) method, also increased from 8.5 MPa m$^{1/2}$ to 12 MPa m$^{1/2}$. It has also been known that this silicon nitride showed remarkably high fracture energies, 200–630 J/m$^2$, particularly at high temperatures [34]. In addition, the superplastically deformed silicon nitride showed substantially improved creep resistance at high temperatures, when the stress was applied along the extruding direction. The creep rates of the deformed specimen in tensile creep tests at 1200 °C was found to be about one order of magnitude lower than that of the original one [35].

Superplastic sinter-forging is a process where sintering and forging (or deformation) are carried out concurrently [36]. Compacted powder mixture of $\alpha$-Si$_3$N$_4$, 5 wt% $Y_2O_3$, and 3 wt% $Al_2O_3$ was sinter-forged with a pressure of 49 kN at 1750 °C for 3 h in 0.1 MPa of nitrogen. The microscopic study revealed that the obtained samples have similar microstructures to the superplastically deformed ones, but the grain size is relatively small compared to them (see Figure 9 (b)). When the stress was applied perpendicular to the compressive direction, the fracture strength determined by three-point flexure tests was as high as 2.1 GPa with the SEVNB fracture toughness of 8.3 MPa m$^{1/2}$.

As already stated, the fracture toughness of ceramic materials generally varies inversely with the fracture strength. The fracture toughness and strength values for a variety of silicon nitrides are plotted in Figure 10. It can be known that the antagonistic relation exists in the silicon nitrides fabricated by the two techniques shown here, though both the properties are substantially improved compared to other silicon nitrides. The superplastically sinter-forged specimen shows higher fracture strength but
lower fracture toughness than those of the superplastically deformed one [32], due to substantially small grain size. The former also exhibits higher strength than that with smaller grain size (~50 nm) [37], which is very likely attributable to the steep $R$-curve behavior and improved fracture toughness caused by the grain alignment, as well as to the reduced sizes of flaws during the deformation.

4. GRAIN BOUNDARY PHASE CONTROL

4.1. Fracture Resistance

Significant improvements in the fracture toughness of silicon nitrides are realized also by tailoring the chemistry of the intergranular amorphous phase. The interface between the β-silicon nitride grains and the intergranular glassy phase should be debonded to some appropriate extent so that the fibrous β-silicon nitride grains can effectively contribute to the toughening mechanisms such as crack bridging and crack deflection. The interfacial debonding in silicon nitrides can be adjusted by the composition of sintering additives which ultimately altered the composition of the intergranular glassy phase [21,39–41]. Chemical bonding across the interface determines the strength of the interface while the internal residual stress arising from thermal expansion mismatch imposed on the interface alters the debonding length. Sun et al. [21] described improvements in the fracture resistance of the seeded and tape-cast silicon nitride ceramics by controlling the chemistry of the intergranular glassy phase. They prepared the seeded and tape-cast silicon nitrides with different yttrium—aluminum ratios and investigated their $R$-curve behavior in relation to the microstructural features. Although the different sintering additives generally result in microstructures with different grain morphologies and sizes, the seeding method is effective in regulating the grain morphology and size, and then is an idealistic approach for investigating the effect of sintering additives on the fracture behavior. The different yttrium—aluminum ratio in the sintering additives leads to the different composition of the intergranular glass. The fracture resistance also varied as a function of the yttrium—aluminum ratio in the sintering additives, as shown in Figure 11. The stress was applied in the direction parallel to the fibrous grain alignment. All the behaviors obtained for different compositions exhibited steeply rising $R$-curves; i.e., the fracture resistance of the materials increased rapidly with initial crack extension. It should be noted, however, that the long crack (or plateau) toughness values increased systematically with increasing the yttrium—aluminum ratio. The in situ scanning electron microscopy (SEM) observation of cracks interacting with the microstructures also revealed that the interfacial debonding behavior between the large fibrous grains and the intergranular glass depended on the sintering additives employed. As shown in Figure 12, despite the similar angles of incidence, the fibrous grain in the sample of 4.0 wt% $Y_2O_3$—2.8 wt% $Al_2O_3$ failed transgranularly while...
crack deflection and interfacial debonding occurred in that of 5.0 wt% Y2O3 and 2.0 wt% Al2O3. For the smaller angles of incidence, interfacial debonding occurred in both the samples; however, the latter sample showed longer debonding lengths. They also estimated the critical debonding angle, beyond which no crack deflection could occur at the interface, at the different yttrium–aluminum ratio and found the angle to increase with increasing the ratio, being about 60, 70, and 75 degrees at the ratios of 4:2.8, 5:2, and 6.25:1, respectively. This indicates that, with the higher yttrium–aluminum ratio, the interface between the fibrous grains and the surrounding glass has a lower interfacial debonding energy and is more readily debonded, leading to the stronger R-curve behavior and the higher plateau toughness. Becher and co-workers [42–46] further investigated the influence of intergranular glass on the interface between the prismatic faces of β-silicon nitride grains and the glass and exhibited that, for example, the interfacial strength increases with increasing the aluminum and oxygen contents of the epitaxial β-SiAlON layer that forms on the Si3N4 grains. This was to the formation of a network of strong bonds (cross bonds) that span the glass–crystalline interface. The in situ high-resolution electron microscopy observations revealed that the debonding path can occur at the interface between the grains and the intergranular glass film or within the film, depending on the film’s composition [46]. Furthermore, the recent studies [47–50] clarified more precisely rare-earth adsorption behaviors at intergranular interfaces in silicon nitride ceramics, and its effects on the phase transformation, microstructure evolution, and mechanical properties in silicon nitrides. For example, Shibata et al. [47,48] showed that the rare earths have different tendencies to segregate to the grain surfaces in silicon nitrides using atomic-resolution scanning transmission electron microscopy and first-principles calculations, which is essential for forming elongated grains and a toughened microstructure.

4.2. Heat Resistance

It is essentially important to control the chemistry of the intergranular phase in silicon nitrides, not only to improve the fracture toughness but also to give the supreme heat or creep resistance. The glassy phases, which are present at grain boundaries of many ceramics including silicon nitrides, are softened at high temperatures, leading to severe degradation of the mechanical reliabilities. Then, the highly refractory phase is required to the interface when ceramics are used for high-temperature applications. In recent decades, great deal of efforts to strengthen the grain boundary phase for improving heat resistance of silicon nitride have been devoted, leading to significant improvements of the high-temperature mechanical reliability. For example, some grades of silicon nitrides sintered with lutetia sintering additive showed excellent heat and creep resistances even at temperatures above 1400 °C [51–54]. The X-ray diffraction (XRD) analyses for these materials revealed Lu2Si2O7 and Lu4Si2N2O7 as secondary phases.

Kondo et al. [55] fabricated highly heat resistant silicon nitride by using the sinter-forging technique with a lutetia
sintering additive. The specimen exhibited the fracture strength of ~700 MPa at 1500 °C, which is remarkably high compared with high-temperature strength reported so far for silicon nitride. Such superior high-temperature strength was attributed to grain alignment as well as to the refractory grain-boundary glassy phase and the existence of glass-free grain boundaries.

Zeng et al. [56] tried to make highly refractory phase at grain boundaries in the seeded and tape-cast silicon nitrides, by using Lu2O3−SiO2 sintering additive systems. The slurry of α-Si3N4 powder, 3 wt% β Si3N4 seed crystals, 1 wt% SiO2, and 9 wt% Lu2O3 was tape-cast; sintering was carried out in 0.9 MPa N2 atmosphere at 1950 °C for 6 h. Typical examples of the microstructure and grain boundary phase are shown in Figure 13. Because of the higher sintering temperature than the previous works [20–23], very large fibrous grains, whose length and diameter were typically several tens of μm and ~5 μm, respectively, were grown from fibrous β-Si3N4 seed crystals and tend to align along the casting direction. Again, it should be noted that many of them are not completely aligned but substantially inclined in the tape planes. Highly refractory Lu2Si2O7 phase was detected as secondary phase at many grain boundaries.

Figure 14 shows the fracture strength determined by three-point flexure tests and fracture energies measured by chevron-notched beam (CNB) tests [57] at room temperature and 1500 °C, in comparison with those of the unseeded material, which was obtained in the same procedures except for not adding β-Si3N4 seed crystal. The stress was applied in two directions parallel (strong direction) and perpendicular (weak direction) to the fibrous grain alignment. The strength measured in the strong direction at 1500 °C was equivalent to the room-temperature one, while the unseeded material shows substantial strength degradation. The strength in the weak direction is low; however, it is comparable to that of the unseeded one at 1500 °C. The excellent fracture strength at high temperatures is very likely attributable to the improved fracture resistance. The fracture energy of the unseeded material was about 100 and 450 J/m² while that of the seeded one in the strong direction was about 300 and 800 J/m², at room temperature and 1500° C, respectively. The large fracture energies apparently were attributable to enhanced crack shielding effects of bridging and frictional pullout of the very large fibrous grains. It should be noted, however, that even in the weak direction the seeded material showed substantially high fracture energies particularly at the high temperature. This is presumably due to the large number of the substantially inclined fibrous grains, which give rise to grain bridging and frictional pull out effects, even in the weak direction. The inclined fibrous grains are also beneficial to toughening in the strong direction due to the so-called wedge effects [58], as shown in Figure 15. Compared to the completely aligned grains, further energy is required for the grains to be pulled out, owing to the additional bending applied to the grains. This mechanism, however, works effectively when the grains are strong enough, or the surrounding phases are weak enough, for the inclined grains not to be fractured during crack opening by such additional bending force. Thus, the large fracture energy can be obtained in the seeded materials which contain the large fibrous grains and at high temperatures where the grain boundary phases are softened.
5. POROUS STRUCTURE CONTROL

In structural materials, pores are generally believed to deteriorate the mechanical reliability; however, this is not always true whenever the microstructure is carefully controlled. This section shows how the mechanical properties change when the size, shapes, and orientation of pores as well as grains are controlled, taking examples of porous silicon nitrides. Particular emphasis is placed in unique or improved properties including fracture energy, fracture toughness, strain tolerance (fracture strain), and thermal shock resistance through such porous structure control.

5.1. Isotropic Porous Structure

A variety of processes have been developed to produce porous silicon nitride ceramics. Sintering powder compacts to a fixed degree of densification, so-called partial sintering, is one of the most frequently employed approaches to fabricate porous ceramic materials. In the case of oxide materials, due to their good sinterability, the density or porosity can be adjusted by the heating characteristics, such as temperature and holding time. The sintering of silicon nitride ceramics is difficult, however, because of strong covalent bonding between silicon and nitrogen atoms, so that sintering additives are necessary for consolidating silicon nitride ceramics by liquid-phase sintering. On the other hand, the difficulty of sintering silicon nitride ceramics is beneficial for controlling density or porosity through adjusting the additives and the sintering process.

A fibrous silicon nitride microstructure is developed during sintering at adequately high temperatures in porous material, and excellent mechanical properties can be obtained compared with those of porous oxide ceramics obtained by partial sintering. Yang et al. [61] fabricated porous silicon nitride ceramics using Yb$_2$O$_3$ sintering additives and investigated the microstructures and mechanical properties. Sintering was performed at various temperatures between 1600 and 1850 °C under a nitrogen gas pressure of 0.6 MPa. A liquid-phase sintering technique and powder compaction similar to the fabrication of dense silicon nitride were used to make the fabrication process simple and cost-effective. The sintering additive, Yb$_2$O$_3$, is known to form crystalline Yb$_4$Si$_2$O$_7$N$_2$ at grain boundaries or triple junctions [62,63] which presumably enhances the high-temperature mechanical properties of silicon nitride ceramics. The addition of Yb$_2$O$_3$ also is known to accelerate the fibrous grain growth of β-Si$_3$N$_4$ [64], which is advantageous for the crack shielding effects of grain bridging and pullout. Furthermore, because of the high melting point of Yb$_2$O$_3$ and the high viscosity of Yb$_2$O$_3$-related glass, this additive is unsuitable for the densification of silicon nitride, a characteristic that is conversely beneficial for the fabrication of porous silicon nitride.

Yang et al. [61] fabricated porous silicon nitride ceramics using Yb$_2$O$_3$ sintering additives and investigated the microstructures and mechanical properties. Sintering was performed at various temperatures between 1600 and 1850 °C under a nitrogen gas pressure of 0.6 MPa. A liquid-phase sintering technique and powder compaction similar to the fabrication of dense silicon nitride were used to make the fabrication process simple and cost-effective. The sintering additive, Yb$_2$O$_3$, is known to form crystalline Yb$_4$Si$_2$O$_7$N$_2$ at grain boundaries or triple junctions [62,63] which presumably enhances the high-temperature mechanical properties of silicon nitride ceramics. The addition of Yb$_2$O$_3$ also is known to accelerate the fibrous grain growth of β-Si$_3$N$_4$ [64], which is advantageous for the crack shielding effects of grain bridging and pullout. Furthermore, because of the high melting point of Yb$_2$O$_3$ and the high viscosity of Yb$_2$O$_3$-related glass, this additive is unsuitable for the densification of silicon nitride, a characteristic that is conversely beneficial for the fabrication of porous silicon nitride.

The microstructures of the porous silicon nitrides sintered at different temperatures are shown in Figure 16. The microstructure of the sample sintered at 1600 °C consisted solely of fine, equiaxed grains. The average grain size was almost the same as that of the starting powder, indicating
little phase transformation, and the XRD analysis identified almost only α-Si₃N₄ peaks. When the sintering temperature increased above 1700 °C, however, the formation and development of β-Si₃N₄ grains were observed, indicating enhanced phase transformation and grain growth. At 1700 °C, very fine, fibrous β-Si₃N₄ grains were obtained, and with increasing the sintering temperature, the microstructure becomes coarse while the porosity changes relatively little. The fracture strength, σ_f, and fracture toughness (critical stress intensity factor), K_IC, are shown also in Figure 16 as a function of sintering temperature. Each data point marks an average of five or six measurements. The fracture strength determined by three-point flexure and the fracture toughness was converted from the fracture energy measured by chevron-notched beam (CNB) tests using the following relationship \[ \gamma_{\text{eff}} = K_{\text{IC}}^2 (1 - \nu^2)/2E \] (1) where \( E \) is the Young’s modulus and \( \nu \) is the Poisson’s ratio. The obtained fracture toughness is an averaged value when the toughness varies with crack extension (R-curve).

When the sintering temperature is 1600 °C and the microstructure is equiaxed, the mechanical properties are low; however, the properties are markedly improved along with the fibrous grain formation above 1700 °C. For example, the sample sintered at 1700 °C has a strength of 380 MPa, which is about 10 times larger than that at 1600 °C. As the sintering temperature further increased and the microstructure became coarser, the strength decreased while the fracture toughness increased. Nonetheless, the toughness values are still lower than 4 MPa m^{1/2}.

Mechanical properties of porous silicon nitrides have also been investigated as a function of porosities by using partial hot-pressing process, where the porosity is determined by the configuration of the carbon mold and the powder amount, leaving other parameters such as characteristics of a powder mixture and sintering additives fixed [59]. Two series of porous material with different porosities were prepared: one was fine-grained material series sintered at 1700 °C and the other was coarse-grained at 1850 °C. Figures 17 (a) and (b) showed porosity dependences of Young’s modulus, \( E \), fracture strength, \( \sigma_f \), fracture energy, \( \gamma_{\text{eff}} \), and fracture toughness, \( K_{\text{IC}} \), for the

![FIGURE 16](image-url) Microstructures and mechanical properties (flexural strength and fracture toughness) of porous silicon nitrides sintered with 5 wt.% Yb₂O₃ at 1600 °C, 1700 °C, 1800 °C and 1850 °C. “P” denotes porosity. (Reprinted from Ref. [61] with permission of Elsevier. All rights reserved)

![FIGURE 17](image-url) Young’s modulus, \( E \), fracture strength, \( \sigma_f \), fracture energy, \( \gamma_{\text{eff}} \), fracture toughness, \( K_{\text{IC}} \), and fracture strain, or strain tolerance, \( \varepsilon_f \) as a function of porosity for (a) fine-grained and (b) coarse-grained porous silicon nitrides. The values are normalized by the respective values of the dense material: \( s_f = 1.1 \) GPa, \( E_0 = 330 \) GPa, \( K_{\text{ICO}} = 6.3 \) MPa m^{1/2}, and \( \gamma_{\text{eff0}} = 70 \) J/m² for the fine-grained material and \( s_f = 1.0 \) GPa, \( E_0 = 330 \) GPa, \( K_{\text{ICO}} = 7.0 \) MPa m^{1/2}, and \( \gamma_{\text{eff0}} = 90 \) J/m² for the coarse-grained one.
fine- and coarse-grained material series, respectively. The values are normalized by the respective values of the dense material: \( \sigma_{f0} = 1.1 \) GPa, \( E_0 = 330 \) GPa, \( K_{IC0} = 6.3 \) MPa \( m^{1/2} \), and \( \gamma_{eff0} = 70 \) J/m\(^2\) for the fine-grained material and \( \sigma_{f0} = 1.0 \) GPa, \( E_0 = 330 \) GPa, \( K_{IC0} = 7.0 \) MPa \( m^{1/2} \), and \( \gamma_{eff0} = 90 \) J/m\(^2\) for the coarse-grained one. For the fine-grained material, the fracture energy first increases somewhat with increasing the porosity up to 10\(\text{–}20\%\) and then decreases while all the other properties continuously decrease. The degree of decrease in the Young’s modulus is larger than that in the strength in the porosity range of 20\(\text{–}30\%\); therefore, the fracture strain, or strain tolerance, \( \varepsilon_f \), given by \( \varepsilon_f = \sigma_f/E \) is substantially high in this range. On the other hand, for the coarse-grained material, the fracture energy shows similar tendency, but is relatively high compared to that of the fine-grained material. While all the other properties continuously decreased, the degree of decrease in the Young’s modulus is smaller than that in the strength, resulting in the low fracture strain. The existence of pores lowers the strength, Young’s modulus, and fracture toughness, but may increase the fracture energy and fracture strain depending on the porosity etc. in the isotropic porous silicon nitrides.

5.2. Anisotropic Porous Structure with Fibrous Grain Alignment

Taking the seeding and tape-casting technique, Inagaki et al. [66,67] developed a porous silicon nitride, where the fibrous grains were uniaxially aligned (hereafter denoted by “anisotropic porous silicon nitride”) and investigated the effects of pores on the mechanical properties. Since the pores exist around the aligned fibrous grains, enhanced operations of the grain bridging and pullout can be anticipated when the crack propagates perpendicularly to the direction of grain alignment. Furthermore, the presence of pores surrounding the silicon nitride grains causes cracks to tilt or twist, namely crack deflections [68,69]. As starting materials they used \( \beta \)-Si\(_3\)N\(_4\) seed crystals, which were obtained by heating the powder mixture of \( \alpha \)-Si\(_3\)N\(_4\), 5 wt\% Y\(_2\)O\(_3\), and 10wt\% SiO\(_2\) in a silicon nitride crucible at 1850 °C for 2 h [19]. The slurry containing the fibrous seed crystals mixed with 5 wt\% Y\(_2\)O\(_3\) and 2 wt \% Al\(_2\)O\(_3\) was tape-cast so that the seed crystals in the sheets were aligned along the casting direction. After the green sheets were stacked and bonded under pressure, sintering was performed at 1850 °C under a nitrogen pressure of 1 MPa. The texture of the anisotropic porous silicon nitride with porosity of 14% was shown in Figure 18; the material consists only of fibrous grains, pores, and grain boundary phase. The fibrous grains of silicon nitride tend to be aligned along the casting direction, but many of the grains are substantially tilted. The pores, whose shapes are mostly plate-like along the same direction, exist among the grains.

The effect of porosity on the mechanical properties of the anisotropic porous silicon nitride is investigated in both directions parallel to and perpendicular to the grain alignment. Figure 19 shows the porosity dependencies of fracture strength and fracture energy when the stress is applied in the parallel direction to the grain alignment of the porous silicon nitride, in comparison with those of the isotropic coarse materials shown in Figure 17 (b). The strength was measured by three-point flexural tests while the fracture energy was determined by the CNB technique. The numbers of measurements for the former and latter are six and three, respectively, and the plots in the figure are their averages. The fracture strength became larger as the porosity decreased. In the porosity ranging below 5%, the

![FIGURE 18](https://example.com/figure18.png)  
Microstructures of anisotropic porous silicon nitride prepared by tape-casting fibrous seed crystals (parallel to the casting plane). The porosity is 14%. (Reprinted from Ref. [66] with permission of John Wiley and Sons. All rights reserved)

![FIGURE 19](https://example.com/figure19.png)  
Porosity dependence of fracture strength and fracture energy of anisotropic porous silicon nitrides, in comparison with those of the isotropic materials shown in Figure 17 (b). Stress is applied parallel to the alignment direction. Bar is the standard deviation.
strength attained above 1.5 GPa. This value was almost comparable to that of fibrous-grain-aligned dense silicon nitride fabricated through superplastic forging [32]. When fibrous grains were aligned, pores around the grains promoted debonding between interlocking fibrous grains without fracture. Therefore, even small amount of pores enhanced the grain bridging improving the strength. On the other hand, fracture energy of porous silicon nitride ranged from 300 to 500 J/m² in the porosity range below 20%. These values were considerably high compared with other studies even for dense silicon nitrides [34,57]. Large fracture energy was mainly due to the effect of bridging crack by aligned fibrous grains and/or pullout of the grains. Debonding of grain boundaries by the existence of pores and aligned grains bridging the crack enhanced these crack shielding effects. Figure 20 shows the micrograph of the ligament area of fractured surface; the protruding grains and holes from which the fibrous grains have been pulled out clearly support these speculations. As porosity increased nearly 15%, fracture energy became larger; however, fracture energy decreased monotonously in the porosity range over 15% with increase of the porosity. This drop of fracture energy is presumably due to the reduction of substantial bridging and/or pullout area.

The porosity dependences of the fracture toughness estimated from the fracture energy using Eqn 1 as well as the fracture strength are shown in Figure 21, in comparison with those of the isotropic coarse materials. High fracture toughness above 17 MPa m¹/² as well as high strength above 1.5 GPa were attained in the porosity range below 5%. Especially, fracture toughness of a specimen with a very small porosity less than 5% exceeded that of a fully dense specimen and became close to 20 MPa m¹/².

Figure 22 shows the dependencies of fracture strength and fracture energy on porosity in the perpendicular direction to the grain alignment of the porous silicon nitride (weak direction), in comparison with those of the isotropic porous silicon nitrides. While the fracture strength is lower than that of the isotropic material, the fracture energy is substantially higher in the whole porosity range, due to the wedge effects caused by tilted large fibrous grains as shown in Figure 15.

5.3. Thermal Shock Resistances

When ceramic materials are subjected to rapid change of temperature, substantial thermal stress arises, sometimes...
resulting in cracking or complete failure. This failure is called as thermal shock fracture. Particularly porous ceramic components used at high-temperature such as hot gas filters are susceptible to this fracture. Thus, to designers and end users of ceramic components, it is essentially important to know properly the resistance properties against thermal shock for the material. This section intends to describe how thermal shock resistances vary when the porous structures change in porous silicon nitrides.

Generally mechanical strength of ceramics after thermal shock remains constant until the temperature difference, $\Delta T$, reaches a critical value, $\Delta T_c$, which is determined by the shape and sizes of specimen and heat transfer conditions, in addition to the material properties. When $\Delta T > \Delta T_c$, a crack propagates catastrophically or quasi-statically depending on the initial crack size, leading to abrupt or gradual decrease of strength. Therefore, two parameters, “thermal shock fracture resistance”, $R$, and “thermal shock damage resistance”, $R''$, are often used to characterize the resistance against thermal shock crack initiation and propagation, respectively. As defined by Hasselman [70,71], $R$ and $R''$ are given as follows:

$$ R = \sigma_f (1 - v)/E \alpha \propto \Delta T_c $$  \hspace{1cm} (2)

$$ R'' = E_{\text{ef}}/\sigma_f^2 (1 - v) $$  \hspace{1cm} (3)

where $\alpha$ is the coefficient of thermal expansion. It should be noted that the effects of fracture strength and Young’s modulus on $R$ and $R''$ are completely opposite; when the fracture strength is high and the Young’s modulus is low, $R$ is high but $R''$ is low. From the properties including the fracture strength, Young’s modulus, and fracture energy, it is possible to estimate $R$ and $R''$ of the porous silicon nitrides as a function of porosity. Figure 23 (a) and (b) show the estimated $R$ and $R''$ for the isotropic fine- and coarse-grained porous silicon nitrides which were sintered at 1700°C and 1850°C, respectively. For the fine-grained materials, the thermal shock fracture resistance, $R$, first increases with increasing the porosity up to 20–30%, and then decreases while the thermal shock damage resistance, $R''$, remains almost constant. For the coarse-grained materials, $R$ decreases and $R''$ increases monotonously with increasing the porosity. It should be noted that completely different tendencies are obtained for both the $R$ and $R''$ between the fine- and coarse-grained materials.

In order to experimentally characterize the $R$ and $R''$, thermal shock tests were performed by dropping the heated specimens from a resistance furnace into a container of water at 20 °C [72]. The specimen was covered with mullite blocks on all sides except one face exposed to the water quenching. After heating in air at a rate of 5 °C/min to a preset temperature and holding at this temperature for 30 min, the specimen was quenched into a water bath at 20 °C. The fracture strengths of the quenched specimens were determined at room temperature by three-point flexure tests. Figure 24 shows the results of the thermal shock tests for the isotropic fine- and coarse-grained porous silicon nitrides, whose porosity is 30% and 28%, respectively. The fine-grained material shows strength retention up to 1200 °C and sharp drop above this temperature, indicating relatively good $R$ and poor $R''$. On the contrary, in the coarse-grained material, the strength starts to gradually decrease around 600 °C, indicative of relatively poor
The thermal shock fracture resistance, $R$, and thermal shock damage resistance, $R''$, of the anisotropic porous silicon nitride as a function of porosity, in comparison with those of isotropic dense fine-grained material.

$R$ and good $R''$. This is consistent with the results of estimated $R$ and $R''$ shown in Figure 23. It should be noted again that this notable difference in thermal shock fracture behavior arises solely from the different sintering temperatures of 1700°C and 1850°C.

Figure 25 shows the estimated $R$ and $R''$ of the anisotropic porous silicon nitride as a function of porosity, in comparison with those estimated for dense fine-grained silicon nitride whose property is shown in Figure 17 (a). The thermal shock fracture resistance, $R$, remains high in the porosity region up to 20% and then starts to decrease. On the other hand, the thermal shock damage resistance, $R''$, increases with increasing porosity up to 30% followed by the plateau region afterward. It should be noted that high values of both $R$ and $R''$ are realized in the porosity range of 15~25%, which are much higher than those of the dense material. Figure 26 shows the results of the thermal shock tests for the anisotropic porous silicon nitride (14% porosity) in comparison with the above dense material. The anisotropic porous material shows the excellent $R$ with $\Delta T_e$ larger than 1200 °C, while that of the dense material is about 1200 °C, indicating good agreement with the results of estimated $R$ shown in Figure 25.

REFERENCES


Chapter 9.4

Determination of the Mechanical Reliability of Brittle Materials*

Stephen W. Freiman† and Jeffrey T. Fong**

† Freiman Consulting, 8212 Inverness, Hollow Terrace, Potomac, Maryland 20854, United States, ** Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Chapter Outline

1. Introduction 675
2. Lifetime Prediction Expressions 676
3. Measurement Procedures 676
4. Uncertainty Calculations 677
5. Summary 679
Acknowledgments 679
References 679

1. INTRODUCTION

Were it not for the existence of subcritical crack growth, designing ceramic components to withstand known stresses would be relatively straightforward, requiring only a statistical estimate of the strength of the material. However, most ceramics are subject to slow crack growth under stress in moisture-containing environments, a phenomenon that can lead to mechanical failure after a period of time at stresses lower than the short-term fracture strength [1]. Examples of components that are vulnerable to delayed failure include optical fibers (SiO₂), multilayer capacitors (BaTiO₃), infrared transmitting windows (ZnS, ZnSe), and piezoelectric transducers, to name but a few. The stresses that could lead to failure can arise, not just due to mechanical loads, but also as a result of thermal gradients, phase transformations, or the presence of electric fields.

What is needed is a method to assure that the most severe flaw in a component is not subjected to a stress such that it will grow to a critical size before the desired lifetime of the part. Three methods for assuring such reliability are available. The most direct is nondestructive evaluation of a part prior to it being placed into service, which in principle would allow the identification of a flaw that could lead to failure at short times. Unfortunately, because of the small sizes of the flaws in most ceramics, no procedures with the necessary resolution to distinguish one critical flaw from the many others in the distribution are available today.

A second alternative is proof testing in which the part is loaded to a stress exceeding that expected in service, and then rapidly unloaded. While this procedure can be effective if carried out correctly, it is expensive to conduct on every part, and a significant loss of parts during the proof test should be expected. In addition, the stresses must be applied in a dry environment, exactly as would be expected in service, and rapid unloading from the proof stress is essential [2].

We are therefore left with the necessity of using fracture mechanics expressions relating crack growth rates to material parameters and known loads, coupled with a statistical analysis of the experimental data as a means of determining reliability. An essential point is that each of the parameters needed for the reliability assessment has a measurement uncertainty associated with it. It is crucial to determine not only the predicted lifetime but also to know to what degree of confidence it can be specified. Previous studies have addressed this issue in various ways [3,4].

This chapter provides a summary of suggested testing methodology, analysis techniques, and statistics needed to assure safe operation of brittle components over specific times, including a discussion of the methods by which the results of laboratory-scale specimens can be used to predict the reliability of a system of parts.
2. LIFETIME PREDICTION EXPRESSIONS

Our ability to predict the reliability or lifetime of a ceramic part rests on the use of a mathematical expression to represent the crack growth data. Following chemical rate theory, which has been shown to be applicable to the environmentally enhanced crack growth process [1], the following exponential relationship between crack velocity, \( V \), and the stress intensity factor, \( K_I \), is derived:

\[
V = V_0 \exp(bK_I).
\]  

(1)

\( V_0 \) is a material-dependent constant containing terms such as the relative humidity of water in the particular environment and the temperature, and \( b \) is a constant proportional to the activation volume for the process [1]. In principle, Eqn (1) can be integrated to yield a time for a flaw to grow from its initial size to criticality. However, there are two problems: First, a closed-form solution for the integration is difficult to achieve. While such integration has been reported [5], the integration limits are in terms of flaw size, which from a practical point of view cannot be determined with any degree of accuracy. Second, the effect of a residual stress term associated with a flaw has not been incorporated into such an expression [6].

An empirical power law is used instead:

\[
V = V_0 \left( \frac{K_I}{K_{IC}} \right)^N.
\]  

(2)

The exponent \( N \) represents the slope of the crack growth curve. In water, \( N \) ranges from as low as 15 for oxide glasses to \( \geq 50 \) for ceramics such as silicon nitride and silicon carbide. \( K_{IC} \) is the critical stress intensity factor, used here as a scaling parameter.

However, most flaws in actual components will experience not only a far-field stress but are also surrounded by a stress field generated by the impact-type process by which they were introduced, for example, machining and finishing. Eqn (2) is then modified to the following [7]:

\[
V = V_0 \left( \frac{K_I}{K_{IC}} \right)^{N'}.
\]  

(3)

where \( N' = (3N + 2)/4 \) for "point flaws," that is, elliptical cracks typically created by machining or other impact processes; \( N' \) is smaller than \( N \) and so provides a more conservative estimate of flaw behavior in a part.

It is useful to put \( K_I \) in terms of the stress, \( \sigma \), in the component through

\[
K_I = Y\sigma a^{\frac{1}{N'}}
\]  

(4)

where \( a \) is the flaw depth in the surface, and \( Y \) is a geometric constant dependent on loading, flaw geometry, and location. Here, \( \sigma \) is the stress in the specimen or part.

Fuller et al. [7] showed that a time to failure, \( t_f \), due to cracks emanating from "point flaws":

\[
t_f = \frac{4^{-N'-3}}{(3N-4)^{-2}} \left( \frac{2K_{IC}^2}{Y^2V_0} \right) S^{N'} \sigma^{-N'},
\]  

(5)

where \( S \) is the strength of the as-finished material. The other terms have been previously defined. However, because of the need to separately measure \( K_{IC} \) and \( V_0 \) using large cracks in fracture mechanics specimens, this form of the expression is unwieldy to apply. Not only might these parameters be difficult to obtain in this way, for example, because of amounts of material required but the results also could differ from those applicable to the small flaws of interest.

Fuller et al. [7] also showed that the term in the parentheses in Eqn (5) can be expressed as

\[
\frac{2K_{IC}^2}{Y^2V_0} = \left( \frac{3N-2}{(N'+1)(4N'-3)} \right) S_v^{2-N'},
\]  

(6)

where \( \lambda \) is obtained from a plot of stressing rate versus strength measurements made on indented specimens, that is, dynamic fatigue tests, and \( S_v \) which we call the reference strength, is the strength of a set of indented specimens in which the as-created cracks are not allowed to grow subcritically. The right-hand side of Eqn 6 contains only parameters that can be obtained from strength measurements.

We now combine Eqn (5) with Eqn (6):

\[
t_f = \frac{\lambda}{N'-1} \left( \frac{S}{S_v} \right)^{N'-2} \sigma^{-N'}.
\]  

(7)

Both \( S \) and \( S_v \) will be taken as the minimum values of strength within the measured distributions.

3. MEASUREMENT PROCEDURES

3.1. Initial Strength Distribution

The initial strength, \( S \), is based on the flaw distribution in the material just before a part is put into service. In carrying out a reliability analysis, a key assumption is that no flaws more severe than those in the initial flaw distribution are introduced during the lifetime of the part. In addition, it is assumed that flaws grow only due to environmentally enhanced crack growth.

It would be advantageous if the strength distribution of actual parts could be obtained under loading conditions that simulate the service loading conditions. However, this can be difficult and costly. Consequently, tests are usually conducted on small pieces of the same material. A key assumption is that the processing procedures and surface treatments, for example, machining and polishing of these specimens, are identical to that seen by the component.

Because of the propensity of brittle materials to fail from surface flaws, flexural tests are the primary test...
method. Such tests can be conducted in either uniaxial or biaxial loading. Because failure from sharp edges is typically a problem with brittle materials, the decision of whether to conduct uniaxial or biaxial tests is usually based on whether edges in the part will be subject to significant stresses. If uniaxial testing, for example, four-point flexure, is chosen, edges should be rounded or chamfered to minimize failure from them. Testing must be conducted at a high loading rate in an inert environment, for example, dry gaseous nitrogen, in order to eliminate slow crack growth effects [8]. In engineering practice, about thirty specimens are recommended to assure that an adequate statistical distribution can be established. However, this number is rooted in the statistics literature to be compatible with the rigorous estimate of confidence levels [9]. The actual number of specimens that are needed will depend on the scatter in the data and the standard deviation (sd) desired [10].

At times, a component may experience a stress state that could cause it to fail from internal flaws such as pores or inclusions. In this case, flexural tests will not be effective, and direct tensile tests will be required.

In the simplest case, all the flaws leading to failure are of the same type and come from one population, that is, one source. In many instances, however, there are multiple flaw populations involved leading to more complex strength distributions. Fractographic analysis of the broken specimens, particularly those in the low-strength region, is important to determine the actual cause of failure and to separate unusually low strengths from those in the general population [11].

3.2. Crack Growth Parameters

While values of the crack growth parameters, \( N \) and \( \lambda \), can be obtained from direct measurements of crack growth rates, dynamic fatigue tests employing cracks that are of comparable size to actual flaws in the component are both more convenient, as well as more relevant to the practical failure issue. As noted earlier, indentations surrounded by a stress distribution lead to a more conservative value of the environmental exponent, that is, \( N' \) rather than \( N \).

In the dynamic fatigue test, indented flexural bars are loaded to failure at a constant stressing rate. The relevant expression from which the crack growth parameters are extracted is [12]:

\[
\sigma_{f}^{N'+1} = \lambda \frac{d\sigma}{dt}
\]

A linear regression of the log of the failure strength, \( \sigma_f \), on the log of the stressing rate, \( \dot{\sigma} \), yields the parameters \( N' \) and \( \lambda \). The stressing rate tests should be carried out at the harshest temperature and environment that will be experienced by the part. It is necessary to take the data over as wide a range of stressing rates as possible, but not less than three orders of magnitude, in order to obtain the required accuracy.

ASTM Standard C-1368-06 [13] provides more details of the testing and analysis procedures for this test. The Standard also describes conditions by which specimens can be preloaded to positive tensile stresses before the stressing rate test itself to reduce the time needed to carry out tests at the lowest rates.

A crucial assumption in obtaining the data in this way is that the strength versus stressing rate curve can be extrapolated to lower stressing rates over a number of orders of magnitude, implying that the mechanism of crack extension does not vary at low stressing rates. We know that for certain materials, there is a downturn in crack growth rates at low stress intensity factors [1]; this would then lead to a conservative prediction of failure time. In a very few cases, however, crack growth deviates to a plateau at small stress intensity factors [1]. The procedure outlined here should not be used for these materials.

3.3. Reference Strength

Because indented specimens were used to determine \( N' \), measurements are needed to calculate \( S_0 \), the upper limit to the indented strength. These strength measurements must also be made at a high loading rate in an inert environment.

3.4. Stresses in the Component

The stress distribution in the component is typically obtained through a finite element analysis.

4. UNCERTAINTY CALCULATIONS

It is not sufficient to simply calculate a time to failure; one must also know the uncertainty in this calculation, particularly its lower bound. Each of the measured parameters described in the previous section will contain some degree of uncertainty. These uncertainties, which can be calculated from standard statistical models [11], must be combined to yield an uncertainty in the predicted time to failure. Here, we describe a three-step strategy to estimate the uncertainty of the failure time in the full-scale structure.

The first and most important step is to determine the uncertainty in the minimum strength, \( S \), of the specimens. This is the most critical parameter in the data set because it is a measure of the most severe flaw. Second, by combining uncertainties, we can determine the uncertainty in the minimum lifetime of the specimen set as determined from Eqn (7). Third, we then determine the uncertainty in the lifetime of the part itself.
4.1. Step 1

To determine the uncertainty in the minimum $S$, one fits the measured strength distribution to a particular mathematical expression. The Weibull function, which is typically used in fracture analysis, is only one of many possible expressions that could be employed but has been used extensively because it seems to represent the underlying physics governing brittle fracture, namely, weakest link theory [14,15]. Historically, a two-parameter Weibull distribution, which allows for failure at zero applied stress, has been used to fit strength data and is the basis for the ASTM standard [16]:

$$P = 1 - \exp \left[ - \left( \frac{\sigma}{\sigma_0} \right)^m dV \right].$$

(9)

In this expression, $P$ is the probability of failure, $\sigma$ is the applied stress, and $\sigma_0$ the Weibull scale parameter. However, this approach is unduly conservative; flaws in components that would fail at extreme stresses could be visually observed, or possibly the component would break before being placed into service. Other, more complex, probability distributions have been put forward as well [17,18]. It is suggested that the three-parameter Weibull distribution be used. The expression is given by:

$$P = 1 - \exp \left[ - \int \left( \frac{\sigma - \sigma_n}{\sigma_0} \right)^m dV \right].$$

(10)

Here, $\sigma_n$ is the minimum in the strength distribution, or the location parameter in the statistical literature. To some degree, assuming that the Weibull model is the only possible choice impedes our ability to fit the experimental data with a minimum of uncertainty. It is advisable to use goodness-of-fit tests to establish the best statistical fit to the data. One can then choose the acceptable level of confidence desired for the lower bound in this calculation [19–21].

4.2. Step 2

From Eqn (7), we can see that the time to failure, $t_f$, is a function of five parameters, namely, $\lambda$, $N'$, $S$, $S_v$, and $\sigma$. Provided that the coefficient of variation of each of the parameters is <0.1, the variance of the time to failure, $\text{var}(t_f)$, where $\text{var}(t_f) = (\text{sd}(t_f))^2$ [2], can be derived using established error propagation formulas as a function of the variances of those five parameters [22]. For example, the expression for the sd of $t_f$, ignoring covariances, is given by

$$\text{sd}(t_f) = t_f \sqrt{\frac{\text{var}(\lambda)}{\lambda^2} + \frac{\text{var}(N')}{(N' + 1)^2} + (N' - 2)^2 \frac{\text{var}(S)}{S^2}} + \frac{\text{var}(S_v)}{(S_v^2)} + (N')^2 \frac{\text{var}(\sigma)}{\sigma^2} \frac{m}{C_0}.$$  

(11)

Note that the sd in $t_f$ increases in direct proportion to the time to failure.

If the propagation of error calculation yields sds in failure times that are unacceptably large, there are three options to consider. Although the parameters $N'$, $\lambda$, and $S_v$ are functions of the material selected, and therefore would not be subject to change, it may be possible to increase the strength of the material and thereby shift the position of the Weibull distribution as well as increase $\lambda$ by better machining and polishing procedures. Second, it may be possible to reduce the stress state in the component through a different mechanical design. A third option is to simply increase the size of the initial strength data set. The more specimens that are tested, the better is the knowledge of the entire flaw size distribution, and therefore, the smaller is the sd.

4.3. Step 3

The preceding uncertainty analysis was based on data taken on specimens. There will likely be many more manufactured parts than specimens tested, and they may have larger surface areas (or volumes) under stress. The two-parameter Weibull distribution allowed one to compare the stressed area or volume of a part to that of a specimen through the following expression:

$$\frac{\sigma_1}{\sigma_2} = \left( \frac{A_1}{A_2} \right)^{\frac{1}{m}},$$

(12)

where $\sigma_1$ and $\sigma_2$ are the strengths of the specimen and the component, respectively, and $A_1$ and $A_2$ are the corresponding stressed areas; $m$ is the Weibull modulus. However, Eqn (12) is only applicable to a two-parameter Weibull distribution. It has no meaning for a three-parameter Weibull distribution, because the lower limit in this distribution will not vary in the same way.

We suggest that a better approach is to employ the concepts termed “tolerance limit” [13] and “coverage” [13]. This approach yields the probability of failure for the entire population of parts.

The confidence interval for the measurements taken from the series of test specimens describes the boundaries on the mean of the specimen population, $\overline{S} \pm d_1$. For example, the 95% confidence limits describe the boundaries on the mean value 19 times out of 20. If the underlying population is at least approximately normally distributed, the formula for calculating $d_1$ is given by Nelson et al. [13]:

$$d_1 = t(0.025; n - 1)s \left( 1 + \frac{1}{n} \right)^{-\frac{1}{2}},$$

(13)

where $t$ is Student’s $t$ distribution, which is a function of the desired uncertainty (here 0.05) and the degrees of freedom $(n - 1)$; $s$ is the sd and $n$ is the number of specimens.
We need to be concerned only with the lower limit on strength; hence, the use of 0.025 in Eqn (13).

What is really needed is the uncertainty with which we know the minimum strength within the entire population of parts that might be produced. This is the tolerance limit. The tolerance limit, \( d_2 \), makes a statement about the strength limits of a given proportion of the entire population of parts. We can calculate this limit to predict the minimum time to failure of a specific proportion of the set of the full-scale structure or components. The tolerance limit, \( d_2 \), is calculated from:

\[
d_2 = \gamma s \mu
\]

Both \( \gamma \) and \( \mu \) are obtained from statistical tables [13] and depend on the degree of assurance, which we desire, that is, the coverage. Coverage is defined as the percentage of the total number of parts (or in other words the total area of all parts under stress) to which we want the tolerance limit to apply. This value of uncertainty will have much larger bounds than that for the specimen population alone. This approach is spelled out in more detail in the article by Fong et al. [21].

5. SUMMARY

This article presents a methodology for assuring the reliability of brittle ceramics in which environmentally enhanced crack growth can lead to failure after time under load. It is shown that times to failure are a function of five parameters, the initial strength of the material, \( S \), the strength of indented specimens, \( S_0 \), crack growth parameters, \( N' \) and \( \lambda \), and the stress in the part, \( \sigma \). The tests suggested to obtain the most accurate values of these parameters are described.

Note also that although Weibull distributions have typically been used to characterize strength data, this choice is not necessary for the accurate prediction of failure. While Weibull distributions have a basis in “weakest link theory” accepted as applicable to the failure of brittle materials, the choice of the particular distribution should be made on its “goodness of fit” to the experimental data.

We have described a straightforward approach to predicting failure times for brittle materials under stress, which are susceptible to slow crack growth. The methodology is general and would be applicable to estimations of reliability of parts in which other properties such as electrical or optical could degrade over time. What is needed is an expression relating a breakdown time to measurable parameters.

ACKNOWLEDGMENTS

The authors wish to thank James J. Filliben and N. Alan Heckert of the Statistical Engineering Division, NIST, for their technical assistance during the preparation of this article. They also wish to gratefully acknowledge the help of Howard Cohl of the Applied and Computational Mathematics Division, NIST, and Jack Mecholsky in reviewing the article.

Disclaimer: The views expressed in this article are strictly those of the authors and do not necessarily reflect those of their collaborators, their affiliated institutions, or National Institute of Standards and Technology (NIST). Any commercial product mentioned is for information only; it does not imply recommendation or endorsement by NIST nor does it imply that the products mentioned are necessarily the best available for the purpose.

REFERENCES


1. INTRODUCTION

Ceramic materials fail due to the extension of processing defects such as pores, inclusions, or cracks. Under rapid loading, unstable extension of these defects causes failure of the component. Under constant or cyclic loading, the defects may extend slowly until a critical size is reached, where unstable extension occurs. The subcritical extension of the defects under constant loading is called static fatigue and subcritical extension under cyclic loading is called cyclic fatigue. The defects can be described as cracks with a sharp tip and, therefore, the methods of fracture mechanics—in most cases of linear fracture mechanics—can be applied. This concept of sharp cracks has been applied successfully to describe the failure behavior of ceramics, even in cases where the flaws are three-dimensional, e.g., pores.

The basic relations of fracture mechanics are:

- A relation between strength $\sigma_c$ and flaw size $a$ in the form of
  $$ \sigma_c = \frac{K_{ic}}{\sqrt{aY}} $$

  where $K_{ic}$ is the fracture toughness of the material and $Y$ a constant depending on the shape of the crack.

- A relation between crack growth rate $\frac{da}{dt}$ and the stress intensity factor $K$ in the form of
  $$ \frac{da}{dt} = AK^n $$

  with the material parameters $A$ and $n$.

- A relation between the crack extension during one cycle and the range of the stress intensity factor $\Delta K$ in the form of
  $$ \frac{da}{dN} = A(\Delta K)^n $$

  with the material parameters $A$ and $n$, which are different from the values for static fatigue.

- A relation between crack growth rate in the creep range and the loading parameter $C^*$ in the form of
  $$ \frac{da}{dt} = A(C^*)^n $$

  with the material parameters $A$ and $n$.

In this chapter the basic principles of fracture mechanics are explained, the methods to determine the material parameters are presented, and details beyond the simple equations of $1-3$ are described.

2. LINEAR-ELASTIC FRACTURE MECHANICS

The basics of fracture mechanics are outlined for metals in several books [1–3]. For ceramic materials a detailed overview is given in [4]. Here, a comprehensive representation of the principles will be given.

Figure 1 shows a plate of thickness $B$ and width $W$ containing an edge crack of depth $a$. Three loading modes result in high stresses at the crack tip:

- mode I: tension normal to the crack plane, $\sigma_{yy}$
- mode II: shear loading in crack direction, $\tau_{xy}$
- mode III: ‘out-of-plane’ shear loading, $\tau_{yz}$.
Taking into consideration the singular stress term and the first regular term, the near-tip stress field can be described by

\[ \sigma_{ij} = \frac{K}{\sqrt{2\pi r}} f_{ij}(\varphi) + \sigma_{ij,0} \]  

where \( r \) and \( \varphi \) are polar coordinates with the origin at the crack tip (Figure 1). \( K \) is the stress intensity factor which depends on the applied load, on the size of the crack, and on the geometry of the component. The angular functions \( f \) for mode I are:

\[ f_{xx} = \cos \left( \frac{\varphi}{2} \right) \left[ 1 - \sin \left( \frac{\varphi}{2} \right) \sin \left( \frac{3\varphi}{2} \right) \right] \]  

\[ f_{yy} = \cos \left( \frac{\varphi}{2} \right) \left[ 1 + \sin \left( \frac{\varphi}{2} \right) \sin \left( \frac{3\varphi}{2} \right) \right] \]  

\[ f_{xy} = \cos \left( \frac{\varphi}{2} \right) \sin \left( \frac{\varphi}{2} \right) \cos \left( \frac{3\varphi}{2} \right) \]  

The first regular term in Eqn (5) represents the constant \( \sigma_{ij,0} \)-stress contribution appearing at the crack tip, which is called the T-stress:

\[ \sigma_{ij,0} = \begin{pmatrix} \sigma_{xx,0} & \sigma_{xy,0} \\ \sigma_{xy,0} & \sigma_{yy,0} \end{pmatrix} = \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix} \]  

For mode II:

\[ f_{xx} = \sin \left( \frac{\varphi}{2} \right) \left[ 2 + \sin \left( \frac{\varphi}{2} \right) \sin \left( \frac{3\varphi}{2} \right) \right] \]  

\[ f_{yy} = \sin \left( \frac{\varphi}{2} \right) \cos \left( \frac{\varphi}{2} \right) \cos \left( \frac{3\varphi}{2} \right) \]  

\[ f_{xy} = \cos \left( \frac{\varphi}{2} \right) \left[ 1 - \sin \left( \frac{\varphi}{2} \right) \cos \left( \frac{3\varphi}{2} \right) \right] \]  

In mode II and mode III the first regular stress term is zero. The stress component in thickness direction, \( \sigma_z \), is given by

\[ \sigma_z = \begin{cases} \nu(\sigma_y + \sigma_x) & \text{for plane strain} \\ 0 & \text{for plane stress} \end{cases} \]  

(\( \nu = \) Poisson’s ratio). Plane stress conditions are fulfilled at the specimen surfaces, i.e., near \( z = \pm B/2 \). In the center of thick specimens, plane strain conditions occur.

The stress intensity factor can be written as

\[ K_1 = \sigma \sqrt{a Y(a/W)}. \]  

In this relation \( \sigma \) is a characteristic stress in the uncracked body, e.g., the remote tensile stress in a tensile test or the outer fiber bending stress in a bending test. The geometric function \( Y \) depends on the crack size normalized to a characteristic dimension of the component. In the case of a curved crack front, e.g., a semi-elliptical surface crack, the stress intensity factor varies along the crack front.

It follows from Eqn (5) that the stress state at the crack tip is characterized by the stress intensity factor \( K_1 \). Consequently, the crack propagation behavior also depends on \( K_1 \).

If a component or a test specimen with a crack is loaded, \( K_1 \) increases with increasing load until unstable crack propagation occurs at a critical value of \( K_1 \). This critical value is the fracture toughness \( K_{ic} \). It is a material property and can be determined experimentally.

The description of crack propagation behavior with stress intensity factors is based on the consideration of stresses near the crack tip. An alternative is the consideration of energies. The crack resistance \( G_{ic} \) is introduced, which is the energy necessary to increase the crack area by unit area.

The specific surface energy \( \gamma \) is the energy necessary for the generation of a unit area of new crack surface. Since two crack surfaces are created during crack extension, \( G_{ic} = 2\gamma \) is obtained.

The energy for crack extension is provided by two sources, i.e., the work \( A \) done by the external forces and the elastically stored energy in the component \( U \). The energy
The energy release rate $G_I$ is the energy which is released in the case of a virtual increase of the crack by the unit area. Therefore,

$$G_I = \frac{dA}{dS} \frac{dU}{dS},$$

(12)

where $S$ is the crack area. The elastically stored energy $U$ enters with a negative sign, since a decrease of the elastically stored energy provides a positive energy contribution for crack extension. With increasing load $G_I$ increases, too. Crack extension is initiated if $G_I = G_{Ic}$ is reached.

Further analysis is illustrated in Figure 2. For a specimen with a crack of depth $a$ the load $F$ is plotted as a function of the displacements $V$ of the points, where the load is applied. A straight line will be found for linear material behavior. The ratio between displacement and applied load is called the compliance $C$:

$$C = \frac{V}{F},$$

(13)

which depends on the crack depth $a$. After a crack extension $\Delta a$, the compliance has increased. In Figure 2a the case of crack extension under constant total displacement is plotted. The load decreases by an amount $\Delta F$. The shaded area corresponds to the elastically stored energy released during crack propagation. The work done by the externally applied forces is zero, because the total load point displacement is kept constant. In Figure 2b the case of crack growth under constant load $F$ is illustrated. The displacement increases by $\Delta V$. The work done by the external forces is

$$\Delta A = F \Delta V.$$  

(14)

The elastically stored energy in the specimen increases by the contribution $\Delta U$. Since $\Delta A > \Delta U$, energy remains available for crack extension.

In the following considerations, the energy release rate shall be computed for the general case (Figure 2c), where both the displacement and the load may change during crack extension.

The work done by the external load is given by

$$A = \int F dV$$

(15)

and, therefore,

$$\frac{dA}{dS} = \frac{dA}{dV} \frac{dV}{dS} = F \frac{d(CF)}{dS} = F^2 \frac{dC}{dS} + FC \frac{dF}{dS}$$

(16)

is derived. The elastically stored energy is

$$U = \frac{1}{2} FV = \frac{1}{2} F^2 C$$

(17)

and

$$\frac{dU}{dS} = FC \frac{dF}{dS} + \frac{1}{2} F^2 \frac{dC}{dS}.$$  

(18)

From Eqs (12), (17) and (18) it follows that

$$G_I = \frac{F^2 dC}{2}.$$  

(19a)

For a plate of thickness $B$ with a through-the-thickness crack, Eqn (19a) can be replaced by

$$G_I = \frac{F^2 dC}{2B da} = \frac{F^2}{2BW} \frac{dC}{d(a/W)}.$$  

(19b)

If the compliance is known as a function of the crack area, the energy release rate for the cracked component can be calculated with Eqn (19).

The loading situation may alternatively be characterized by the stress intensity factor as well as by the energy

FIGURE 2 Load—displacement curve under crack extension; (a) constant displacement conditions, (b) constant load conditions, (c) general case with change in load and displacement.
release rate. Irwin has shown that a simple relation exists between the two loading quantities, namely,

$$K_1^2 = G_1 E'$$

(20)

with

$$E' = \begin{cases} E/(1 - \nu^2) & \text{for plane stress} \\ E/((1 - \nu^2)^2) & \text{for plane strain.} \end{cases}$$

(21)

Unstable crack propagation occurs, if

$$K_1 = K_{lc}$$

(22)

or if

$$G_1 = G_{lc}.$$  

(23)

Equation (20) also holds for the failure condition. Consequently, we find for plane stress

$$K_{lc}^2 = G_{lc} E' = \frac{2\gamma E}{1 - \nu^2}.$$  

(24)

3. METHODS FOR THE DETERMINATION OF FRACTURE TOUGHNESS $K_{lc}$

The principal procedure of fracture toughness measurements consists in the generation of a crack in a test specimen, the measurement of load at failure, and the calculation of $K_{lc}$ from failure load or failure stress, respectively, and from crack depth using the relations

$$K_{lc} = \sigma \sqrt{a} Y \quad \text{or} \quad K_{lc} = \frac{F}{B \sqrt{W}} Y^*.$$  

(25)

with the crack length included in the geometric function $Y^*(a/W)$.

3.1. The Edge-Cracked Bending Bar

The specimens commonly have a thickness of $B = 3 \text{ mm}$ and height of $W = 4 \text{ mm}$ and are usually loaded in a four-point bending test arrangement (Figure 3). The inner roller span is mostly $S_2 = 20 \text{ mm}$, the outer one $S_1 = 40 \text{ mm}$.

Measurements of fracture toughness for ceramic materials are often performed with specimens containing narrow saw cuts. The advantage of edge-notched bars consists of the relatively simple introduction of the notch. For a fine-grained ceramic, a very narrow notch is necessary.

A procedure proposed by Nishida et al. [5] allows introducing notches with notch root radii in the range of $1-10 \mu m$. In the first step, a conventional saw cut is introduced. Then, a sharp V-shaped notch is produced by using a razor blade and diamond paste (Figure 4). In extensive investigations, the usefulness of this method was confirmed [6].

The fracture toughness can be computed from the maximum load $F_{\max}$ and the relative crack length $\alpha = a/W$:

$$K_{lc} = \frac{F_{\max} S_1 - S_2}{B \sqrt{W}} \frac{3 \Gamma_M \sqrt{\alpha}}{2(1 - \alpha)^{3/2}}$$

(26)

with [7]

$$\Gamma_M = 1.1215 \sqrt{\pi} \left[ \frac{5}{8} - \frac{5}{12} \alpha + \frac{1}{8} \alpha^2 + 5\alpha^2(1 - \alpha)^6 \right.$$

$$\left. + \frac{3}{8} \exp\left(-6.1342 \frac{\alpha}{1 - \alpha}\right) \right].$$  

(27)

Sometimes, also three-point bending tests are used to determine $K_{lc}$. This test may be of special interest, if only a small number of specimens are available. Then, the fragments remaining from four-point bending tests can be used again. In [7] a stress intensity factor solution is given, which holds for all $S_1/W$ ratios in the range of $S_1/W > 2$. For the special value of $S_1/W = 2.5$ (obtained, e.g., with $S_1 = 10 \text{ mm}$, $W = 4 \text{ mm}$) Eqn (26), with $S_2 = 0$, becomes

$$\Gamma_M = \sqrt{\pi} \left[ 0.3738\alpha + (1 - \alpha) \left(1.044 - 2.116\alpha \right. \right.$$

$$\left. + 3.733\alpha^2 - 3.147\alpha^3 + 0.935\alpha^4 \right].$$  

(28)

Experimental investigations on the influence of the notch width and the notch root radius $\rho$ showed the dependency given schematically in Figure 5a. Beyond a critical notch root radius $\rho_c$, the measured nominal $K_{lc}$-value nearly increases with $\rho^{1/2}$. Only for $\rho < \rho_c$ will a correct $K_{lc}$-value be measured. The critical value $\rho_c$ depends on the tested material. In any case, it has to be ensured that the saw cut is narrow enough. In order to obtain real toughness data from the results measured with a finite notch radius, an extrapolation technique may be applied. The basis for such consideration is a stress intensity factor solution for small cracks in front of a narrow notch.

Figure 6 illustrates a small crack of length $\ell$ ahead of a circular notch of radius $\rho$ and length $a_0$. For an edge crack along the notch root a simple relation for the stress intensity factor $K$ was derived based on the boundary collocation procedure [7]:

$$K/K^* = \tanh \left[ \beta \sqrt{\ell/\rho} \right], \quad \beta \equiv 2.24.$$  

(29)
In this relation $K^*$ is the stress intensity factor of an edge crack with the total length $a = a_0 + \ell$. This relation is applied for the evaluation of fracture toughness of ceramic materials with the natural defects at the notch root modeled by straight through-the-thickness cracks. In practical applications, crack-like defects in ceramics are described well by semi-circular or semi-elliptical cracks, but in general, the exact geometry and size of the small crack ahead of the notch is unknown.

Damani et al. [8] assumed $\ell$ to be related to micro-structural features or machining damage, e.g., grain size, pores, or scratches. Damani et al. [8] also proposed the use of the property $\beta\sqrt{\ell}$ in Eqn (29) as a free fit parameter. In this way, it can be taken into consideration that real cracks in front of a notch have a shape in between an edge crack and a semi-elliptic crack. As an example, Figure 5b shows results from [9] for a silicon nitride. From these data a fit yields $K_{ic} \approx 5.2$ MPa$\cdot\sqrt{m}$. 

FIGURE 4 V-notch introduced into a bending bar (Si$_3$N$_4$) [6], right: Detailed view of the notch root.

FIGURE 5 Influence of the notch root radius on the measured fracture toughness, left: Schematic representation, right: Results on silicon nitride [6,9].

FIGURE 6 Crack in front of a narrow notch.
3.2. Specimens with Chevron Notches

The most popular specimen with a chevron notch is shown in Figure 7. By two saw cuts, a triangularly shaped cross section is produced. During loading, a sharp crack develops from the tip of the notch and the length of the crack front increases continuously. The geometrical data of the specimen and crack are:

- specimen thickness $B$,
- specimen width $W$, and
- notch parameters $a_0$ and $a_1$ ($a_0 = a_0/W$, $a_1 = a_1/W$).

Between the length of the crack front $b$ and the crack depth $a$ the relation

$$b = B \frac{a - a_0}{a_1 - a_0} = B \frac{a - a_0}{a_1 - a_0}$$

holds with $a = a/W$.

The normally obtained load–displacement curve is illustrated in Figure 8. The record is curved and exhibits a maximum. The nonlinearity is caused by crack propagation and a related increase in compliance. At maximum load, the crack has propagated (starting from $a = a_0$ to $a = a_{\text{max}}$).

If $C_{\text{Ch}}$ is the compliance of the chevron specimen, the available energy under load $F$ is

$$\Delta P = \frac{F^2}{2W} \frac{dC_{\text{Ch}}}{d\alpha} \Delta a.$$  \hspace{1cm} (31)

Crack propagation is initiated, if $\Delta P$ equals the energy required for crack extension, namely $G_{\text{lc}} b \Delta a$. This finally provides

$$K_{\text{lc}} = \frac{F}{B\sqrt{W}} \left[ \frac{EB}{2(1 - \nu^2)} \frac{\alpha_1 - \alpha_0}{\alpha - \alpha_0} \frac{dC_{\text{Ch}}}{d\alpha} \right]^{1/2} = \frac{F}{B\sqrt{W}} Y^*.$$  \hspace{1cm} (32)

The geometric function $Y^*$ contains two terms which change with $\alpha$. The derivative of the compliance with respect to $\alpha$ increases with $\alpha$ and the geometry term $(\alpha_1 - \alpha_0)/(\alpha - \alpha_0)$ decreases with increasing $\alpha$. If during crack extension the material resistance is constant ($K_1 = K_{\text{lc}} = \text{const}$), we must also have $FY^* = \text{const}$. The minimum $Y_{\text{m}}^*$ of $Y^*$ corresponds to the maximum load $F_{\text{max}}$, and the fracture toughness can be expressed as

$$K_{\text{lc}} = \frac{F_{\text{max}}}{B\sqrt{W}} Y_{\text{m}}^*.$$  \hspace{1cm} (33)

In order to allow $K_{\text{lc}}$ to be determined, the geometric function $Y_{\text{m}}^*$ has to be known for different specimen geometries ($W$, $B$) and notch geometries ($a_0$, $a_1$). From compliance measurements and finite element computations, the following relation was determined for the four-point bending specimen [10]:

$$Y_{\text{m}}^* = \left( 3.08 + 5.00a_0 + 8.33a_0^2 \right) \left\{ 1 + 0.007 \left( \frac{S_1 S_2}{W^2} \right) \frac{\alpha_1 - \alpha_0}{1 - \alpha_0} \frac{S_1 - S_2}{W} \right\}$$

where $S_1 = \text{outer roller distance}$, $S_2 = \text{inner roller distance}$, valid for $0.12 \leq \alpha_0 \leq 0.24$, $0.9 \leq \alpha_1 \leq 1$.

The main advantage of specimens with chevron notches is the fact that no sharp precrack has to be introduced. The specimen produces the sharp crack during loading. Moreover, the evaluation of fracture toughness does not need the actual crack length, since (33) only contains the maximum load $F_{\text{max}}$, and the geometric function $Y_{\text{m}}^*$. Further test specimens with chevron notches (short rod, short bar) are addressed in [4].

In the case of a material with a rising crack resistance curve exhibiting a so-called R-curve behavior (see Section 6), the resistance against crack extension increases with increasing crack extension. Consequently, the product of $F$ and $Y^*$ is no longer constant, but increases with crack extension. The maximum load $F_{\text{max}}$ does not occur exactly at the minimum $Y_{\text{m}}^*$ value of the geometric function $Y^*$. Nevertheless, it can be shown that application of Eqn (33) gives a toughness value which lies on the rising R-curve [11]. The crack extension at $F_{\text{max}}$, however, depends on the specimen size. This leads to the influence of the specimen size on the measured $K_{\text{lc}}$.

For an idealized notch with a disappearing notch root radius crack extension would immediately start after loading. Due to the finite notch width, crack propagation starts at a finite initial load $F_1$. Curve (a) in Figure 8 shows the desirable case $F_1 < F_{\text{max}}$. Curve (b) illustrates the case $F_1 > F_{\text{max}}$ where a sudden decrease in load indicates a limited amount of unstable crack extension. Then, the load increases again and passes the maximum $F_{\text{max}}$. In case of curve (c) in Figure 8, crack arrest is not reached and the maximum load cannot be measured. Consequently, fracture toughness cannot be determined in such a case.

**FIGURE 7** Four-point bending specimen with chevron notch.
3.3. Specimen with Knoop Indentation

A very simple way to produce a sharp crack is the hardness indentation test with a diamond pyramid. In case of a Knoop pyramid, a nearly semicircular surface crack develops under load. During unloading, the damaged region closes.

In Figure 9a the characteristic dimensions of a Knoop indentation are represented. The hardness impression on the surface has a length $L$ and a width $b$. The crack profile which becomes visible on the fracture surface (see Figure 9b for a soft PZT ceramic) is characterized by depth $a$ and width $2c$. The depth of the impression is $x$ and the depth of the damaged zone $z_0$. For some materials this quantity can be detected on the fracture surface. It was found that $z_0$ and $b$ are nearly identical [12]. After unloading, the damaged region is under compressive stresses which are balanced by tensile stresses near the crack tip region.

Commonly, the surface cracks are introduced into bending bars which are then broken in a four-point bending test. The computation of the fracture toughness $K_{IC}$ from the maximum load can be performed using the equation given by Newman and Raju [13].

The relation of Newman and Raju is appropriate for aspect ratios of $a/c < 1$:

- at the deepest point:
  \[ K_1 = \sigma \sqrt{\pi a} \frac{M}{\Phi} H_2; \]  
  \[ (35) \]

- at the specimen surface:
  \[ K_1 = \sigma \sqrt{\pi a} \frac{M}{\Phi} H_1 \sqrt{a/c(1 + 0.35a^2)} \]  
  \[ (36) \]

with the outer fiber bending stress $\sigma$

\[ \sigma = \frac{3(S_1 - S_2)F}{2W^2B} \]  
\[ (37) \]

and the quantities

\[ H_1 = 1 - [0.34 + 0.11(a/c)]a \]  
\[ (38) \]

\[ H_2 = 1 - [1.22 + 0.12(a/c)]a + \left[ 0.55 - 1.05(a/c)^{0.75} + 0.47(a/c)^{1.5} \right] a^2 \]  
\[ + 0.47(a/c)^{1.5} \]  
\[ (39) \]

\[ M = 1.13 - 0.09 \frac{a}{c} + \left[ -0.54 + \frac{0.89}{0.2 + a/c} \right] a^2 \]  
\[ + \left[ 0.5 - \frac{1}{0.65 + a/c} + 14(1 - a/c)^{24} \right] a^4. \]  
\[ (40) \]

$\Phi$ is the complete elliptical integral of the second kind, which can be approximated by

\[ \Phi \approx \left[ 1 + 1.464(a/c)^{1.65} \right]^{1/2}. \]  
\[ (41) \]

Fracture will be caused at the point with the maximum stress intensity factor, i.e. for cracks with $a/c > 0.8$ at the surface point B.

A problem in the correct evaluation of $K_{IC}$ is the removal of the residual stresses caused by the damaged zone directly under the Knoop indenter. The residual stresses are responsible for a positive residual stress intensity factor $K_{res}$ which is superimposed on the externally applied stress intensity factor and leads to a lower failure stress in the bending test. The residual stress intensity factor may be
reduced during load application, but will not disappear completely before failure. Since $K_{lc}$ is computed with the maximum externally applied load, the resulting value will be too small. To determine correct maximum externally applied load, the resulting value will not be removed before the bending test is carried out. This can be done by surface grinding or by annealing.

A more indirect method of toughness determination is based on crack evaluation during a Vickers indentation test. The relevant references for this test are reported in [4].

4. MODE-II AND MIXED-MODE FRACTURE TOUGHNESS

4.1. Determination of Toughness

Besides the mode-I fracture behavior, fracture under mode II and mixed-mode loading also are of interest. The results are plotted in a mixed-mode $K_{I}-K_{II}$ diagram. Three types of test specimens will be described in the following subsections.

Whereas cracks orthogonal to the specimen surfaces are applied in $K_{lc}$ determination, oblique edge cracks (or notches) are appropriate for mixed-mode fracture.

Figure 10a illustrates the geometrical data of a four-point bending test and Figure 11a represents the related geometric functions $Y_I$ and $Y_{II}$ for several values of $a/W$ and angles $\beta$ (see also [7]). $Y_I$ and $Y_{II}$ are defined by

$$K_I = \sigma Y_I \sqrt{a}, \quad K_{II} = \sigma Y_{II} \sqrt{a}, \quad \sigma = \frac{3S_1 - S_2}{2BW^2} F. \quad (42)$$

While the curves for mode-I stress intensity factors depend significantly on the relative crack depth $a/W$, the mode-II contributions are nearly independent of $a/W$.

For more details see [4].

Figure 10a illustrates the geometrical data of a four-point bending test and Figure 11a represents the related geometric functions $Y_I$ and $Y_{II}$ for several values of $a/W$ and angles $\beta$. As an example, the normalized geometric functions $Y'_I = Y_{II}(1 - a/W)^{1/2}$ and $Y'_I = Y_I (1 - a/W)^{3/2}$ are plotted in Figure 11b.

For the determination of the pure mode-II fracture toughness $K_{IIc}$ with edge-notched (or precracked) beams, asymmetric bending arrangements (Figure 12) are recommended [14–16], which show a negligible mode-I contribution. The geometric functions $Y_I$ and $Y_{II}$, defined by

$$K_I = \frac{F}{BW} \left(1 - \frac{d}{L}\right) Y_I \sqrt{a}, \quad K_{II} = \frac{F}{BW} \left(1 - \frac{d}{L}\right) Y_{II} \sqrt{a},$$

are also shown in Figure 12.

The geometric function $Y_{II}$ is nearly independent of the value of $d/W$ in the range of $0.375 < d/W < 0.625$. An analytical approximation can be given by

$$Y_{II} = 6.949\alpha - 9.092\alpha^2 + 25.624\alpha^3 - 46.601\alpha^4 + 30.261\alpha^5, \quad \alpha = a/W. \quad (44)$$

Mixed-mode studies have often been performed with controlled surface flaws in bending bars [17–19]. A single Knoop surface flaw is placed on the tensile surface at an orientation angle $\Theta$ with respect to the specimen length axis (Figure 13). The externally applied uniaxial stress $\sigma^*$ gives rise to a stress $\sigma_n$ normal to the crack and a shear stress $\tau$ in the crack plane:

$$\sigma_n(x) = \sigma^* \sin^2 \Theta, \quad \tau(x) = \sigma^* \sin \Theta \cos \Theta. \quad (45)$$

In a bending test $\sigma^*$ decreases linearly with the distance from the surface. In [17–19] the Knoop crack was assumed to be a semicircular crack, which is nearly in agreement with experimental results, and the decrease of $\sigma^*$ was neglected ($a/W < < 1$).

The stress intensity factors used were for a crack in an infinite body, where surface effects are ignored:

$$K_I = 2\sqrt{a/\pi} \sigma_n \quad (46)$$

$$K_{II} = \frac{4}{2 - \nu} \sqrt{a/\pi \tau \cos \gamma} \quad (47)$$

**FIGURE 10** (a) Four-point bending test with an oblique notch, (b) three-point bending test with an eccentric notch.
with the angle $\gamma$ defined in Figure 13b. The advantage of this test method is the existence of a sharp crack. Nevertheless, there are some problems which aggravate the analysis. It has to be ensured that all residual stresses generated during the indentation test have been removed by annealing or surface grinding. Since the mode-II and mode-III contributions vary along the crack-tip line, it seems impossible to determine without doubt that particular mixed-mode stress intensity factor combination which was responsible for failure.

4.2. Effective Stress Intensity Factors

Under mixed-mode loading, failure of a component containing a crack occurs, if a combination of the stress intensity factors and the effective stress intensity factor $K_{\text{eff}}$ equals the fracture toughness $K_{\text{Ic}}$. The most popular failure criterion is that of the coplanar energy release rate [20]. Under plane strain conditions, it reads

$$K_{\text{eff}} = \sqrt{K_I^2 + K_{II}^2 + \frac{1}{1-\nu}K_{III}^2} = K_{IC} \quad (49)$$

Such a criterion makes sense only for positive $K_I$. For $K_I < 0$ the crack faces are under compression and no singular mode-I stress field exists. Nevertheless, a mode-II stress intensity factor can occur, caused by the superimposed shear loading. Due to friction between the crack faces, $K_{II}$ has to be calculated with an effective shear stress [21].
\( K_{II, eff} = \tau_{eff} Y \sqrt{a}, \tau_{eff} = \left\{ \begin{array}{ll} \tau & \text{for } |\tau| + \mu |\sigma_n| > |\tau| \\ |\tau| & \text{for } |\mu \sigma_n| < |\tau| \\ 0 & \text{for } |\mu \sigma_n| > |\tau| \end{array} \right. \)  

(50)

with the geometric function \( Y \) for constant shear stresses. Equation (50) is only correct for cracks with completely closed crack faces and constant stresses over the crack size. This is often the case, but in case of strongly varying stresses (occurring e.g., near point or line contacts), this condition is not fulfilled. As outlined in [22] for edge cracks, the condition

\[ K_{II, eff} = K_{II, appl} + \mu K_{I, appl} \]  

(51)

(where subscripts “appl” designate the applied stress intensity factors caused by the stresses in the uncracked body at the location of the crack) remains exactly valid for small (natural) cracks fully or partially closed, if \( a/W \rightarrow 0 \).

5. FRACTURE OF PIEZOELECTRIC CERAMICS

When applied as actuators, piezoelectric ceramics are subjected to electrical and mechanical loads. Both types of loading will influence the failure behavior. Whereas the mechanical loading results in the stress intensity factors \( K_I, K_{II}, \) and \( K_{III} \), another (electric) stress intensity factor \( K_{IV} \) appears in cracked piezoelectric components due to the electric field [23–26].

Two limit cases have to be distinguished, namely, an electrically conducting and an insulating crack. The conducting crack is assumed to be filled with a medium of high permittivity \( \varepsilon_r \rightarrow \infty \) (or has thin conducting layers on its surface). The insulating crack is filled with an idealized medium \( \varepsilon_r \rightarrow 0 \). Mechanical stiffnesses are supposed to disappear in both cases.

An insulated crack in an infinite body may be considered. With the coordinate system shown in Figure 1 the dielectric displacements near the crack tip are given by

\[ \begin{aligned} D_x &= -\frac{K_{IV}}{2\pi \epsilon_0} \sin\left(\phi/2\right) \\ D_y &= \frac{K_{IV}}{2\pi \epsilon_0} \cos\left(\phi/2\right) \end{aligned} \]  

(52)

i.e. the same singularity occurs as in case of mechanical stresses. The electric field intensity factor \( K_{IV} \) is defined as

\[ K_{IV} = D_y \sqrt{\pi a} \]  

(53)

with the electric displacements in \( y \)-direction far away from the crack, \( D_y \). The energy release rate, related to \( K_{IV} \), is

\[ G_{IV} = \frac{1}{2\epsilon_0 \epsilon_r} K_{IV}^2 \]  

(54)

with \( \epsilon_0 = 8.86 \times 10^{-12} \text{ As/Vm} \).

Two specific experimental results may be addressed. An insulated crack in a CT specimen was investigated by Park and Sun [27]. As shown in Figure 14a, the polarization \( P \), the applied electric field \( E = U/H \), and the mechanical loading were perpendicular to the crack plane. The critical stress intensity factor \( K_{IC} \) is plotted in Figure 14b as a function of the electric field. For positive field \( E \) a decrease in \( K_{IC} \) and for negative field an increase of \( K_{IC} \) was found.

The failure criterion for conducting cracks was studied by Heyer et al. [28,29] in bending tests. In this case the polarization was oriented in the crack plane direction. The electric boundary conditions are illustrated in Figure 15a. Figure 15b represents the results where the critical mechanical mode-I stress intensity factor \( K_{IC} \) is plotted versus the electric intensity factor \( K_E \), describing the singularity of the electric field ahead of the crack tip, and defined by the electric field in the crack plane \( E(x-a) \) as

\[ E = \frac{K_E}{\sqrt{2\pi(x-a)}} \]  

(55)

FIGURE 14 Influence of an electric field on the fracture toughness of a piezoelectric ceramic as reported by Park and Sun [27], (a) test device, (b) fracture toughness (normalized to toughness in the absence of an electric field).
The symbols in Figure 15b cover the range of crack length of $0.9 \leq a \leq 2.2$ mm.

### 6. R-CURVE BEHAVIOR

In the case of ideally brittle materials, the fracture toughness is independent of the crack extension $\Delta a$. In this case, a so-called flat crack resistance curve (R-curve) is obtained. In many ceramics a different behavior is observed. The crack growth resistance increases with increasing crack extension. The necessary energy for unit area crack extension — now called $G_{IR}$ — increases from an initial value $G_{I0}$. The stress intensity factor $K_{IR}$ increases from $K_{I0}$. The crack propagation behavior is no longer described by a single value $K_{Ic}$, but now characterized by the $K_{IR}$-$\Delta a$ curve. In a first approximation it was assumed that the crack growth resistance curve is a pure material property, i.e. independent of the specimen geometry, the crack size, and the loading (tension, bending, etc.).

In the interpretation of the rising crack growth resistance it is assumed that during crack extension the stress intensity factor at the crack tip $K_{I0}$ is constant and identical to the value at the onset of crack propagation. The stress intensity factor applied during crack extension, $K_R$, is reduced by a shielding stress intensity factor $K_{sh}$ (see Figure 16):

$$K_R + K_{sh} = K_{I0}, \quad K_{sh} < 0$$  \hspace{1cm} (56)

For coarse-grained materials, the shielding term is predominantly caused by crack border interactions in the wake of the growing crack (see section 10). For materials such as zirconia undergoing phase transformation in front of a crack tip (mostly accompanied by a change in volume), the shielding stress intensity factor is caused by residual stresses which reduce the stresses at the tip. The same holds for domain switching in PZT materials, which occurs under constant volume. In all cases $K_{sh}$ is negative and increases with increasing crack extension.

The first observed increasing R-curve was published by Hübnner and Jillek [30] for alumina. Later, rising R-curves were found for many other materials. In [4] the authors presented results for a wide range of materials and covered several standard test methods. In this contribution some recent procedures and results are shown.

### 6.1. Experimental Results

Some experimental results are shown in Figure 17. The increase in the crack growth resistance with increasing grain size can be seen from Figure 17a for alumina. Figure 17b shows that the slope of the crack growth resistance curve is not independent of the initial crack size. Figure 17c reveals that large regions with a linear increase in crack growth resistance are possible. In Figure 17d an example is given, where after an increase in the crack growth resistance a decrease follows. Results for hot-pressed silicon nitrides are shown in Figures 17e and 17f. The R-curves are initially very steep, reaching saturation already at a crack extension of about 15 $\mu$m.

![FIGURE 15](image1.png) (a) Conducting crack under combined electrical and mechanical loading. (b) fracture criterion determined by Heyer et al. [28,29].

![FIGURE 16](image2.png) Schematic representation of an R-curve.
FIGURE 17  (a) Crack resistance curves for alumina with different grain sizes [31], (b) influence of the initial crack size on R-curves for alumina [4], (c) MgAl\textsubscript{2}O\textsubscript{4} spinel from DCB tests [32] (d) R-curves of several specimens of 9 mol\% Ce-TZP with a mean grain size of $d_m = 2.5\ \mu m$ [33], (e) R-curves for a hot-isostatically pressed nitride with 5 wt\% Y\textsubscript{2}O\textsubscript{3} and 2 wt\% MgO and a silicon nitride with 3 wt\% Y\textsubscript{2}O\textsubscript{3} and 3 wt\% Al\textsubscript{2}O\textsubscript{3} [34], (f) same as (e) for short crack extensions.
6.2. Methods for Determination of R-Curves

6.2.1. Controlled Bending Tests

For the experimental determination of crack resistance curves for specimens with macrocracks, the actual load \( F \) and the actual crack length \( a \) have to be determined. \( K_{IR} \) can be calculated according to

\[
K_{IR} = \sigma a Y(a/W) = \frac{F}{B\sqrt{W}} Y^*(a/W).
\]

with \( Y^* = Y\sqrt{a/W} \).

The crack length can be measured optically with a microscope. An alternative is to use the compliance

\[
C = \frac{V}{F},
\]

where \( V \) is the load point displacement.

In Figure 18a very rigid bending test device is shown [33]. R-curve experiments are carried out by three-point bending with a roller span of 40 mm. The load is recorded with a quartz load cell and the load point displacement monitored with an LVDT. The load rig is mounted horizontally on a microscope to measure the crack length, yielding an accuracy of 1 mm. This enables to compare the crack length obtained via compliance with the physical crack length visible on the surface.

Figure 18 also represents a load versus displacement curve together with partial unloadings, from which the actual compliance can be determined.

The relation between compliance and crack length can be determined experimentally for a given specimen geometry. It is also possible to calculate this relation from the function \( Y \) or \( Y^* \). From Eqn (19) and (20), it follows that

\[
C = C_0 + \frac{2}{BE'} \int_{0}^{a} Y^2(\alpha')\alpha' d\alpha',
\]

where \( Y \) and \( Y^* \) are the geometric functions for the stress intensity factor under the special loading situation and \( C_0 \) is the compliance of the specimen in the absence of a crack or notch. From the measured difference in compliance between an uncracked and a cracked specimen, \( \Delta C = C - C_0 \), \( \alpha = a/W \) can be determined.

For larger crack extensions the compliance method underestimates the real crack length. This can be explained by bridging effects. A partially bridged crack leads to a higher stiffness of the cracked component than a crack without bridging interactions. The increased stiffness in the presence of bridging effects results in a lower compliance and, consequently, in a shorter crack length.

6.2.2. Stable Fracture Test Under Opposite Roller Loading

A test for stable crack propagation over a large range of crack extension is illustrated in Figure 19 [35]. A pre-notched bar is loaded via four rollers by the load \( P_{appl} \) which is applied via a half-sphere to the upper rollers. As rollers, cylindrical pins (hardened steel) may be used. After a certain increase of load, the test is suspended. After unloading, the actual crack length can be measured at the side surfaces using an optical microscope. Additional application of a penetration dye allows measuring the crack depth on the fracture surface of a broken specimen. From the load \( P \) applied at the moment of unloading and the crack depth \( a \), the mode-I stress intensity factor is obtained as

\[
K_I = \sigma^* Y_1 \sqrt{W}, \quad \sigma^* = \frac{P_{appl}}{WB}.
\]
with the geometric function $Y_1$. The mode-I stress intensity factor is shown in Figure 20a as a function of $a/W$ with $d/W$. Figure 20b shows an R-curve obtained for an alumina ceramic.

In the special case of $d/W = 1$, the geometric function $Y_1$ is fitted for $a = a/W \leq 0.6$ by

$$Y_1 = 0.905\alpha^{1/2} - 3.358\alpha^{3/2} + 3.857\alpha^{5/2} + 1.4425\alpha^{7/2} - 3.873\alpha^{9/2}$$  \hspace{1cm} (61)

Knowledge of the stress intensity factor solution enables predicting crack growth under increasing load. For a material without an R-curve effect, the crack propagation phases are illustrated in Figure 21 for the special case of $d/W = 1$.

If the initial crack depth $a_0$ is smaller than the value $a_m$ corresponding to the maximum stress intensity factor (Figure 21a), crack extension is unstable up to $a_1$. Then, stable crack propagation takes place. For an initial crack depth $a_0 \geq a_m$, exclusively stable crack growth occurs (Figure 21b). In case of a sufficiently steep R-curve, stable crack extension appears also for $a_0 < a_m$ [35].

6.2.3. R-Curves from Strength Measurements on Knoop-Damaged Specimens

The R-curve in the range of short crack extension may also be determined from strength measurements on artificially damaged tensile or bending specimens [36]. Let us consider specimens with a Knoop crack of initial depth $a_0$ and width $2c_0$ under tension. During the increase in applied load, the crack starts to extend, if $K_A, K_B > K_{I0}$, where $K_A$ is the stress intensity factor at the deepest point and $K_B$ are the stress intensity factors at the surface points of the semi-ellipse. Catastrophic failure follows after a certain amount of stable crack growth. The failure condition in a strength test after stable crack growth

$$K_{IA} = K_{IR}(\Delta a), \quad K_{IB} = K_{IR}(\Delta c)$$ \hspace{1cm} (62a)

$$\max \left( \frac{\partial (K_{IA} - K_{IR}(\Delta a))}{\partial a}, \frac{\partial (K_{IB} - K_{IR}(\Delta c))}{\partial a} \right) \geq 0$$ \hspace{1cm} (62b)

In the following evaluation it is assumed that the R-curve is a function of $\Delta a$ and independent of $a_0$.

Rectangular specimens of a PZT ceramic with $3 \times 4 \times 45$ mm$^3$ were damaged by Knoop indentations with different indentation loads $P$. The initial crack depth $a_0$ is shown in Figure 22a as a function of the indentation load $P$. After annealing above the Curie temperature, tensile tests were performed with the results plotted in Figure 22b [36].

The stress intensity factors computed with the strength as the critical stress $\sigma = \sigma_c$ must provide an envelope for the R-curve. For a first rough computation the aspect ratio was assumed to remain constant ($a/c = a_0/c_0 = \text{const.} = 0.95$) during crack extension. Since the geometric functions $F_A$ and $F_B$ differ slightly for this aspect ratio, the average value

![FIGURE 20](image-url) (a) Geometric function $Y$ according to Eqn (61) for 4-point loading, (b) R-curve for alumina with 4% glass content.

![FIGURE 21](image-url) Crack growth under increasing load; (a) initial crack size $a_0 < a_m$ and (b) $a_0 = a_m$. 

\[ F = \frac{1}{2} (F_A + F_B) \] was used as an approximation. With these simplifications the applied stress intensity factors for several initial crack depths were computed from

\[ K_{\text{appl}} = \sigma_c F \sqrt{\pi a} \]  (63)

with \( \sigma_c \) taken from Figure 22b.

The resulting curves \( K_{\text{appl}} = f(\sigma_c, \Delta a) \) are entered in Figure 22c with the initial crack depth as the parameter. The R-curve resulting from the envelope condition is introduced as a thick line. A strong increase of \( K_{\text{IR}} \) with crack extension \( \Delta a \) is obvious.

For further computation steps the assumption of a constant aspect ratio was dropped. The resulting higher approximations for the R-curve were found to be nearly identical with the simply obtained first estimation [36].

6.3. Determination of \( K_{I0} \)

The increasing crack resistance (R-curve) behavior of ceramic materials is of high interest for technical applications. Numerous investigations are available in literature, in which the material behavior was studied for long crack extensions. In recent years, increasing attention was drawn to the starting point \( K_{I0} \) of an R-curve, also called “crack tip toughness” [37,38]. Several methods for the determination of \( K_{I0} \) are mentioned below.

6.3.1. Extrapolation of R-curves to \( \Delta a \to 0 \)

In the case of a linear increase in crack resistance, the \( K_{I0} \)-value can be determined easily by extrapolating to \( \Delta a = 0 \). For instance, this is possible for the MgAl\(_2\)O\(_4\) spinel from Figure 17c. This curve shows a pronounced initial straight line behavior which extends over about a 6 mm crack extension. We find for this range

\[ K_R = K_{I0} + C \Delta a, \quad C = -0.214 \text{ MPa}\sqrt{\text{m}}/\text{mm} \]  (64)

and the stress intensity factor at the onset of crack propagation \( K_{I0} = 1.45 \text{ MPa}\sqrt{\text{m}} \) (see Figure 17).

6.3.2. \( K_{I0} \) from Near-tip Crack Opening Displacements

A microscopic observation of the near-tip crack opening displacement during a stable crack growth test yields \( K_{I0} \) [37,39]. In [37] crack opening displacements in “round compact tension” (RCT) specimens (Figure 23a) were evaluated. Figure 23b shows the measured crack opening profile for a coarse-grained Al\(_2\)O\(_3\) versus the distance \( a-x \) from the crack tip. The near-tip COD data are related to the actual crack tip stress intensity factor \( K_{\text{tip}} \) by

\[ \delta_{\text{tip}} = \frac{\sqrt{8}}{\pi} \frac{K_{\text{tip}}}{E'} \sqrt{a-x} \]  (65)

The dashed curve in Figure 23b represents this dependency. The analysis of the near-tip data yields \( K_{I0} \approx 2.3 \text{ MPa}\sqrt{\text{m}} \).

6.3.3. \( K_{I0} \) from Load versus Displacement Curves

In principle, the initial stress intensity factor \( K_{I0} \) can also be determined from controlled bending fracture tests. Unfortunately, the often-used evaluation of the load versus displacement curve is very insensitive to small crack extensions. An improvement of the sensitivity of this procedure was proposed in [40]. Figure 24 represents the test arrangement. Two strain gauges (1) and (2) are placed at the bending bar in the compressive zone.

One of them is applied at a location remote of the notch (1) and the other one directly opposite the notch (2). Each strain gauge is the active arm of a quarter-bridge circuit. Under bending loading, the strain gauge (2) will produce a higher signal (U2) than the other sensor.
(U₁) due to the higher stresses in the notched cross section. Nevertheless, the ratio of the two signals U₂/U₁ is independent of the actual load (as long as linear-elastic material behavior is considered), but depends on changes in geometry.

Let us assume that the strains at location 2 are higher by a factor of \( \lambda > 1 \) than the strains at location 1, i.e. \( U₂/U₁ = \lambda \). In the two amplifiers I and II the strain gauge signals are differently amplified. The amplifications of amplifier I (\( k_I \)) and amplifier II (\( k_{II} \)) are conveniently chosen as

\[
\frac{k_I}{k_{II}} = \lambda
\]  

(66)

Consequently, the output signals of both amplifiers are identical. The difference of the single output signals is given on the input connectors of an additional amplifier III. For this purpose, the authors use the amplifier of an X–Y recorder. As long as linear material behavior is investigated, no signal is available at the output of amplifier III. The optimum ratio \( \lambda \) can be found experimentally by application of a small load and by adjusting the magnification of the amplifier until the output signal vanishes. If a crack starts to grow at a certain increased load, the geometry changes and the signal of strain gauge 2 increases more than that of strain gauge 1. The consequence is a signal at the output of amplifier III, which can be recorded.

### 6.4. Reasons for R-curve Behavior

The most important reasons of the R-curve behavior are mentioned in the following subsections.

#### 6.4.1. Bridging Effects

Coarse-grained materials show an R-curve effect which is caused by crack–border interactions in the wake of an advancing crack. This fact has been proved experimentally by re-notching experiments [41,42] and by in-situ observations under the electron microscope [43,44].

In Figure 25a such a bridging interaction influenced by a friction force \( F \) is illustrated schematically. The crack face interactions localized at single grains can transfer loads which can be modeled in a more homogeneous way by so-called bridging stresses \( \sigma_{br} \) that depend on the crack opening displacement \( \delta \). The bridging stresses shield the crack tip from the external loads as illustrated schematically in Figure 25b.

For bridging stresses which are caused by friction effects, Mai and Lawn [45] proposed a relation

\[
\sigma_{br} = \begin{cases} 
\sigma_o(1 - \delta/\delta_c) & \text{for } \delta/\delta_c \leq 1 \\
0 & \text{for } \delta/\delta_c > 1 
\end{cases}
\]  

(67a)

with a characteristic displacement \( \delta_c \) at which the bridging stress disappears. If the critical displacement
\( \delta_c \) is \( \Gamma \)-distributed, an averaged bridging relation is obtained [46]:

\[
\sigma_{br} = \sigma_0 \exp\left( -\frac{\delta}{\delta_0} \right). \tag{68a}
\]

with a material-specific displacement \( \delta_0 \).

In the case of unbroken microstructural elements, the bridging stress may increase up to a critical displacement \( \delta_c \), leading to a relation of

\[
\sigma_{br} = \frac{\sigma_0}{\delta_0} \delta_c \exp\left( -\frac{\delta}{\delta_0} \right) \quad \text{for} \quad \frac{\delta}{\delta_0} \leq 1.
\]

(67b)

If again the critical displacement is \( \Gamma \)-distributed, an averaged bridging relation is obtained:

\[
\sigma_{br} = \sigma_0 \frac{\delta}{\delta_0} \exp\left( -\frac{\delta}{\delta_0} \right). \tag{68b}
\]

The two bridging laws, Eqns (68a) and (68b), are illustrated in Figure 26.

The bridging stress intensity factor results from the distribution of bridging stresses along the crack, \( \sigma_{br}(x) \) in Figure 25b, as

\[
K_{br} = \int_0^a h(x, a) \sigma_{br}(x) \, dx. \tag{69}
\]

6.4.2. Phase Transformation Effects in Zirconia Ceramics

Due to the singular stress field near a crack tip in transformation-toughened zirconia, the material undergoes a stress-induced martensitic transformation and the tetragonal phase changes to the monoclinic phase (t- to m-ZrO\(_2\)). This transformation occurs when the local stresses reach a critical value, and the result is a crack tip transformation zone.

In this section the stress intensity factor analysis according to McMeeking and Evans [47] will be addressed briefly. Their procedure is based on an Eshelby technique [48]. Let us assume the transformation zone ahead of the crack (see Figure 27) to be removed from the original material. The transformed material is assumed to be restored to the original shape of the non-transformed material by applying surface traction \( T \) to the contour of the transformation zone. The zone now fits again into the body and equilibrium can be obtained by nullifying the surface traction with a layer of body forces. The surface traction then result in a residual stress intensity factor \( K_{res} \):

\[
K_{res} = \int_{\Gamma} T \cdot h \, dS,
\]

where \( \Gamma \) is the contour line of the transformation zone and \( dS \) is a line length increment. The vector \( h \) represents the weight function \( h = (h_x, h_y, h_z)^T \) with the components \( h_x \) and \( h_y \).

In an early study McMeeking and Evans [47] investigated the case that transformation was initiated when the hydrostatic tensile stress reached a characteristic value \( \sigma_{hyd,c} \). For the special case of small-scale transformation conditions (transformation zone size negligible compared
to crack size and component dimensions), McMeeking and Evans [47] computed the transformation zone, neglecting the perturbation of the stress field due to transformations.

The phase transformation zone for plane strain becomes

$$\omega = \frac{(1 + \nu)^2}{4\sqrt{3\pi}} \frac{K_1}{\sigma_{\text{hyd},c}}^2, \quad r = \frac{8}{3\sqrt{3}} \omega \cos^2(\theta/2). \quad (71)$$

Figure 27 illustrates the transformation zone for a non-extending crack (a) and after a crack extension of $\Delta a$ (b). Caused by the martensitic transformation, a volumetric expansion strain of about 4.5% occurs. These strains cause tensile stresses at some distance ahead of the crack tip and compressive stresses along the length $\Delta a$ at the crack line. The compressive stresses lead to a shielding stress intensity factor which has to be overcome during crack propagation, i.e. the applied stress intensity factor must be increased to maintain stable crack growth.

In the special case of a pure dilatational transformation the shielding stress intensity factor tends asymptotically to $K/p \omega^{1/2} = 0.22$, where $p$ is the pressure necessary to restore the untransformed zone shape.

In much more complicated numerical studies (e.g., [49]), the influence of shear stresses and strains on the transformation criterion and on the zone calculation was also taken into consideration.

Compared to the nearly circular phase transformation zones which are found in MgO-doped zirconia, the zone shape of Ce-partially stabilized zirconia differs strongly. The left part of Figure 28 illustrates a typical elongated transformation zone for a zirconia (Ce-TZP) alumina composite as reported by Yu et al. [50]. A typical zone for a 9 mol% Ce-TZP with a mean grain size of 1.4 $\mu$m is shown in the right part of Figure 28, as reported by Rauchs et al. [33]. In this case, an irregular zone contour is visible.

### 6.4.3. Domain Switching in Piezoelectric Ceramics

Piezoelectric ceramics such as barium titanate or lead zirconate/titanate zirconate (PZT) are paraelectric above a critical temperature, the Curie temperature, and piezoelectric below this temperature. If during cooling the Curie temperature is reached, a spontaneous phase transformation occurs. In the case of BaTiO$_3$, this is a transformation from the cubic to the tetragonal phase. The cubic lattice parameter is $a = 4.009 \text{ Å}$; in the tetragonal state the ratio of the lengths of the $c$-axis to the $a$-axis is $c/a = 1.01$ at room temperature. Phase transformation is associated with an electrical polarization in the direction of the $c$-axis. In each grain six polarization directions are possible. The strains are subdivided into domains, where each domain has a direction of polarization. The domains are separated by domain boundaries. The overall polarization of a polycrystal after cooling below the Curie temperature is zero. Applying an electrical field leads to domain switching in the direction of the electrical field and, therefore, to a polarization of the sample. Domain switching can also be caused by mechanical stress. Applying a compressive stress in the direction of polarization leads to a switching of the domains in a direction perpendicular to the stress and, therefore, to a contraction of the specimen. Applying a tensile stress perpendicular to the orientation of polarization can lead to domain switching in the direction of the applied stress and to an elongation of the specimen. In an unpolarized specimen domain switching is also possible for those domains which are in a suitable orientation. Therefore ‘plastic’ deformation is possible in polarized and unpolarized materials.

A simple explanation of the toughening effect may be given by the influence of domain orientation ahead of the tips of cracks [51]. Let us consider a volume element at some distance from the tip of a natural crack. In an unpoled tetragonal material we will find 1/3 of all domains
orientated in the x-direction, 1/3 in the y-direction, and 1/3 in the z-direction (Figure 29a). If a mechanical load is applied normally to the crack plane (strongly magnified near the crack tip), most of the domains will switch into the stress direction (Figure 29b) with antiparallel orientation sense. The increase of domains in the y-direction causes a residual stress field which reduces the influence of the externally applied stress, interpreted as an increase of resistance against crack propagation [52,53].

In the case of PZT poled perpendicularly to the stress direction, nearly all domains may be orientated in the x-direction (Figure 29c). Under externally applied tensile stresses, an orientation of the domains in the y-direction can also be observed, similar to the orientation of domains in the unpoled material (Figure 29d). Whereas in the case of unpoled material 2/3 of all domains can switch into the y-direction, all domains are able to switch in the case of ⊥-poled PZT. From this point of view, an increased domain switching and increased crack resistance have to be expected. Due to domain switching from the poled state, depolarization ΔP_x takes place. This depolarization is strongly concentrated in the vicinity of the crack tip. The highly stressed and depoled zone is in series with the nearly unstressed bulk material which remains in the poled state. Therefore, the short-circuited electrodes — far away from the depoled zone — are without effect and the electric boundary conditions are given by a constant dielectric displacement, D = constant. Due to depoling, an electrical field ΔE_x is generated

\[ D_x = εE_x + P_x = \text{const} \implies ΔE_x = -ΔP_x/ε \]

(ε = dielectric permittivity), which inhibits depoling by domain switching (Figure 29d), i.e. switching in the poled state needs a higher stress to reach the same degree of domains orientated in the y-direction.

If the PZT is poled in the y-direction, the number of domains in the x- and z-directions able to switch into the y-direction is very small. Therefore, we do not expect any effect of reorientation (Figure 30a and 30b). Hence, the lowest crack resistance is expected.

The case of a crack with the polarization being parallel to the crack tip line is shown in Figure 30c and d. In this case all domains are able to switch into the y-direction. The highest crack resistance has to be expected.

From the previous considerations, we expect that initial polarization with its differing domain reservoir in combination with the influence of the internal electric field will lead to the following ranking of fracture toughness:

\[ K_y < K_x < K_{\text{unpoled}} < K_z \]  

(73)

where \( K_x, K_y, \) and \( K_z \) are the toughness values for poling in x-, y-, and z-directions, respectively. This ranking is in agreement with measurements by Lucato et al. [54] (see Figure 31).

6.4.4. Microcrack Zone and Crack Branching

Further reasons for R-curves are the energy-consuming effects which occur ahead of a crack tip also in materials without phase transformations. Ahead of a crack tip a disturbed material zone extends with increasing load. Due to the high stresses, part of the grain boundaries aligned perpendicularly to the maximum principal stress will fail and a zone of microcracks develops [55,56]. Its shape is similar to Figure 27a. In the energy considerations the microcracks contribute to the total crack surface which becomes \( ΔS_{\text{eff}} > BΔa \) and, consequently, increases the effective critical energy release rate. In the stress intensity factor description the microcracks reduce the effective Young’s modulus ahead of the crack tip, the crack tip stresses, and the effective stress intensity. From both points of view, an increased crack resistance must be concluded. An increase of the disturbed zone with increasing crack extension leads to an R-curve effect.

Also in the case of crack branching, the energy consideration yields an increase in the crack surface with increased energy consumption.

**FIGURE 29** Orientation of domains ahead of a crack in an unpoled material (a) under externally applied mechanical load (b), (c), (d) Influence of mechanical load on the domains for PZT poled in x-direction [51].
6.5. Determination of Bridging Stresses

The case of the computation of R-curves on the basis of known bridging stresses has been discussed in [4]. In this section a similar problem will be treated which is founded on the same basic relations.

6.5.1. Bridging Stresses from the R-Curve

Let us assume a measured R-curve which is based on bridging interactions in the crack wake. The specimen may be loaded by an external load, which leads to a stress distribution \( \sigma_{\text{appl}}(x) \) at the location of the crack in the uncracked component. In case the material exhibits a bridging zone with crack-surface interactions, the total stress is the sum of the applied stress and the bridging stress \( \sigma_{\text{br}}(x) \), i.e.:

\[
\sigma(x) = \sigma_{\text{appl}}(x) + \sigma_{\text{br}}(x).
\]

The bridging stresses are responsible for the bridging stress intensity factor \( K_{\text{br}} \) which is given in the representation of the fracture mechanics weight function \( h \) [57] as

\[
K_{\text{br}} = \int_0^a h(x,a)\sigma_{\text{br}}(x)\,dx
\]

and for the applied stress intensity factor \( K_{\text{appl}} \):

\[
K_{\text{appl}} = \int_0^a h(x,a)\sigma_{\text{appl}}(x)\,dx.
\]

The total stress intensity factor is then given as

\[
K_{\text{tip}} = K_{\text{appl}} + K_{\text{br}}.
\]

As a simple example, the bridging stresses may be determined from the R-curve for MgAl2O4 spinel (Figure 17c), which was obtained in a double cantilever beam (DCB) test [32]. The weight function for the DCB specimen reads [7]:

\[
h = \sqrt{\frac{12}{H}} \left[ \frac{a-x}{H} + 0.68 \right] + \frac{2}{\pi(a-x)} \exp \left( -\sqrt{\frac{12(a-x)}{H}} \right),
\]

where \( H \) is the half height of the specimen.
The total displacements of the crack surface, \( \delta \), result from the relation [4]:

\[
\delta(x) = \frac{1}{E} \int_x^a h(x, a') \left[ \int_0^{a'} h(x', a') [\sigma_{\text{appl}}(x')] + \sigma_{\text{br}}(\delta(x')) \right] \, da'.
\]

Due to the long linear range of the R-curve, a very simple attempt is possible to obtain the bridging stresses. From Eqns (64), (75), and (78) it can be concluded that for crack extensions within the straight part of the R-curve, the distribution of the bridging stress must be reciprocal to the weight function, i.e.,

\[
\int_0^a h(x, a) \sigma_{\text{br}} \, dx = C(a - a_0) \iff h(x, a) \sigma_{\text{br}} = C \sigma_{\text{br}} = C/h
\]

The evaluation of bridging stresses from R-curve data is based on the solution of Eqns (75)–(80). The numerical strategy consists of the following steps:

1. For small displacements (corresponding to crack extensions where \( \Delta a \leq 5 \, \text{mm} \)) the bridging relation results from Eqn (80) with the displacements computed by Eqn (79). The result is shown in Figure 32a.

2. For larger displacements the function \( \sigma_{\text{br}}(\delta) \) is represented by the stress values at the sampling points \( \delta_1 \) and \( \delta_2 \), which are interpolated by cubic splines (Figure 32b).

3. For any chosen value of \( \sigma_{\text{br}}(\delta_1) \) and \( \sigma_{\text{br}}(\delta_2) \) (see Figure 32b) the bridging stresses in the whole range of \( \delta \leq 4 \, \mu \text{m} \) were computed, the integral equation was solved, and the crack profile and bridging stress distribution were obtained. From Eqn (75) the related bridging stress intensity factor, \( K_{\text{br}} \), was calculated and from Eqn (56) the R-curve values, \( K_R \), were obtained for a number of selected \( \Delta a \). The computed \( K_R \) values were compared with the experimental data. The bridging stress values at \( \delta_1 \) and \( \delta_2 \) were changed systematically by use of an optimization computer subroutine, until minimum deviations between the computed and measured \( K_R \) were found. The corresponding bridging stresses at \( \delta_1 \) and \( \delta_2 \) then establish the bridging stress relation \( \sigma_{\text{br}}(\delta) \) for larger crack opening displacements (Figure 32c).

Application of Eqns (74)–(76) and (79) on the R-curves for the silicon nitrides shown in Figures 17e and f result in the bridging stresses given in Figures 33a and b. Figure 33a represents the bridging stress distribution in the near-tip wake and Figure 33b shows the bridging relation \( \sigma_{\text{br}} = \varphi(\delta) \) [58].

6.5.2. Bridging Stresses from Crack Profiles

In order to determine the relation between bridging stresses \( \sigma_{\text{br}} \) and the crack opening displacement (COD) \( \delta \), \( \sigma_{\text{br}} = \varphi(\delta) \), the CODs observed during stable crack propagation can be evaluated. Such evaluations were performed in [39] on coarse-grained alumina with the ‘round-compact-tension’ (RCT) specimen and by Kobayashi et al. [59,60] on reinforced alumina with the ‘compact-tension’ (CT) specimen. The procedure is, in principle, based on the relations of section 6.5.1. In Eqn (79), the left-hand side is known as the result of the displacement measurements. The unknown quantity is the bridging stress distribution \( \sigma_{\text{br}}(\delta) \):

\[
\delta_{\text{measured}}(x) = \frac{1}{E} \int_x^a h(x, a') \left[ \int_0^{a'} h(x', a') [\sigma_{\text{appl}}(x')] + \sigma_{\text{br}}(\delta(x')) \right] \, da'.
\]

FIGURE 32 Determination of the bridging relation: (a) solution for small displacements from the linear part of the R-curve, (b) representation of the relation for large displacements by cubic splines (two sampling points), and (c) final result.
The integral equation can be solved numerically and provides the bridging relation

\[ \sigma_{br} = f(\delta) \]

As an example of an application, results from [39] are reported. In Figure 34a the crack opening displacements of Figure 23b are plotted once more. The crack opening profile for \( K_{\text{appl}} = 3.75 \text{ MPa} \cdot \sqrt{\text{m}} \), which would occur in the absence of any bridging effect, was computed with Eqn (76) and is introduced in Figure 34a as the dashed curve. The evaluation of these data via Eqn (78) yields the bridging stress law plotted in Figure 33.

6.5.3. Bridging Stresses from Post-Fracture Tensile Tests

In an effort to directly characterize the wake behavior, Hay and White [61–63] developed the post-fracture tensile (PFT) test which gives the crack closure stress versus the crack opening displacement relationship.

In addition to the half-thickness side groove machined into the DCB specimen (Figure 35a) for crack guidance, two separate grooves are cut to facilitate tensile loading of the PFT specimens, as shown in Figure 35b. Next, several slices are taken from the cracked portion of the DCB specimen (indicated schematically in Figure 35a), thereby isolating

FIGURE 33 (a) Bridging stress distribution for the silicon nitrides obtained from the R-curves of Figures 17e and 17f, (b) related bridging stress versus displacement.

FIGURE 34 Crack profile \( \delta \) for coarse-grained \( \text{Al}_2\text{O}_3 \) (symbols: Measured data, solid curve: Average dependency, dashed curve: \( \delta_{\text{app}} \)); (b) bridging stresses derived from (a) [39].
the wake region. Tensile tests were performed on these specimens with the crack-face separations measured with a laser interferometric displacement gauge. In these tests the wake stiffness and peak load capacity depend on the starting crack-opening displacement (COD) conditions. The PFT specimens cut from near the crack tip exhibit the greatest peak load and stiffness. From a series of load—displacement curves, the relation $\sigma_{br} = f(\delta)$ is derived [32].

The significance of the data obtained by this procedure relies on the successful isolation of the wake without damaging the wake constituents. By using the DCB specimen, this latest incarnation of the PFT technique allows for careful control of the COD profile and for the efficient use of material.

Figure 36a shows the PFT results for a 99.7% Al$_2$O$_3$ ceramic with 12 $\mu$m mean grain size together with an evaluation of the R-curve using the relations of Section 6.5.1. By interpolation of these curves (Figure 36b) a sufficient “wide range” bridging relation can be concluded [32].

6.5.4. Bridging Stresses from Raman Spectroscopy

Pezzotti et al. [64–66] measured the closure stresses in Al$_2$O$_3$, in an alumina/molybdenum composite and in Si$_3$N$_4$ with Raman microprobe spectroscopy. In all cases, a large scatter was observed. Some trends, however, could be detected. In Al$_2$O$_3$ with a grain size of 20 $\mu$m local stresses up to 90 MPa occurred [64]. The average closure stress of 16 MPa was constant up to 350 $\mu$m behind the crack tip. The R-curve increased linearly with increasing crack extension and could be described by the constant stress measured. In Si$_3$N$_4$ close to the crack tip, closure stresses of 1 GPa were measured [65]. The average stress could be described by a two-step curve: 310 MPa up to 100 $\mu$m and 124 MPa up to 800 $\mu$m. In the alumina/molybdenum composite, the maximum local stress was 300 MPa. The average stress was constant up to 500 $\mu$m behind the crack tip [66].
6.6. R-curves from Natural Flaws

So far R-curves measured with specimens containing macrocracks were considered. Only a few experimental results exist with respect to R-curves from natural flaws (see [67]). Some of these results are presented here. In Figure 37 results of Steinbrech and Schmenkel [68] for Al₂O₃ with a grain size of 13 μm are shown in form of the energy release rate as a function of crack extension. In bending tests semicircular cracks were found applying a dye penetration technique. The initial flaw size was 40 μm. The stress intensity factors of the cracks were re-evaluated by Fett and Munz [69]. For the natural cracks, crack extension starts at $G_0 = 5 \text{ N/m}$ and for the macrocracks at $G_0 = 17 \text{ N/m}$, corresponding to $K_0 = 1.34 \text{ MN/m}^{3/2}$ and $K_0 = 2.47 \text{ MN/m}^{3/2}$. The natural crack curve approaches the macrocrack curve at a crack extension of about 200.

Results for zirconia with 9 mol% MgO of Marshall and Swain [70] are shown in Figure 38 and for zirconia with 12 mol% CeO₂ of Ramachandran et al. [71] in Figure 39. Also in these transformation-toughened ceramics the R-curves for natural cracks are below the curves for macrocracks.

7. SUBCRITICAL CRACK GROWTH

In the preceding subsections stable crack extension for a rising crack growth resistance curve and unstable crack extension were considered. A third type of crack extension is called subcritical crack growth (SCG), where a crack is growing at $K_{\text{appl}} < K_{\text{c}}$ or in terms of the crack tip stress intensity factor $K_{\text{tip}} < K_{\text{I0}}$.

The subcritical crack growth is governed by the crack tip stress intensity factor $K_{\text{tip}}$. For a given material and environment there is a unique relation between the crack growth rate $\nu$ and $K_{\text{tip}}$:

$$\nu = \frac{da}{dt} = f(K_{\text{tip}}).$$  \hspace{1cm} (82)

In Figure 40 a $\nu - K_{\text{I}}$ curve is shown in the log–log representation. At low crack growth rates an extended range (region I) occurs with a straight line. In this region the crack growth rates fulfill a power-law relation:

$$\nu = A K_{\text{tip}}^n = A^* \left(\frac{K_{\text{tip}}}{K_{\text{I0}}}\right)^n$$  \hspace{1cm} (83)

with the parameters $A$, $A^*$, and $n$ depending on the material, the temperature, and the environment. In rare cases...
(e.g., glass) a threshold value can be detected, below which no subcritical crack growth is found [72,73]. Sometimes, this threshold value is also denoted as $K_{I0}$. In order to avoid confusion and to distinguish between the onset value of stable crack extension and the threshold in SCG, the symbol $K_{th}$ is used here. At a relatively high crack growth rate, a plateau (region II) in $v$ may occur with crack growth rates independent of $K_I$. Such a plateau was found for glass.

If a $v-K$ curve is determined as shown in Section 7.1, it always has to be assessed what the relation $v(K)$ represents. In the case of small natural cracks in the order of some 10 μm, which will fail at high stresses after a very short crack extension, the difference between $K_{appl}$ and $K_{tip}$ may be negligible. Regarding artificial macroscopic cracks this requirement is not fulfilled for materials with an R-curve behavior (see section 7.1).

As has been outlined in [4] for small cracks, a representation of subcritical crack growth in terms of the applied stress intensity factor $K_{appl}$ is approximately possible:

$$v = A K_{appl}^n = A^* \left( \frac{K_{appl}}{K_{Ic}} \right)^n$$  \hspace{1cm} (84)

where $A$ and $A^*$ are different from those in Eqn (83) and the exponent $n$ is larger than that in Eqn (83). It should be emphasized once more that a description according to Eqn (84) is not possible for SCG curves obtained with macroscopic cracks for materials with a strong R-curve behavior.

It is common use for ceramic materials to represent subcritical crack growth by $v(K_{appl})$ curves. Therefore, in the following relations always $K_I$ is used instead of $K_{tip}$ and $K_{appl}$ in order to simplify the notation.

7.1. Determination of $v-K$ Curves with Macroscopic Cracks

Many types of specimens have been used to determine the $v-K$ relation. As an example of a test specimen, the double cantilever beam (DCB) specimen may be addressed (Figure 41). The specimen consists of a rectangular bar with a slot and is loaded by a pair of forces $F$ that load the specimen in the pure mode I. If the two halves of the specimen are assumed to be clamped at the location of the crack tip, the elementary bending theory provides for the two loading point displacements:

$$V = \frac{8 Fa^3}{EBH^3},$$

and the energy release rate and stress intensity factor resulting from Eqns (19) and (20):

$$G_I = \frac{F^2 dC}{2B da} = \frac{12F^2 a^2}{EB^2H^3}, \quad K_I = \sqrt{\frac{12aF}{1 - \nu^2}} \frac{B H^3}{2}$$

If additional shear deformations are included and the condition of “fixed clamps” is modeled more accurately, it results (see e.g., [74]):

$$K_I = \frac{\sqrt{12aF}}{(1 - \nu^2)^{1/2}BH^{3/2}} \left[ \left( 1 + \frac{a_0}{a} \right)^2 + \frac{1 + \nu}{3} \left( \frac{H}{a} \right)^2 \right]^{1/2},$$

with good approximation, $a_0 = H/3$.

Accurate measurements of the crack growth rate as a function of $K_{appl}$ for materials with a rising R-curve showed first a decrease and after some crack extension an increase in the crack growth rate with increasing $K_{appl}$. This can be explained by Figure 42 where, for a test with constant load, $K_{appl}$ increases according to $K_{appl} = \sigma Y \sqrt{a}$. The increase in the shielding stress intensity factor is larger for small crack extensions and therefore $K_{tip}$ first decreases corresponding to the decrease in the crack growth rate. Therefore, tests with macroscopic cracks cannot be recommended for materials with a rising crack growth resistance curve.

![Figure 41](image-url)
7.2. Evaluation of Lifetimes under Constant Load

Under conditions of subcritical crack growth the lifetime of a material is finite. The lifetime depends on the stress in the component, which may change with time. In cases where the simple crack growth relation given in Eqn (84) can be applied, the lifetime can be computed analytically.

The lifetime, $t_f$, under an arbitrary loading history, $\sigma(t)$, is given by (see e.g., [4])

$$\int_0^{t_f} \sigma(t)^n \, dt = \int_{a_i}^{a_c} \frac{da}{AY\sigma_c^2}$$

(88a)

If the subcritical crack extension is caused by the same flaws as the strength, the initial crack size $a_i$ can be replaced by the strength $\sigma_c$ (see [4]) leading to

$$\int_0^{t_f} \sigma(t)^n \, dt = B\sigma_c^{-2} \left[ 1 - \left( \frac{\sigma_c}{\sigma_c} \right)^{n-2} \right]$$

(88b)

with a material parameter $B$ comprising some fracture mechanical quantities:

$$B = \frac{2K_{ic}^2}{A^2Y^2(n-2)}.$$  

(89)

In tests with constant load, $\sigma(t) = \text{const.}$, one obtains

$$t_f = B\sigma_c^{-2} \sigma_c^{-n} \left[ 1 - \left( \frac{\sigma}{\sigma_c} \right)^{n-2} \right],$$

(90)

resulting in a strong stress dependency of the static lifetimes as a consequence of the high exponent $n$ for ceramics.

For a load sufficiently lower than the critical one, Eqn (90) is simplified to

$$t_f = B\sigma_c^{-2} \sigma_c^{-n}.$$  

(91)

or in logarithmic form:

$$\log t_f = -n \log \sigma + \log \left[ B\sigma_c^{-2} \sigma_c^2 \right].$$

(92)

From this relation, the parameters $n$ and $\log(B\sigma_c^{-2})$ can be obtained by linear regression of $\log(t_f)$ versus $\log(\sigma)$.

As an application of this method, static lifetime measurements are reported in Figure 43 for hot-isostatically pressed $\text{Al}_2\text{O}_3$. They were carried out in four-point bending tests in a concentrated salt solution at 70 °C [4]. In Figure 43a the lifetime results are given in a Wöhler-like diagram and in Figure 43b in Weibull representation [75]. The data of Figure 43 result in

$$n = 20, \log(B\sigma_c^{-2}) = 45.51 \quad \text{(for } B\sigma_c^{-2} \text{ in MPa}^2\text{ h}),$$

The inert strength $\sigma_c$ and the time to failure $t_f$ are properties with a large amount of scatter. Both are related to the scatter of the initial flaw size $a_i$. The scatter of the fracture strength can be described by the Weibull distribution. The cumulative density function $F$ for the two-parameter Weibull distribution is given by

$$F(\sigma_c) = 1 - \exp \left[ - \left( \frac{\sigma_c}{\sigma_0} \right)^m \right],$$

(93)

where $m$ and $\sigma_0$ are the Weibull parameters.
In a plot of $\ln\ln(1/(1-F))$ versus $\ln(n)$ a straight line results and the parameter $m$ is obtained from the slope. The scatter of lifetimes $t_i$ is related to the scatter of strength and represented by

$$F(t_i) = 1 - \exp\left[-\left(\frac{t_i}{t_0}\right)^m\right] \quad (94)$$

with

$$m^* = \frac{m}{n-2} \quad (95)$$

$$t_0 = B\sigma_0^{-\frac{n-2}{n}} \sigma^{-n}. \quad (96)$$

One can find the value $m^*$ as the slope in the plot $\ln\ln(1/(1-F))$ versus $\ln t_i$. Equation (95) can be used to obtain the exponent $n$. An analytical evaluation of the Weibull parameters has to be performed using the maximum likelihood method [76].

In order to evaluate the Weibull parameters for these lifetime distributions, the maximum likelihood method has to be applied. Due to the large scatter and the limited test duration, a number of specimens which fail during load application or survive the chosen time limit (e.g. 1000 h) will be obtained in many cases (especially for small $m$, i.e. large $n$ and small $m$). In such a case, a test series which is truncated at a lower limit $t_{i1}$ and at an upper limit $t_{ia}$ [77] has to be evaluated.

For a number of $n_i$ spontaneous failures with $t_i < t_{i1}$, $n_u$ survivals, where $t_i > t_{ia}$, and $n_v = n - n_i - n_u$ valid test results with $t_{i1} < t_i < t_{ia}$, the equations for the determination of $m^*$ and $t_0$ read

$$n_v + \sum_{i=1}^{n_i} \ln \frac{t_{i1}}{t_0} + n_i \ln \frac{t_{i1}}{t_0} - \frac{1}{t_0^{m^*}} \left[ \sum_{i=1}^{n_i} \left( \frac{t_i}{t_0} \right)^{m^*} \ln \frac{t_i}{t_0} \right]$$

$$= \frac{1}{2} \left[ \sum_{i=1}^{n_i} \left( \frac{t_i}{t_0} \right)^{m^*} \ln \frac{t_{i1}}{t_0} + n_i \ln \frac{t_i}{t_0} - \frac{1}{t_0^{m^*}} \left( \sum_{i=1}^{n_i} \left( \frac{t_i}{t_0} \right)^{m^*} \right) \right] \quad (97a)$$

$$t_0^{m^*} = \frac{1}{n_i + n_v} \left( \sum_{i=1}^{n_i} t_i^{m^*} + n_u t_{ia}^{m^*} + \frac{1}{2} \sum_{i=1}^{n_i} t_i^{m^*} \right) \quad (97b)$$

which have to be solved numerically.

The lifetime procedure described before is based on the power law relation of Eqn (84). An alternative method, also based on a statistical evaluation, was proposed in [4]. The evaluation, however, does not require the assumption of a special prescribed type of crack growth relation. The crack growth rate at the beginning of a lifetime test (subscript $i$) is

$$\nu(K_{li}) = -\frac{2}{t_i \sigma_c^2} \left( K_{lc} / Y \right)^2 \frac{d[\log(\sigma/\sigma_c)]}{d[\log(t_i \sigma^2)]} \quad (98)$$

with the stress intensity factor directly after load application.

In Eqn (98) the lifetime $t_i$ and the inert strength $\sigma_c$ of a specimen have to be known. This is in principle impossible, since either a strength test or a lifetime test can be performed with one specimen. Therefore, a statistical procedure has to be applied.

In a first series of tests, $N$ samples are tested at high loading rates in an inert environment to give the distribution of $\sigma_c$. The $N$ values of strength are arranged in increasing order. In a second series, also involving $N$ specimens, the lifetimes $t_i$ are measured. The results are also arranged in increasing order. The $r$th value of lifetime $t_{ir}$ is associated with the $r$th value of inert bending strength $\sigma_{cr}$. By plotting $\log(t_{ir}, \sigma^2)$ versus $\log(\sigma/\sigma_{cr})$, one obtains a dependency, from which — by use of a suitable smoothing procedure — the actual derivative

$$\frac{d[\log(\sigma/\sigma_c)]}{d[\log(t_i \sigma^2)]}$$

results. By introducing this derivative into (Eqn (98)), the crack growth rate $\nu(K_{li})$ is obtained. The corresponding $K_{li}$ follows from $K_{li}/K_{lc} = \sigma/\sigma_c$. Some results at high temperature are compiled in Figure 44. Figure 44a shows results for boron-doped sintered $\alpha$-silicon carbide and Figure 44b for MgO-doped HPSN.

### 7.3. Dynamic Strength Test

In the dynamic strength test developed by Charles [80], the specimens are loaded with a constant loading rate $\dot{\sigma}$ until failure. The failure stress $\sigma_f$ depends on the loading rate as

$$\sigma_f^{n+1} = B\sigma_c^{n-2} \dot{\sigma}(n+1) \left[ 1 - (\sigma_f/\sigma_c)^{n-2} \right]. \quad (99)$$

For very low loading rates Eqn (99) is simplified to

$$\sigma_f^{n+1} = B\sigma_c^{n-2} \dot{\sigma}(n+1) \quad (100)$$

From measurements of strengths at different stress rates $\dot{\sigma}$, $n$ and $A$ of Eqn (84) can be evaluated by application of Eqn (100). In the logarithmic form this equation reads

$$\ln \sigma_f = \alpha + \beta \ln \dot{\sigma} \quad (101)$$

with

$$\alpha = \frac{1}{n+1} \ln B\sigma_c^{n-2}(n+1), \quad \beta = \frac{1}{n+1} \quad (102)$$

A regression analysis then results in the exponent $n$ and the lifetime-relevant quantity $B\sigma_c^{n-2}$. In this context, it should be mentioned that in the relation for predicting lifetimes under arbitrary loading the parameter $B\sigma_c^{n-2}$ can be directly introduced and an evaluation of the coefficients $A$ and $A^*$ of Eqn (84) is not necessary. The evaluation of the dynamic bending strengths is recommended to be performed via the median values. But sometimes also a regression of all single data points is carried out.
In the study group “Strength and Lifetimes“ of the DKG/DVM (Deutsche Gesellschaft für Keramik e.V. /Deutsche Gesellschaft für Materialkunde e.V.) a round robin was applied to dynamic bending tests including 9 different ceramics [81]. As an example Figure 45 shows the results for SSN and Al₂O₃ in water at room temperature. The different symbols indicate different labs. The data introduced at 100 MPa/s represent the inert strength measurements in air. The evaluation was performed on the basis of the median strength values and also by consideration of the single data points.

As can be seen from the data of Figure 45, large strength scatter occurs. In order to allow an assessment regarding the confidence of the determined parameters, a statistical analysis should be made. A possibility is to calculate the confidence intervals of the relevant parameters. Therefore, some statistics relations necessary for the evaluation of the measurements are reported. With the abbreviation of

\[ \sum_{i=1}^{N} \hat{z}_i \]

the estimates \( \hat{\alpha} \) and \( \hat{\beta} \) for the parameters \( \alpha \) and \( \beta \) result from \( N \) tests as

\[ \hat{\beta} = \frac{[xy]N - [x][y]}{[x^2]N - [x]^2} \quad \hat{\alpha} = \frac{[y]}{N} - \hat{\beta} \frac{[x]}{N} \]

with their standard deviations

\[ \varepsilon_{\hat{\alpha}} = \varepsilon \sqrt{\frac{N[2]}{N[x^2] - [x]^2}} \quad \varepsilon_{\hat{\beta}} = \varepsilon \sqrt{\frac{N}{N[x^2] - [x]^2}} \]

**FIGURE 44** Subcritical crack growth data obtained by Eqn (13.1.8.95), (a) sintered \( \alpha \)-silicon carbide (boron-doped) [78] and (b) HPSN (MgO-doped: Circles, Y₂O₃-doped: Squares) [79].

**FIGURE 45** Dynamic bending strength for alumina and SSN, values at 100 MPa/s: Inert strength data obtained in air, solid, and open circles symbolize different labs [81].
and \( y = \log \sigma_t \) and \( x = \log \dot{\sigma} \). In (Eqn. (105)), \( \epsilon \) is the residual standard deviation defined by
\[
\epsilon = \sqrt{\frac{\left( y - \hat{\alpha} - \hat{\beta} x \right)^2}{N - 2}}.
\] (106)

In a simplification we assume that \( \hat{\alpha} \) and \( \hat{\beta} \) might be normally distributed. Then, the confidence intervals to the confidence level \( 1 - \gamma \) can be determined easily. At a value of \( \gamma = 0.1 \), the true values of \( \alpha \) and \( \beta \) are located with a probability of 90\% in the intervals of
\[
\hat{\alpha} - t_{0.05} \epsilon_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{0.95} \epsilon_{\hat{\alpha}},
\]
\[
\hat{\beta} - t_{0.05} \epsilon_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{0.95} \epsilon_{\hat{\beta}}
\] (107)
with the upper confidence limit factor \( t_{0.05} \) determined at \( \gamma/2 \) and the lower confidence limit factor \( t_{0.95} \) at \( (1 - \gamma/2) \).

In Figure 46 the two straight-line parameters \( \hat{\alpha} \) and \( \hat{\beta} \) are plotted for the median value evaluation and the evaluation of all single data points. From this diagram an excellent agreement can be seen. From these two parameters the strength \( \sigma(0) \) at 1 MPa/s loading rate and the exponent \( n \) were determined.

In order to obtain an impression of the agreement between different laboratories involved in the tests (for each test a lab of a research institute and one from industry), the 90\% confidence intervals for \( \sigma(0) \) and \( n \) are plotted in Figure 47. It follows from Eqns (100)–(102): log \( \sigma(0) = \alpha \) and \( n = (1/\beta) - 1 \). The horizontal and vertical dimension sizes of the rectangles represent the confidence intervals obtained by two labs. An agreement within the 90\% confidence intervals can be stated, if the rectangular areas touch the straight line with slope 1. Apart from \( \sigma(0) \) for the SSN ceramic, sufficient agreement was found between the labs. In contrast to this result large confidence intervals for \( n \) are visible.

In Figure 47 the two straight-line parameters \( \hat{\alpha} \) and \( \hat{\beta} \) are plotted for the median value evaluation and the evaluation of all single data points. From this diagram an excellent agreement can be seen. From these two parameters the strength \( \sigma(0) \) at 1 MPa/s loading rate and the exponent \( n \) were determined.

In order to obtain an impression of the agreement between different laboratories involved in the tests (for each test a lab of a research institute and one from industry), the 90\% confidence intervals for \( \sigma(0) \) and \( n \) are plotted in Figure 47. It follows from Eqns (100)–(102): log \( \sigma(0) = \alpha \) and \( n = (1/\beta) - 1 \). The horizontal and vertical dimension sizes of the rectangles represent the confidence intervals obtained by two labs. An agreement within the 90\% confidence intervals can be stated, if the rectangular areas touch the straight line with slope 1. Apart from \( \sigma(0) \) for the SSN ceramic, sufficient agreement was found between the labs. In contrast to this result large confidence intervals for \( n \) are visible.

In Figure 46 the two straight-line parameters \( \hat{\alpha} \) and \( \hat{\beta} \) are plotted for the median value evaluation and the evaluation of all single data points. From this diagram an excellent agreement can be seen. From these two parameters the strength \( \sigma(0) \) at 1 MPa/s loading rate and the exponent \( n \) were determined.

In order to obtain an impression of the agreement between different laboratories involved in the tests (for each test a lab of a research institute and one from industry), the 90\% confidence intervals for \( \sigma(0) \) and \( n \) are plotted in Figure 47. It follows from Eqns (100)–(102): log \( \sigma(0) = \alpha \) and \( n = (1/\beta) - 1 \). The horizontal and vertical dimension sizes of the rectangles represent the confidence intervals obtained by two labs. An agreement within the 90\% confidence intervals can be stated, if the rectangular areas touch the straight line with slope 1. Apart from \( \sigma(0) \) for the SSN ceramic, sufficient agreement was found between the labs. In contrast to this result large confidence intervals for \( n \) are visible.
7.4. Cyclic Fatigue

Under the assumption of the same mechanism being responsible for constant and cyclic loading, lifetimes can be predicted from the results obtained in tests with constant load by using Eqn (88) and introducing the relevant loading history $\sigma(t)$ under the integral sign. Unfortunately, this method is only suited for a few materials as has been confirmed by cyclic experiments on glass and porcelain [82]. In general, the crack growth rates in ceramics are larger and the lifetimes are shorter than those predicted from constant load tests.

Therefore, it is indispensable to determine the crack growth rates under cyclic loading experimentally. Similar to static fatigue tests one will find power law relations over wide regions of crack growth rates $\frac{da}{dN}$ ($N = \text{number of cycles}$). Two representations of cyclic fatigue results are well established, namely, a description in terms of the range of the stress intensity factor

$$\frac{da}{dN} = A(\Delta K)^n = A^*(\frac{\Delta K}{K_{ic}})^n$$ \hspace{1cm} (108)

or in terms of the maximum stress intensity factor during a cycle $K_{max}$

$$\frac{da}{dN} = A(K_{max})^n = A^*(\frac{K_{max}}{K_{ic}})^n$$ \hspace{1cm} (109)

with the maximum stress intensity factor

$$K_{max} = \sigma_{max}\sqrt{a}Y$$ \hspace{1cm} (110)

with the maximum stress $\sigma_{max}$. The parameters $A$ or $A^*$ and $n$ depend on the R-ratio and the frequency. The R-ratio and the range $\Delta K$ are given by

$$R = \frac{K_{min}}{K_{max}}, \hspace{0.5cm} \Delta K = K_{max} - K_{min}$$ \hspace{1cm} (111)

Integration of the crack growth relations from the initial crack size $a_i$ to the critical crack size $a_c$ yields the number of cycles to failure $N_i$:

$$N_i = \frac{B\sigma_{c}^{n-2}}{(\Delta \sigma)^n} = \frac{B\sigma_{c}^{n-2}}{\sigma_{max}^{n}(1-R)^n}$$ \hspace{1cm} (112)

with $B$ defined by Eqn (89). In a plot of $\log \Delta \sigma$ or $\log(\sigma_{max})$ versus $\log(N_i)$ a straight line with a slope of $-1/n$ is expected.

In principle, all procedures used for static tests are also applicable for cyclic fatigue experiments. Since small cracks are of highest importance for lifetime predictions, the evaluation of crack extension parameters may be outlined here for tests with natural cracks. For this purpose let us consider a piezoelectric material. In their application as actuators, piezoelectric ceramics are subjected to changing electrical and mechanical loads. These changes can lead to an electrical and mechanical fatigue which is connected with a premature failure of the component. In [83] fatigue experiments were performed in pulsating tensile tests on unpoled and poled commercial PZTs (PIC 151, PI Ceramic). The specimens were poled parallel ($\parallel$) and perpendicular ($\perp$) to the tensile direction. Figure 48a shows the Weibull plots of the inert strength $\sigma_c$.

The lifetime results obtained for $R = \sigma_{min}/\sigma_{max} = 0$ and a frequency of 1 Hz are summarized in Figure 48b in the Weibull representation. Similar to the crack growth rate in static tests, Eqn (98), it holds for cyclic tests [84]:

$$\frac{da}{dN} = \frac{2K_{ic}^2}{N_i\sigma_{c}^2Y^2}d[\log(\frac{\sigma_{max}}{\sigma_{c}})]$$ \hspace{1cm} (113)

FIGURE 48 (a) Tensile strength tests on a PZT ceramic in different poling states and (b) numbers of cycles to failure in Weibull representation (symbols as in a).
For the computation of $K_{\text{max}}$ the proportionality between stress and the stress intensity factor can be used, i.e. $K_{\text{max}}/K_{\text{IC}} = \sigma_{\text{max}}/\sigma_{\text{c}}$.

In order to allow the determination of the logarithmic derivative in Eqn (113), an auxiliary diagram $\lg(\sigma_{\text{max}}/\sigma_{\text{c}})$ versus $\lg(N\sigma_{\text{c}}^2)$ has to be plotted. This is illustrated in Figure 49a. The derivative can be determined from a fitted curve which is not necessarily a straight line. The crack growth rates obtained with Eqn (113) are entered in Figure 49b. It is obvious that crack growth rates can be described well by power law relations for the investigated materials at $v < 10^{-6}$ m/s.

To compare crack growths in tension and bending tests, Figure 50a shows bending results from [85] as straight lines and results from pulsating tensile tests for the $\perp$-poled material (symbols). The results for the two different loading types (bending/tension) are in good agreement. A change of the crack growth exponent with the R-curve is obvious. This is in agreement with results for other ceramics (Figure 50b and 50c).

8. THE LOADING PARAMETER $C^*$

In the preceding considerations behavior of cracks was addressed in terms of linear-elastic fracture mechanics. Whereas at low and moderate temperatures the description by stress intensity factors has to be applied, this approach may fail at high temperatures. Fracture at elevated temperatures may be caused by crack propagation which starts from existing flaws or from creep-induced damage caused by the formation and coalescence of pores. In the presence of noticeable creep, the $C^*$-integral proposed by...
Landes and Begley [88] is the crack driving parameter (for details see Riedel [89]).

The \( C^* \) parameter is a path-independent energy-rate line integral. In the two-dimensional case (see Figure 51), it is defined by

\[
C^* = \int_{\Gamma} \left[ W^* dy - T_{ij} \frac{\partial v_j}{\partial x} ds \right] \tag{114}
\]

\[
W^* = \int_0 W_{ij} \, \dot{e}_{ij}, \quad T_i = \sigma_i n_j, \tag{115}
\]

where \( W^* \) is the strain energy density rate, \( T_i \) is the traction vector along the contour \( \Gamma \), \( n_j \) is the outward normal to \( \Gamma \), \( v_j \) the displacement vector, and \( ds \) is a line length increment. The stress tensor is \( \sigma_{ij} \) and the strain rate tensor is \( \dot{e}_{ij} \). The integration path \( \Gamma \) extends from the lower crack surface counterclockwise around the crack tip and ends at the upper crack surface.

A premise for the applicability of \( C^* \) as a loading parameter is the occurrence of secondary creep, which mostly is described by the power-law relation proposed by Norton [90] and Baily [91]:

\[
\dot{e}_s = D \sigma^n. \tag{116}
\]

with exponents in the range \( n \approx 1-4 \) for ceramic materials.

The \( C^* \) integral governs the stresses and strains in front of a crack tip. Under secondary creep condition, an HRR (Hutchinson, Rice, Rosengren) stress field appears ahead of the crack tip [92,93] and the stresses and strain rates are given by

\[
\sigma_{ij} = \left( \frac{C^*}{I_0 Dr} \right)^{\frac{1}{\sigma}} \sigma_{ij}(\theta), \quad \dot{e}_{ij} = \left( \frac{C^*}{I_0 Dr} \right)^{\frac{\dot{\sigma}}{\sigma}} \dot{e}_{ij}(\theta), \tag{117}
\]

where \( I_0(n), \dot{\sigma}(\theta), \) and \( \dot{e}(\theta) \) are dimensionless functions which can be taken from [94].

\( C^* \) can be expressed by the energy rate difference of two identically loaded structures containing cracks of depths \( a \) and \( a + da \), respectively, i.e.

\[
C^* = -\frac{1}{B} \frac{dU^*}{da}, \quad U^* = \int F \dot{\delta} \tag{118}
\]

where \( (B = \text{specimen thickness}) \). This results directly from (114), if the outer contour of the component is used for the path \( \Gamma \).

Equation 118 is the basis for a number of procedures used to determine \( C^* \). The experimental data necessary for the evaluation of Eqn (118) are the force \( F \), the displacement rates \( \dot{\delta} \) of the load application points, and the crack length \( a \).

In the dynamic loading case with constant displacement rates \( \dot{\delta} \) at the load application points, the \( C^* \) parameter can be computed as

\[
C^* = \frac{F \dot{\delta}}{BW} \left[ -n \frac{W \, dF}{(n + 1) \, da} \right]. \tag{119}
\]

where \( F \) is the load and \( W \) is the specimen thickness.

Similar to the description of subcritical crack growth driven by the stress intensity factor \( K \), a power law relation is also possible for creep crack growth. A relation was proposed by Riedel [89] in which the exponent is no longer a free parameter:

\[
\frac{da}{dr} = A(C^*)^{\frac{n}{n+1}}, \tag{120}
\]

Compared with the high crack growth exponents for subcritical crack growth, the creep crack growth exponents are small, namely, less than 1. Experimental results on ceramics (see Figure 52) for an Al₂O₃ ceramic with about 3% glass content confirm Eqn (120) [4].
REFERENCES


Fracture Mechanics Measurements

Stephen W. Freiman
Freiman Consulting, 8212 Inverness, Hollow Terrace, Potomac, Maryland 20854, United States

Chapter Outline
1. Introduction 717
2. Fracture Mechanics Background 718
3. Fracture Mechanics Specimens 719
4. Double Torsion Test 721
5. Tests Based on Flexural Loading 724
6. Indentation Methods 726
7. Double Cleavage Drilled Compression 728
8. Interpretation and Use of Fracture Mechanics Data 729
9. Summary 730
Acknowledgments 730
References 730

GLOSSARY
$\gamma_f$ Critical fracture energy
$E$ Young’s modulus
$G$ Strain energy release rate $\equiv 2\gamma_f$
$\nu$ Poisson’s ratio
$K_I$ Stress intensity factor in mode I loading
$K_{II}$ Stress intensity factor in mode II loading
$K_{III}$ Stress intensity factor in mode III loading
$K_{IC}$ Critical stress intensity factor for fast fracture; fracture toughness
$K_{Ipb}$ Critical stress intensity factor using the precracked beam technique
$K_{Ivb}$ Critical stress intensity factor using the chevron notch technique
$K_{Isc}$ Critical stress intensity factor using the surface crack in flexure technique

1. INTRODUCTION

There are two primary outputs that one can obtain from fracture mechanics tests. First, one can determine a material’s fracture toughness, $K_{IC}$, which is a measure of its resistance to the rapid growth of a pre-existing crack. Fracture toughness is an important factor in a material’s tendency to form a crack under a localized impact such as that produced by a dust particle or a water droplet, as well as serving as a guide to a material’s strength, since it helps define the severity of the critical flaw. It can also be used either as a quality control measure or to assess potential improvements in the strength of new materials.

A second important use for fracture mechanics tests is in the determination of the susceptibility of a material to environmentally enhanced crack growth. Most brittle materials are subject to delayed failure, what many term as “stress corrosion,” in which cracks grow slowly under the simultaneous presence of a tensile stress and an environment which can react in some way with the chemical bonds at the crack tip. The development of the science of fracture mechanics has permitted investigators to separate crack growth phenomena from the statistics of brittle fracture. Many of the tests described herein can be used to understand the mechanisms of environmentally enhanced crack growth (see the chapter on Environmentally Enhanced Crack Growth [1] for more details).

This chapter provides a compilation and overview of a broad range of measurement methods available for the quantitative determination of crack growth in brittle materials, which, in simplest terms, are the ones in which plastic deformation, e.g. dislocations and twins, are insufficiently mobile to relieve crack-tip stress concentrations. Brittle materials include some metals, e.g. tungsten, and intermetallics, and some polymers, e.g. polymethylmethacrylate and polystyrene; however, for the most part, the primary materials of interest are inorganic nonmetals such as ceramics, semiconductors, and glasses. The chapter focuses on those tests that can be used on single crystals and amorphous or polycrystalline materials in bulk form. It deals with the most commonly used and accepted fracture mechanics techniques. Fracture mechanics tests applicable to films and coatings bring their own special challenges, and are outside the scope of this document. While the test procedures described herein can be used for two-phase materials, such as particulate-reinforced composites, measurements applicable to continuous-fiber reinforced
composites are more complex and are also excluded. For a more comprehensive description of fracture mechanics tests, the reader is referred to Chapter 4 in “The Fracture of Brittle Materials” [2].

2. FRACTURE MECHANICS BACKGROUND

In 1920, Griffith [3] postulated that small cracks in glass grow when subjected to a large enough tensile stress. He put forth the idea that resistance to the growth of these cracks is determined by the energy required to create the two fracture surfaces produced by their extension. This approach assumes that fracture occurs in an equilibrium manner, i.e. in the absence of any kinetic effects, and that no energy is lost due to plastic flow or heat. It also neglects possible effects of the test environment in which the flaw is growing, e.g. water. Griffith’s expression for glass fracture based upon this approach is given by

$$\sigma_t = \left(\frac{2E\gamma_f}{\pi a}\right)^{1/2}$$  

where $\sigma_t$ is the fracture strength, $E$ is Young’s modulus, $\gamma_f$ is the energy required to form the crack surfaces (i.e. the fracture energy), and $a$ is the critical flaw size, i.e. that at which it begins to extend. While Griffith’s calculations were not entirely correct because of his assumption that the fracture energy of the glass was a surface energy that could be extrapolated from measurements carried out at elevated temperatures, the form of Eqn (1) accurately depicts the relationship between strength, flaw size, and fracture energy.

The use of fracture energy as a measure of a material’s resistance to crack extension continued for many years. A significant development was the introduction of the concept of energy release rate, $G$, where $G$ is defined by the total energy consumed in growing the crack, and is, by definition, equal to twice the fracture energy derived by Griffith, i.e. $G = 2\gamma_f$ [4].

An important later development came about with the science of linear elastic fracture mechanics [5]. Linear elastic fracture mechanics is based on the idea that an applied stress produces a stress concentration at a crack tip. The stress intensity factor, $K_I$, is defined through the following expression [4]:

$$\sigma_y = \frac{K_I}{\sqrt{2\pi r}}$$  

where $\sigma_y$ is the applied, far-field, stress and $r$ is the distance extending outward from the crack tip. Eqn (2) is only valid outside of any zone at a crack tip within which the material is nonlinearly elastic or in which some permanent deformation has taken place. The beauty of linear elastic fracture mechanics is that it can be used to explain fracture in spite of the existence of this singularity.

Crack extension takes place when the far-field stress reaches a level such that the stress intensity factor produces crack-tip stresses large enough to rupture atomic bonds. A material’s fracture toughness is defined as the value of the stress intensity factor, $K_I$, at the point where $K_I = K_{IC}$, the critical stress intensity factor.

In most brittle materials, cracks can extend slowly under stress in moisture-containing environments. The key is the relative humidity of water in the environment, i.e. its partial pressure. No liquid, no matter how small is the solubility of water in it, can be considered to be completely inert. Because of these environmental effects, $K_{IC}$ for brittle materials is frequently defined as the point at which rapid crack extension occurs, usually seen as a drop in the load on a testing machine.

Because there are numerous procedures for determining $K_{IC}$, there are two options to defining fracture toughness: (1) assume that $K_{IC}$ is a fundamental material property, and that each technique represents an attempt to estimate its value by a different method, or (2) assume that fracture toughness is only defined in terms of the measurement itself, and that there is nothing particularly fundamental about each determination.1 The proposal is made that there is a number that we can call $K_{IC}$ for each material, but whose determination will depend on many factors including test procedure, environment, etc. One can show, that given a material whose microstructure and sensitivity to environmentally enhanced crack growth are not factors in its fracture toughness, values of $K_{IC}$ obtained by multiple techniques are very similar (see ASTM Standard C-1421 [6]).

Nonetheless, for a given material, $K_{IC}$ will be a function of environment, testing rate, and test geometry. For most of fracture mechanics tests, $K_{IC}$ is calculated from the maximum recorded load. The crack velocity at which this point occurs will be a function of the compliance of both the specimen and the testing machine, but is usually in the range of $10^{-2}$–$10^0$ m/s. Fortunately, crack growth curves are typically very steep; so any uncertainty in crack velocity translates into only a small variability in fracture toughness. A suggested general definition of fracture toughness is:

*For a brittle material in which no subcritical crack growth occurs, $K_{IC}$ is the critical stress intensity factor at which crack growth begins.*

Another issue is the stress state at the crack tip, i.e. plane strain or plane stress. Whether a specimen is in a state of plane stress or plane strain depends on the constraints at the crack tip. Where there is significant constraint, e.g. thick specimens, and little or no plastic deformation, a state

1. This is the approach taken in ASTM Standard 1421 [6].
of plane strain will exist. In all specimens and materials, the outer surface of the specimen will be in a state of plane stress, but the depth of the plane stress zone could be quite small. In metals, the condition of plane strain is defined in terms of the size of a plastic zone at the crack tip, where the zone size is determined by the ratio of $K_1$ to the yield stress in the material. Given that tensile yield stresses for brittle materials are too large to measure, zone sizes at crack tips will be quite small. It can therefore be assumed that for almost any brittle material and for any specimen size, the crack will be in a state of plane strain.

The issue of plane strain or plane stress arises in particular when one transforms measurements between fracture energy and stress intensity. It should be noted that both approaches are equivalent — neither is more fundamental. The relationship between fracture toughness and fracture energy or strain energy release rate is given by

$$K_1 = \sqrt{G}$$

(3)

$$K_1 = \sqrt{G(1 - \nu^2)}$$

(4)

where $\nu$ is Poisson’s ratio. The $(1 - \nu^2)$ term has been usually ignored in the literature in discussing brittle fracture. For most materials neglecting this term will introduce an ~4% uncertainty into the absolute value of either $G$ or $K_1$ depending on the primary measurement.

3. FRACTURE MECHANICS SPECIMENS

3.1. Double Cantilever Beam

Double cantilever beam (DCB) configurations were the first fracture mechanics specimens to be used to determine the fracture energy of brittle materials. In fact, the identification and use of DCB configurations predate the development of the science of fracture mechanics by almost 20 years [7].

For the typical DCB specimen such as that shown in Figure 1, beam theory can be used to relate $\gamma_f$ to the force, $F$, applied to the end of the specimen. The groove is essential in specimens other than single crystals in which cracks are constrained to grow along only one crystallographic plane.

The energy stored in the specimen is equated to that required to form two new surfaces, leading to the expression [8]

$$\gamma_f = \frac{F^2 L^2}{2EI\ell}$$

(5)

where $L$ is the crack length, $E$ is Young’s modulus, $t$ is the thickness of the specimen at the groove, and $I$ is the moment of inertia of one-half of the specimen about its longitudinal axis. $I$ must be modified to account for the groove, which is typically one-half of the specimen thickness:

$$I = \frac{Wh^3}{12}$$

(6)

Wu et al. [9] derived an expression to accurately account for the presence of the groove:

$$I = \frac{bh^3}{12} + bh(\Delta x)^2 - \left[\frac{1}{12}(b-t)(h-w)^3 + (h-w)(b-t)\left(\frac{w}{2} + \Delta x\right)^2\right]$$

(7)

$$\Delta x = \frac{w(h-w)(b-t)}{2[bh - (h-w)(b-t)]}$$

(8)

As a guide, specimen dimensions equivalent to glass microscope slides work quite well, but other ratios of width to thickness are possible. As with the conventional DCB specimen, data should be taken from the center third of the specimen length.

The analysis of the DCB specimen was later modified by subsequent researchers [8,10] in order to account for effects of shear and the energy stored in the uncracked portion of the specimen. Wiederhorn et al. [10] derived the following modified expression for fracture energy:

$$\gamma_f = \gamma_A\left[1 + 1.32h/L + 0.54(h/L)^2\right]$$

(9)

where $\gamma_A$ represents the value of fracture energy without the modifications.

Ignoring the $(1 - \nu^2)$ term, $K_{IC}$ is then given by

$$K_{IC} = (2E\gamma_f)^{1/2}$$

(10)

One drawback in the use of this particular DCB loading configuration is that $K_1$ is crack-length dependent. Making an accurate determination of $K_1$ requires that one be able to precisely determine the position of the crack tip. In opaque

---

2. For anisotropic materials, $E$ should be taken as the value of the modulus parallel to the longitudinal axis of the specimen.
ceramics, this can prove to be a significant problem. While cracks can be seen in translucent materials, e.g. alumina, under transmitted light, the crack tip can have a wavy appearance, leading to an uncertainty in $K_I$.

According to Wiederhorn et al. [10], measurements should be made for crack lengths greater than 1.5 $h$ from each end of the specimen. Favored specimen dimensions are $h \sim 6$ mm, $b \sim 1$ mm, $t \sim 0.5$ mm, and specimen length $\sim 25$ mm. Specimens having other dimensions can be used, but if $h$ is too small the crack will turn to break off an arm; if $h$ is too large the forces required to propagate the crack can damage the specimen.

This DCB specimen shown in Figure 1 is loaded through a tensile force applied to pins or hooks inserted through holes drilled in the specimen. It is important that the loading pins be able to rotate freely since frictional forces will produce errors in the determination of $\gamma$ or $K_I$. In order to obtain slow crack growth data, loads are applied by hanging weights on wires attached to the loading pins. If the specimen is being used to determine fracture toughness, the pins can be attached by wires to a universal testing machine.

3.2. Applied Moment Loading

In 1973 Freiman et al. [11] described a general expression for fracture energy measured on a DCB specimen that can be written to include an applied bending moment term. The strain energy release rate, $G$, is given by

$$G = \frac{F^2 a^2}{Elt} \left[ \left(1 + \frac{\delta}{a} \right)^2 + \frac{E}{12G} \left( \frac{h}{a} \right)^2 \right]$$

(11)

where $F$ is the force on the moment arm, $G$ is the shear modulus, $E$ is Young’s modulus, and the dimensional parameters are shown in Figure 2.

If the specimen is loaded only by a moment, $G$ is crack-length independent, and is given by

$$G = \frac{M^2}{Elt}$$

(12)

where $M =$ (applied load, $F$) × (moment arm, $l$). $l$ is fixed by the distance between the loading point and the point at which the arms fit into the loading fixture, and $l$ is given in Eqn (6). The expression for $K_I$ is given by

$$K_I = \frac{M}{(lI)^{1/2}}$$

(13)

The AMDCB specimen must also be grooved to ensure that the crack propagates down the center. The groove depth should be roughly one-half of the thickness of the specimen, and should have square corners.

Loads can be applied through a testing machine in the case of the determination of $K_I$ or by dead weight loading when crack velocity data are being obtained. In the latter case, a simple method is to hang a pan containing weights from the triangle. For both types of loading, care must be taken that the pins in the trapeze rotate with a minimum of friction since a small resistance to rotation as the crack grows can lead to significant overestimates of $K_I$ (Figure 3).

Two types of loading arms are typically employed. The simplest are two metal bars, grooved to fit the trapeze, and with a dimple on the opposite side to accommodate the loading point, which are cemented to the top of the specimen. This type of arm works well for materials with a small resistance to crack growth, e.g. glasses, in which the loads on the arms are therefore relatively small.

A second, more complex, type of arm involves an arrangement in which the arms are held to the top of the specimen through a clamping arrangement. Inserts in the arms keep the sample parallel to the loading plane during testing. Inserts can be prepared with a range of openings for the arms, allowing the testing of specimens of different thicknesses. The inserts are clamped against the sample by means of a pointed screw extending through each arm.

3.3. Introducing and Guiding a Crack

Introducing a starting crack into either version of DCB specimen such that it propagates only partially down the length can be burdensome. One method involves using a sharpened screw to put a force on the ungrooved side of the specimen just at the base of the starting notch.
With care, crack pop-in occurs with crack arrest, but this procedure is quite sensitive to the force applied, and results in the loss of a fair number of specimens. A better method is to use a Vickers or Knoop indenter to place a hardness indentation at the base of the notch, and depend on the formation of the small cracks emanating from this indentation to act as the starting crack. The ungrooved side of the specimen should be polished to facilitate observation of a crack as it pops in. Using an intense light directed at an oblique angle to the specimen surface, one can see the crack as it initiates even in opaque specimens of polycrystalline ceramics. Tapering the starting notch can also aid in the introduction of a stable crack.

### 3.4. Determination of $K_{IC}$

Because cracks will grow at different rates depending on the moisture content of the atmosphere, measurements conducted slowly in air are likely to yield smaller values of $K_{IC}$ than those taken in dry environments under rapid loading conditions. Because slow crack growth susceptibility for an untested material cannot be predicted, the conservative approach is to test in relative dry environments, e.g. vacuum pump oil that has been dried in a furnace or in dry nitrogen gas introduced into a plastic bag surrounding the specimen and load train, and to load the specimen to failure as rapidly as the testing machine will permit. On the other hand, if the data are to be used to predict strengths in ambient environments, then determining fracture toughness under these conditions would make sense.

A frequently asked question is the minimum number of specimens that are required to obtain a statistically significant value of $K_{IC}$. The ASTM Standard [6] for fracture toughness demands that only four specimens be tested. However, the choice of the number of specimens will depend on the scatter in data and the accuracy requirements of the particular measurement. Even though the DCB technique has never undergone a rigorous standardization process, values of fracture toughness obtained by this method generally agree with those determined by other techniques.

### 3.5. Measurement of Crack Growth Rates

The DCB specimen offers a very useful configuration for measuring crack growth rates. In transparent materials, such as glasses and some single crystal and polycrystalline materials, observing the motion of the crack is fairly straightforward. A focused light is placed behind (or in front of) the specimen at an angle that will cause the crack face to be illuminated. A traveling microscope is then used to measure the position of the crack front. In translucent or opaque materials, observing crack motion is more difficult. In translucent materials, proper lighting can at times reveal the crack, but accuracy in position is lost. If the ungrooved surface of opaque materials is polished, the crack can at times be seen, but the resolution is questionable. As long as the relative motion of the crack is all that is needed, data can be obtained. However, the scatter in data can be considerable.

### 4. DOUBLE TORSION TEST

The double torsion (DT) specimen was first described by Outwater et al. [12]. It has a number of desirable attributes:

- $K_I$ is crack length independent
- the specimen geometry is simple
- amenable to loading in compression, so use at elevated temperatures and in harsh environments is facilitated
- crack extension can be monitored through changes in specimen compliance, obviating the need for direct observation
A review paper on the DT technique by Shyam and Lara-Curzio [13] provides a wealth of detail on the background, analysis, and use of this technique, including over 90 references. An earlier review of the technique by Tait et al. [14] also contains useful information.

As seen in Figure 4, the DT specimen is loaded by subjecting one end of a rectangular specimen to a torque through the application of a four-point loading system. Steel spheres are typically used as the loading points. In the early versions of this test, a center groove was used to prevent the crack from deviating from the center. However, it was subsequently found that through proper alignment and loading, cracks will grow stably down the center of the specimen[13], avoiding the need for grooves.

Fuller [15] and Pletka et al. [16] developed an analysis of the DT specimen, and demonstrated its use in experiments. The driving force on the crack, i.e. the strain energy release rate, $G$, is developed from the expression

$$ G = \frac{P^2}{2d} \frac{dC}{da} $$

where $P$ is the applied load, $C$ is the elastic compliance of the specimen, and $a$ is the crack length. Then

$$ G = \frac{3P^2W_m^2}{2Wd^3d^G\Psi} $$

wherein $P$, $W_m$, $W$, and $d$ are defined in Figure 4, $G$ is the shear modulus. $\Psi$ is a factor determined by Fuller [15] correcting for the fact that a finite beam is being analyzed, and is given by

$$ \Psi = 1 - .6302\tau + 1.20\tau \exp\left(-\frac{\pi}{\tau}\right) $$

where $\tau$ is given by $(2d/W)$. The need for this correction comes about particularly in the use of thicker beams.

The $K_I$ for such a specimen is given by

$$ K_I^2 = E'G $$

$E' = E/(1 - \nu^2)$ for plane strain. Assuming that plane strain applies,

$$ K_I = PW_m \left[ \frac{3(1 + \nu)}{Wd^3d^G} \right]^{1/2} $$

Note that uncertainties in the value of $\nu$ for a material can lead to some ambiguity in the assessment of $K_I$, especially when applied to a material of uncertain elastic properties.

Finite element analyses of the DT specimen [17,18] have shown that $K_I$ is constant as long as the crack length, $a$, is $>0.55W$, where $W$ is the full width of the specimen, and $(L - a) > .65W$. As noted by Shyam and Lara-Curzio [13], this means that for $L/S = 3$, $K_I$ is constant over the middle 60% of the test specimen. Too short a starting crack leads to an overestimate of the toughness, while too long a crack underestimates it, which has been confirmed experimentally (Figure 5) [16].

As with a number of the other fracture mechanics specimen configurations, the DT specimen must be pre-cracked in order to ensure that an accurate value of fracture toughness is obtained. Some of the same procedures applicable to the DCB test can be used here as well. Another method of starting a crack is to taper a notch at the frontend of the specimen. If the notched specimen is loaded slowly, the pop-in of a crack can be observed through a drop in the load. A second method is to introduce a series of Knoop indentations along the starting notch. The presence of the cracks emanating from the indentation sites will help produce a sharp crack during preloading. Also,
a Knoop indenter can be used to form a straight scratch on the center of one end of the specimen from which a crack will emanate.

4.1. Fracture Toughness Determination

Critical fracture toughness can be determined using the DT specimen by loading a precracked specimen rapidly until the crack grows the length of the specimen. $K_{IC}$ is calculated from Eqn (18) using the largest recorded load, $P$. The same issues of loading rate and moisture as discussed with respect to the DCB test apply here as well.

While this method yields accurate values of fracture toughness, the quantity of material required for each specimen means that other techniques to determine $K_{IC}$ may be better choices. The exception would be if the material is produced in large sheets or can be cast to shape such as with concrete.

4.2. Crack Growth Measurements

Although it is possible to collect crack growth data in a DT specimen by placing it under constant load while monitoring crack length optically [19], two other procedures, a load relaxation and a constant displacement rate testing method, have been shown to be effective in determining crack growth rates without a need to view the crack directly [20].

Because the driving force on the crack is independent of crack length, the change in specimen compliance can be used to determine crack extension. The deflection of the specimen at the loading points, $y$, is given by

$$y = \frac{3Paw^2}{Gt^3W}$$  \hspace{1cm} (19)

The change in crack length, $\Delta a$, can therefore be calculated from a corresponding $\Delta y$ through

$$\Delta a = \frac{Gt^3W}{3pW^2_m} \Delta y$$  \hspace{1cm} (20)

Evans [20] noted that, as shown above, the displacement of the loading points, $y$, is directly related to the applied load, $P$, and the crack length, $a$, through

$$y = P(Ba + C)$$  \hspace{1cm} (21)

where $B$ and $C$ are constants related to the elastic properties of the material and the dimensions of the specimen and the test device. For large initial crack lengths, $a_i$, one can show that

$$V = -\frac{a_iP_i}{P^2} \left( \frac{dP}{dt} \right)$$  \hspace{1cm} (22)

Tests are conducted in a machine capable of constant crosshead displacement by loading a precracked specimen to a predetermined load, $P_i$, at a relatively high rate of loading so that no crack growth occurs prior to reaching the fixed load. The crosshead is stopped, and the relaxation of the load is followed. The data obtained by Evans on polycrystalline alumina in order to verify this procedure are shown in Figure 6.

Caution should be exercised in carrying out this technique to very low crack velocities, i.e. below $10^{-7}$ m/s, because of possible interferences due to compliance variations in the test machine, temperature variations, or electrical perturbations.

Evans [20] also suggested a second method of acquiring crack growth data on the DT specimen. He observed that if one holds the rate of displacement, $\frac{dy}{dt}$, constant then

$$\frac{dy}{dt} = BPV$$  \hspace{1cm} (23)

Since $K_I = AP$, $A$ is a constant involving specimen geometry and elastic properties:

$$\frac{dy}{dt} = B' K_I V$$  \hspace{1cm} (24)

![Figure 6](image_url)
A plateau in load, $P$, is evidence that the crack is growing in a stable manner (Figure 7). The constant loading rate technique is particularly useful in obtaining data under severe environmental conditions in which either viewing the crack directly or the stability of the loading system is questionable.

One drawback in the use of the constant displacement rate method is that the range of crack velocities that can be obtained is limited to those $>10^{-5}$ m/s.

The DT technique is an excellent choice for obtaining crack growth data, particularly under conditions in which observing a crack optically is difficult. It is less likely to be the choice for a critical fracture toughness determination for most brittle materials because of the amount of material required. However, one of the significant users of the DT technique has been the geological community, which has made use of the ability to scale the specimen to large sizes to investigate mineral and rock fracture.

5. TESTS BASED ON FLEXURAL LOADING

A number of fracture mechanics test geometries and procedures are based on a rectangular bar in which a notch or crack has been introduced, and which is subsequently loaded in either three-point or four-point flexure. These tests include the single-edge-notch-bend (SENB) test, the work of fracture (WOF) test, the single-edge-V-notch-bend (SEVNB) test, the surface crack in flexure (SCF) test, and the single-edge-precracked-beam (SEPB) method. Except for the WOF test, these procedures differ essentially through the method by which a starting crack is introduced into the flexural bar. Only some of these are described in this chapter. The reader is referred to Ref. [21] for descriptions of the others.

All of the tests in this section require that a specimen be broken in flexure. Reproducible, accurate results of a flexural test depend on the specimen having the correct dimensions and machining, and it being loaded correctly. The particulars of how to conduct a flexural test for ceramics, including detailed descriptions of loading fixtures, are contained in both an ASTM [21] and an ISO standard [22]. In addition, an excellent review of flexural strength procedures was published by Quinn and Morrell [23].

5.1. Single-Edge-Notch-Bend Test

The first fracture mechanics test for brittle materials involving a flexural specimen was the SENB test (also called the notch-beam test) suggested by Davidge and Tappin [24]. As its name suggests, it consists of a bar with a notch cut into one of the narrow sides (Figure 8).

While there are extensive data in the literature on results of this test, because of the absence of a true crack, SENB data should not be trusted. It is typically assumed that in the case of brittle materials, a machining-induced crack exists at the root of the notch. This assumption is frequently incorrect, and leads to overestimates of the fracture toughness.

5.2. Chevron-Notch-Bend Tests

Munz et al. [25] suggested an improved version of the SENB test, which makes use of a flexural bar containing a symmetrical chevron notch, and is termed the chevron notch bend test (CNBT) (Figure 9). The specimen can be loaded in either three- or four-point bending in a universal test machine.

Critical fracture toughness is calculated from the maximum load reached before failure using knowledge of the specimen compliance, thereby avoiding the need to
determine an initial crack size. This specimen is one of three methods that has been adopted as both an ASTM Standard [6] and an ISO Standard [26].

Equation (25) can be used to calculate fracture toughness, designated as $K_{IVb}$. The subscript reflects ASTM’s position that, because of factors such as a different response to environmentally enhanced crack growth, each of these tests may lead to a different value of fracture toughness:

$$K_{IVb} = Y_{min} \left[ \frac{P_{max}[S_o - S_i]10^{-6}}{BW^{3/2}} \right]$$  \hspace{1cm} (25)

where $S_o$ and $S_i$ are the outer and inner loading points, respectively, $W$ is the specimen height, and $B$ is the specimen width. $Y_{min}$ is defined as the minimum stress intensity factor coefficient which depends on the specific specimen geometry.

In order for a test to be valid, the force–displacement curve must exhibit a smooth maximum as seen in Figure 10a. Curves such as shown in Figure 10b are indicative of a valid test as well, but care must be taken to use the $P_{max}$ point on the diagram in the calculation. The observation of a sharp drop (Figure 10c) warns of an overestimate of the fracture toughness.

The narrower is the notch the more likely it is that a stable crack growth will ensue, and the less will be the scatter in data. It has been observed that some skill is needed in order to properly machine such a specimen. Valid tests are more likely if cracks can be initiated at the chevron tip prior to loading in the test itself. The standard suggests that one way of accomplishing this is to turn the specimen upside down and load it in compression over a few loading cycles.

As a way of better ensuring that a crack grows smoothly from the tip of the chevron, Nishida et al. [27] introduced the idea of sharpening the notch in an SENB specimen through the use of a razor blade impregnated with 1-micrometer diamond paste. The razor blade is drawn back and forth through the notch to produce V-shaped notches that can act as cracks. This test has become known as the SEVNB test.

Kübler [28] expanded the discussion of the SEVNB test providing valuable details on the best procedures for use of the razor blade to sharpen the notch. He reported that notch widths as narrow as two micrometers could be created in a hot-pressed silicon nitride.

The ISO standard [29] for the chevron notch test (ISO 24370) is relatively similar to ASTM Standard C-1421. However, it calls for only two specimen geometries, and does not provide for the possibility of three-point bending.

5.3. Single-Edge-Precracked-Beam

The SEPB test uses a specially designed system to introduce a controlled crack into the surface of a bend specimen. It was put forward by Nose and Fujii [30] based on the so-called “bridge-indentation” procedure [31] for stably extending a crack either from a hardness indentation or from a machined notch. This method now exists as both an ASTM [6] and an ISO standard [32].

The crack starter can be a machined notch, one or more Vickers indentations, or a Knoop indentation with its long axis perpendicular to the longitudinal axis of the specimen. The precrack is then loaded in compression in a specially designed fixture. A sound is created at pop-in, but it may be necessary to use a stethoscope to detect it. The use of an ultra-penetrating fluorescent dye penetrant helps to determine the final crack length before testing. As with other flexural tests, it is necessary to ensure that stable crack growth occurred, as evidenced from the shape of the force–displacement curve (Figure 10) in order to obtain an accurate value of $K_{IC}$.

ASTM Standard C-1421 [6] gives expressions which can be used to calculate $K_{IB}$. Because the method and the fixture are relatively complex, this technique is not used extensively.

5.4. Surface Crack in Flexure

The SCF technique was first suggested by Petrovic and Mendiratta [33]. It has since become incorporated as an established procedure in ASTM Standard C-1421 [6] as well as into an ISO Standard [34]. The test involves making a Knoop hardness indentation in the surface of a flexural bar, thereby forming a crack underneath the indentation site, removing the residual stressed zone surrounding the indent, and fracturing the bar in four-point bending. A schematic of the indented bar is given in Figure 11. The standard points out that the indentation can be made in either long face of the bar.
6. INDENTATION METHODS

By indentation methods is meant those procedures that involve using a hardness indentation coupled with the ensuing crack system and the stress field introduced by the indentation as the primary source of the starting crack in a fracture mechanics test. There are two basic indentation test methods.

In the first, known as indentation fracture (IF) [37], a Vickers indenter is used to create a hardness impression. The length of the cracks emanating from the corners of the indentation is measured and combined with knowledge of the elastic properties and hardness to calculate fracture toughness.

In the second, called indentation strength (IS) [38], the indentation is placed in a bar of the material, and the strength measured in flexure. This test differs from the surface crack-in-flexure method discussed previously in that the residual stress system associated with the indentation itself is not removed and is considered as part of the driving force on the crack.

6.1. Indentation Fracture

In the IF technique, a Vickers hardness indentation is placed in the surface of the material producing cracks emanating from the corners of the hardness impression (Figure 12); the indenter is then removed, and the length of the cracks measured.

One of the serious complications and ultimate shortcomings of this test is the empirical nature of the expressions used to relate “$K_{IC}$” to the measurable parameters, load, hardness, elastic modulus, and crack length. Numerous expressions relating fracture toughness to indentation parameters have been suggested to calculate fracture toughness by this procedure [39–41]. One of the more popular versions of such an expression is given in Eqn (27) [37]:

$$K_{IC} = k \left( \frac{E}{H} \right)^{1/2} \left( \frac{P}{c_{0}^{3/2}} \right)$$

(27)

where $E$ is Young’s modulus, $c_{0}$ is the measured crack length, $P$ is the indentation load, $k$ is claimed to be a material-independent constant, and $H$ is the hardness determined from the expression:

$$H = \frac{P}{a_{0}a^{2}}$$

(28)
The term $K_C$ rather than $K_{IC}$ is used to emphasize that the procedure should not be considered to lead to “standard” fracture toughness. In this expression, $a$ is the size of the indentation impression and $a_0$ is a numerical constant. Anstis et al. obtained a value of $k = 0.016 \pm 0.004$ by averaging fracture toughness data obtained by conventional techniques over a range of materials.

There are three significant problems associated with this test procedure [39–41]:

- The method cannot yield a value of fracture toughness that is consistent with the definition of $K_{IC}$ given previously, namely the stress intensity at which unstable crack growth occurs. Rather than accelerating, the crack decelerates and stops. There is no basis for assuming that this crack arrest value corresponds to a fracture toughness value obtained by accepted methods in which a single crack is grown to failure.
- The crack system itself can be extremely complex depending on the microstructure of the material. More than one crack can be present, and the presence of lateral cracks, which extend more or less parallel to the surface below the indentation, can further interfere with the growth of the radial cracks.
- None of the expressions used to calculate $K_C$ are fundamental in nature.

The question arises of whether this test could be used to rank materials without depending on the numerical accuracy of $K_C$. This too is questionable. Ponton and Rawlings [39,40] compared materials using a number of the expressions proposed to calculate fracture toughness, and concluded that an accurate ranking of materials with respect to fracture toughness depended on the set of materials chosen and the expression used to calculate $K_C$.

Nonetheless, there are situations in which there are no other techniques available to determine fracture toughness. One example is the determination of the fracture resistance of small single crystals, where the size of the available crystals (of the order of a few millimeters or less) precludes the use of standard fracture mechanics tests. Determination of the fracture toughness of single crystals poses an additional problem, however, due to the anisotropic nature of the elastic properties.

Another use of the IF test is in the investigation of the possible role of environment, e.g. water, on crack extension. If environmentally enhanced crack growth takes place, then the cracks will extend to a greater distance than $c_0$, the crack length measured in an inert environment. This factor must be taken into account in any calculation of fracture toughness, but it has been shown that it also possible to use the ratio of crack lengths in various environments to obtain at least a semi-quantitative measure of the material’s susceptibility to subcritical crack growth [42–44].

If one assumes that crack growth follows the relationship

$$V = V_0 \left( \frac{K_I}{K_{IC}} \right)^N$$

then $V_0$ and $N$ are empirical material parameters. Based on the analysis by Gupta and Jubb [42], one can derive the following expression for the ratio of crack lengths in two different environments:

$$\frac{c}{c_0} = \left[ \frac{(3N + 2) V_0 t}{2c_0} \right]^{\frac{2}{(3N + 2)}}$$

where $c_0$ is the initial crack length after indentation and $c$ is the crack length measured at time, $t$. This expression is valid when $(3N + 2)V_0 t/2c_0 \gg 1$.

Despite the fact that some data are lost because of the difficulty in obtaining measurements at very short times, White et al. [43] and Cook and Liniger [44] have demonstrated that this procedure can be used to determine values of $N$ that agree quite well with those obtained from classical fracture mechanics specimens.

This indentation procedure has also been employed as a way of determining effects of residual stresses [45] and applied electric fields [46] on crack growth. It has become a primary technique for studying fracture in piezoelectric materials.
6.2. Indentation Strength

The IS technique to determine fracture toughness was suggested by Chantikul et al. [38], and is similar to the SCF test described earlier. The difference is that instead of removing the stress field due to the indentation, it is combined in a mathematical expression with the far-field stress as part of the driving force on the crack. Fracture strength can be measured in either uniaxial or biaxial flexure.

Chantikul et al. [38] derived the following expression for $K_c$:

$$K_c = \eta_v^R \left( \frac{E}{H} \right)^{1/8} \left( \frac{\sigma P^{1/3}}{s} \right)^{3/4}$$  \hspace{1cm} (31)

where $E$ is Young’s modulus, $H$ is the hardness, $\sigma$ is the fracture strength, and $P$ is the indentation load. Chantikul et al. [38] determined an average value for the constant $\eta_v^R = 0.59$.

The authors claim an accuracy of 30–40% for $K_C$; however, one should note that some materials, glasses in particular, deviate from the trend line, and caution must be exercised in assuming that fracture toughness values for glasses are within this range. There are other factors which can lead to errors in the determination of $K_C$ by this method. These include interactions between the radial and lateral cracks and microstructural effects, including anisotropy. At small indentation loads, one must make sure that failure began from the indentation site and not a pre-existing flaw in the surface or edge of the specimen.

One test to determine whether this method is applicable to a particular material is to logarithmically plot fracture strength as a function of indentation load. The slope of the line should be $-1/3$. However, as shown in Figure 13, deviations from the behavior predicted by Eqn (31) can occur.

Such deviations indicate that there are additional stresses in the material that have not been accounted for. Under no circumstances should a value of $K_C$ be taken from this regime. On the other hand, such deviations can be valuable in obtaining information regarding the influence of microstructural-level stresses on fracture.

One must also be aware, as with other tests, of the possible role of test environment in reducing the apparent value of $K_C$. Testing under inert conditions, e.g. in previously dried vacuum pump oil or nitrogen gas, and at fairly rapid loading rates will reduce the influence of environment.

The essentially same IS procedure is followed in the so-called “dynamic fatigue” test in which the strength of an indented flexural bar is measured over a range of loading rates. The slope of the strength versus stressing rate curve yields a value of $N$, the measure of the materials’ sensitivity to environmentally enhanced crack growth. This test is discussed in more detail in the chapter on Mechanical Reliability in this Handbook.

The IS procedure is also useful for investigating mixed-mode crack growth. Mixed-mode fracture can occur if the face of the flaw is oriented at an angle to the tensile stress. A schematic of such a situation is shown in Figure 14. Both $K_{II}$ and $K_{III}$ components of the stress field can act on the flaw, with $K_{II}$ being greatest at the tensile surface, and $K_{III}$ being largest at the maximum flaw depth. Effects of mode III loading generally will not play a major role in brittle fracture, and are generally ignored.

7. DOUBLE CLEAVAGE DRILLED COMPRESSION

The double cleavage drilled compression (DCDC) specimen was first suggested by Janssen [48], and consists of a square compression rod, usually having approximate dimensions of $150 \times 15 \times 15$ mm with a circular hole drilled through the center (as shown in Figure 15). When
loaded in compression, the presence of the hole modifies the applied compressive stress to create tensile stresses leading to the formation of vertical cracks emanating from the hole.

There have been a few analyses performed to determine the stress intensity factor as a function of the geometric parameters in the specimen [49,50]. The simplest of these obtained through a finite element analysis is given as follows [49]:

\[ K_I = \frac{\sigma r^{1/2}}{1.595 + 0.3530 \left(\frac{a}{r}\right)^{1/2}} \]

where \( r \) is the hole radius, \( a \) is the crack length, and \( \sigma \) is the stress applied to the ends of the specimen. Michalske et al. [49] found that over the crack lengths considered the results were within 10% of the values of fracture toughness obtained through other tests. Because the crack propagates into a decreasing stress field, what is measured is a crack arrest value rather than \( K_{IC} \).

The stability of the crack in terms of both its tendency to arrest and the fact that it grows uniformly on the centerline of the specimen make this configuration attractive for some crack growth studies such as the investigation of interfaces [51].

8. INTERPRETATION AND USE OF FRACTURE MECHANICS DATA

Because of the low fracture toughness of ceramics, the sizes of microstructural features, e.g. grain size, can be nearly equivalent to the flaw sizes which limit the strength of actual components. Such features can become a significant factor in the measurement and interpretation of crack growth. Therefore, a primary consideration in interpreting test results is the fact that many fracture mechanics tests require cracks much larger than the microstructure, e.g. DCB, DT, and SENB. Others such as indentation test methods employ much smaller cracks, of the order of the microstructure. The former tests are likely to yield significantly higher values of crack growth resistance due to microstructural mechanisms that apply only to the larger cracks. Therefore, fracture toughness values obtained with large cracks may not predict the behavior of small flaws, i.e. toughness may not translate into strength.

The presence of microstructure can lead to what are called rising \( R \) (crack resistance) curves, that is, the resistance to crack growth increases with increasing crack size. \( R \)-curves can arise due to a number of factors:

1. A transition from single crystal to polycrystalline-controlled fracture. This type of behavior is frequently experienced in relative large grain size material. Flaws within a grain will be governed by the fracture toughness of a single crystal of that material. This type of behavior is demonstrated quite clearly in Figure 16 for ZnSe [52,53]. The transition from single crystal to polycrystalline-controlled fracture can be very sharp [53] or broad depending on the crystal structure of the material and grain orientation.

2. A number of toughening mechanisms will apply to large cracks and not to small ones. For example, toughening mechanisms such as crack deflection [54,55], microcracking [56], transformation toughening [57], and grain bridging [58] depend on forces acting on the crack flanks. Therefore, until a steady state is reached, the longer the crack, the greater is the restraining force on it. However, even if \( R \)-curve behavior occurs in a material, increasing values of fracture toughness may not translate into higher strength [59].
In addition, the mode of fracture can vary with crack velocity. For example, it has been observed that crack growth in MgF₂ changes from intergranular to transgranular when the crack reaches a particular velocity (Figure 17) [60].

9. SUMMARY

This chapter summarizes the most important features of each of the leading fracture mechanics test procedures currently in use. The effect of the microstructure in ceramics is demonstrated to play an important role in fracture, and must be taken into account in the selection of a test procedure.

ACKNOWLEDGMENTS

The many helpful discussions with George Quinn and Jack Mecholsky are gratefully acknowledged. The aid of Nick Mecholsky in preparing many of the figures is also appreciated.

REFERENCES

[1] Chapter, this volume.
Chapter 9.6 Fracture Mechanics Measurements


Layered Ceramics

Raul Bermejo and Marco Deluca

Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Franz-Josef-Straße 18, 8700 Leoben, Austria

1. INTRODUCTION

The interest for the mechanical behavior of ceramic materials has been always motivated by their possible application as structural components, especially in the cases where properties such as high hardness, chemical stability, low density, and high strength, among others, are sought. Due to their brittleness, ceramics have been used for many decades as structural elements, but usually under compressive loading conditions. Nowadays, most of the new engineering designs need to withstand tensile stresses, which imply potential limitations for ceramics due to their low fracture toughness and the sensitivity of their strength to the presence of defects [1–3]. This is very well known for glass, which is one of the strongest manufactured materials when the surfaces are free from flaws (as for the case of glass fibers), whereas under ordinary conditions “glass” is a synonym for fragility. The brittle fracture of glasses and ceramics is a consequence of the material defects located either within the bulk or at the surface, resulting from the processing and/or machining procedures [4,5]. Under external applied stress, the stress concentration associated with those defects is the common source of failure for ceramic components. If each defect is considered as a crack or a potential source for crack initiation, then it becomes clear that the size and type of these defects determine the mechanical strength of the material [6]. The distribution of defects within a ceramic component yields a statistically variable strength which can be described by the Weibull theory [7,8]. Since flaws are intrinsic to processing and in most cases unavoidable, the reliability of ceramic components in terms of strength is associated with such flaw distribution.

In order to reduce both the defect population and the defect size, many studies have been devoted in the past to improve ceramic processing. The use of colloidal routes has, to some extent, enabled the reduction of the critical size of the flaw causing the failure of the material, thus allowing fabrication of advanced ceramics with relatively high strength [9]. In an attempt to reduce the level of uncertainty in mechanical strength and to overcome the lack of toughness of structural and functional ceramics, several processing routes have arose in the last two decades which do not utilize the conventional “flaw elimination” approach, but rather use the implication of energy release mechanisms to obtain “flaw tolerant” (strength reliable) materials, with improved fracture toughness [9–26]. In this regard, the outstanding mechanical behavior of hierarchical layered structures found in nature has inspired material scientists to reproduce or mimic some of these architectures for improving engineering designs [10,15,16,27–32], for instance, the extraordinary toughness and strength of mollusk shells (Figure 1), which are related to their fine-scale structure, namely a laminate of thin calcium carbonate crystallite layers consisting of 99% calcium carbonate (CaCO₃) and tough biopolymers, arranged in an energy-absorbing hierarchical microstructure [33]. The strength and toughness of such layered structures are significantly higher than those of their constituents [30,34].

In addition to the improvement of mechanical strength, there is another motivation for the production of layered ceramics. The demanding requirements for advanced devices involve in many cases the combination of several material classes (such as metals and ceramics) to fulfill the
performance in a given system. The unique mechanical and functional properties offered by ceramic materials make them good candidates for many advanced engineering applications (e.g. solid oxide fuel cells, piezoactuator and sensor devices, thermal barrier coatings, and conducting plates for wireless communications). The fabrication of components having two different materials is here a challenge from the viewpoint of not only structural integrity (i.e. mechanical resistance) but also functionality. Multi-layer piezoelectric actuators, for instance, are constituted by thin piezoceramic layers with interdigitated metal electrodes [35,36]. In order to ensure the device functionality, the propagation of cracks between adjacent electrodes must be avoided [37,38]. Another example is the case of low temperature co-fired ceramics (LTCCs), where the concept of co-sintering multilayer ceramic substrates, metal electrodes, and vias at relatively low temperatures (i.e. ca. 900 °C) has enabled the improvement of wireless communication systems (e.g. mobile phones and GPS technology) at relative low costs [39]. In LTCCs the mechanical reliability of the ceramic substrate (subjected to thermo-, mechanical-, and electrical loads) relies on avoiding crack propagation which would reduce or even completely hinder the performance of the microelectronic device.

In recent years, as a result of remarkable progress in terms of microstructural design and advanced processing [9,31,40,41], toughness and strength of structural ceramics have been increasingly enhanced by crack shielding from microstructure-related mechanisms [10,42–48]. A direct consequence of these energy-dissipating toughening mechanisms, which aim to reduce the crack driving force at the crack tip, is the development of an increasing crack growth resistance as the crack extends, i.e. R-curve behavior. This concept first arose from studies in metals and alloys in the 1960s and was later applied to single-phase, duplex, and laminar ceramics by numerous authors [11,48–50]. Particular attention has been paid to fiber, platelets, and layer-reinforced ceramics, where the better mechanical performance is associated with the second phase or layer addition as well as with the arrangement of the fibers or the layer assemblage, respectively [32,45, 51–54]. As an extension of this laminar ceramic/fiber-reinforced concept, multilayer designs have also been attempted in many ways aiming to improve both the resistance to crack propagation and the mechanical reliability of ceramic components [10,11,15,18,20,23,55–58]. This approach has been demonstrated to be more cost-effective than the former and more accurate in terms of tailoring mechanical requirements.

The design of composite materials using such multilayer architectures (e.g. ceramic composites such as alumina–zirconia and mullite–alumina among others) has been reported to exhibit increased fracture toughness, higher energy absorption capability, and/or noncatastrophic fracture behavior in comparison with their constituent (monolithic) materials. Among the various laminate designs reported in literature, two main approaches regarding the fracture energy of the layer interfaces must be highlighted. On the one hand, laminates designed with weak interfaces have yielded significant enhanced fracture energy (failure resistance) through interface delamination [10,14,25,30,59–68]. The fracture of the first layer would be followed by crack propagation along the weak interface or within the weaker interlayer. This is the so-called "graceful failure," which prevents the material from catastrophic failure — a fracture scenario where maximum applied load does not lead to unstable crack extension. In this case, the reinforcement mechanisms during fracture remind those from natural systems such as mollusk shells (cf. Figure 1). On the other hand, laminates designed with strong interfaces present crack growth resistance (R-curve) behavior through microstructural design (e.g. grain size and layer composition) and/or due to the presence of compressive residual stresses, acting as a barrier to crack propagation [15,16,20,21,24,55,58,69–71]. The increase in fracture energy in these laminates is associated with energy-dissipating mechanisms such as crack deflection/bifurcation phenomena. Within this context, a commonly used structural design is that associated with the presence of compressive residual stresses and/or phase transformations. Compressive stresses could develop in the laminate during cooling from sintering due to differences in elastic or thermal properties (Young’s modulus, thermal

**FIGURE 1** Fracture of a mollusk shell consisting of CaCO₃ brittle layers and tough biopolymers, arranged in a hierarchical microstructure. (Courtesy of Dr. Deville [33]).
expansion coefficient, etc.) between the layers [46, 56, 72]. The specific location of the compressive layers, either at the surface or internal, is associated with the attempted design approach, based on either mechanical resistance or damage tolerance, respectively. In the former case, the effect of the compressive residual stresses results in a higher, but single-value, apparent fracture toughness together with enhanced strength (the main goal) and some improved reliability [16, 49, 55]. On the other hand, in the latter case, the internal compressive layers are designed to rather act as a stopper to any potential crack growing from processing and/or machining flaws, at or near the surface layers such that failure tends to take place under conditions of maximum crack growth resistance [15, 20, 24, 58]. The utilization of tailored compressive residual stresses acting as physical barriers to crack propagation has succeeded in many ceramic systems, yielding in some cases a so-called “threshold strength,” i.e. a minimum stress level below which the material does not fail [15, 24, 26, 58, 70, 73, 74]. In such layered ceramics, the strength variability of the ceramic material due to the flaw size distribution in the component is reduced, leading to a constant value of strength. The selection of multilayer systems with tailored compressive stresses either at the surface or in the bulk is based on the end application, and is determined by the loading scenarios where the material will work.

The understanding of the conditions under which such reinforcement and energy-dissipating mechanisms occur and the influence of the layered architecture on the crack propagation must be pursued in order to improve modern structural and functional multilayer devices. The motivation of this chapter is to review the main approaches attempted in the field of layered ceramics to improve the mechanical properties of ceramic materials. Among the different possibilities to increase strength and toughness in multilayers, the use of residual stresses as key feature will be addressed aiming to provide some guidelines to design more reliable advanced ceramics.

2. RESIDUAL STRESSES IN LAYERED CERAMICS

In every case where dissimilar materials are sealed together through a relative strong interface and subsequently undergo differential dimensional change, stresses arise between the materials [75]. Hence, a particular challenge in the processing of ceramic laminates is to understand the nature of these residual stresses, especially if they are to be used to enhance their mechanical properties. In ceramic laminates, residual stresses can be due to different factors: some of them are due to intrinsic causes such as epitaxial, variations of density or volume, densification, and oxidation at the surface. Others are extrinsic such as thermal or thermoplastic strains developed during cooling or by external forces and momentums. Among them, the aspect most commonly referred to is the difference in the coefficients of thermal expansion (CTEs) between adjacent layers. During sintering of the laminate, it is considered that stresses between layers are negligible. However, as the temperature decreases, the differences in the CTE (\(\alpha_i\)) may promote a differential strain between layers. In addition to this differential strain, other strain differences, mainly due to phase transformations (\(\Delta \varepsilon_i\)) [46, 57] or reactions (\(\Delta \varepsilon_i\)) [72] inside one layer, should also be considered. As a result, the final differential strain between two given layers A and B after cooling may be expressed as

\[
\Delta \varepsilon = (\alpha_A - \alpha_B) \Delta T + \Delta \varepsilon_i + \Delta \varepsilon_r
\]

where \(\Delta T = T_{\text{ref}} - T_0\) is the difference between the reference temperature, i.e. the temperature at sintering where residual stresses are negligible, \(T_{\text{ref}}\), and the room temperature, \(T_0\).

2.1. Analytical Solution

For ideal elastic materials, neglecting the influence of the external surfaces (where strains may relax) and considering the laminate as an infinite plate the stress field can be determined analytically [12, 75, 76]. In each layer a homogeneous and biaxial residual stress state exists far from the free edges. The stress magnitude in each layer, \(\sigma_{\text{res},i}\), can be defined as

\[
\sigma_{\text{res},i} = \frac{E_i}{1 - \nu_i}(\bar{\alpha} - \alpha_i) \Delta T = \frac{E_i}{1 - \nu_i} \Delta \varepsilon_i,
\]

where \(E_i, \nu_i, \) and \(\alpha_i\) are material properties of the \(i\)th layer (i.e. Young’s modulus, Poisson’s ratio, and coefficient of thermal expansion, respectively). \(\Delta \varepsilon_i = (\bar{\alpha} - \alpha_i) \Delta T\) is the mismatch strain of the \(i\)th layer, where the coefficient \(\bar{\alpha}\) is given as an averaged expansion coefficient of the laminate:

\[
\bar{\alpha} = \frac{\sum_{i=1}^{N} E_i t_i \alpha_i}{\sum_{i=1}^{N} E_i t_i}
\]

with \(t_i\) being the thickness of the \(i\)th layer and \(N\) the number of layers.

Note that the reference temperature \(T_{\text{ref}}\) is, in practice, not easy to determine. It is always lower than the sintering temperature, since – if temperatures are reduced after sintering – the diffusion, which reduces the strain mismatch, does not stop abruptly but it becomes slower and slower. Generally, a normalization based on Eqn (2) and additional residual stress measurements have to be performed to determine \(T_{\text{ref}}\). For typical alumina-based or silicon-based ceramics, this stress-free empirical temperature is general between 1200 °C and 1300 °C [76].
When only two types of layer materials (A and B) are represented in the laminate (this is the case we will consider in the following), Eqn (3) can be written in the form

\[
\sigma = \left( \frac{E_A a_A}{1 - v_A} \sum_{i=1}^{n_A} t_{A,i} + \frac{E_B a_B}{1 - v_B} \sum_{i=1}^{n_B} t_{B,i} \right) \left( \frac{1}{1 - v_A} \sum_{i=1}^{n_A} t_{A,i} - \frac{1}{1 - v_B} \sum_{i=1}^{n_B} t_{B,i} \right).
\]

(4)

The magnitude of the residual stresses depends on the properties of A and B and on the ratio between the total thickness of the layers of type A \(T_A = \sum t_{A,i}\) and B \(T_B = \sum t_{B,i}\). The ratio of total thickness of the layer materials equals to their volume ratio: \(T_B / T_A = V_B / V_A\). It is interesting to note that the magnitude of residual stresses only depends on this volume ratio and not on the thickness of the individual layers \(i\) [77]:

\[
\sigma_{res,i} = f_i \left( \sum_{j=1}^{n_B} t_{B,j} / \sum_{j=1}^{n_A} t_{A,j} \right) = f_i (T_B / T_A)
\]

\[
= f_i (V_B / V_A).
\]

(5)

Finite element simulations of residual stress distributions in two symmetrical layered systems with the same volume ratio show the same magnitude of stresses in both cases, regardless of the combination of layer thicknesses of each material (see Figure 2). This is a very important aspect, which can provide the designer with more flexibility in order to tailor the disposition and thickness of layers when searching for an optimal design for a given level of residual stresses.

2.2. Limitations for Design

Although residual stresses can be a key feature to enhance the mechanical properties of many layered systems, some negative effects of high residual stresses should be considered. For instance, while compressive stresses are beneficial in acting as “shielding” mechanism against crack advance, tensile stresses will cause the cracking of the layer if they overcome its strength. A typical example is “tunneling cracks” that may appear at the surface of the tensile layers [78–81], and which can affect the structural integrity of the laminate (see Figure 3a). They can be avoided by lowering the level of tensile stresses in the corresponding layers.

Another important aspect is the free surface of the material. It is well known that stresses at the free surface of layered materials are different from those within the bulk. In the region far from the free edges (i.e. in an infinite plate), biaxial residual stresses parallel to the layer plane exist, and the stresses perpendicular to the layer plane are negligible [82]. Near free edges, however, the residual stress state is no longer biaxial since the edge surface must be traction-free. As a result, a stress component perpendicular to the layer plane appears at the free edge [82–87]. This stress has a sign opposite to that of the biaxial stresses in the interior. Hence, for a compressive layer sandwiched between two tensile layers, a tensile residual stress perpendicular to the layer plane exists near the free surface of the compressive layer. The amplitude of maximum tensile residual stress at the free surface is related to the biaxial compressive residual stress in the interior. Although such tensile residual stress decreases rapidly from the edge surface to become negligible at
a distance of the order of the compressive layer thickness, defects at the surface may be activated so that cracks may be initiated. An example are the so-called “edge cracks”, initiating from preexisting flaws, which can be encountered at the free edges of compressive layers (see Figure 3b). Preventing edge cracking is important for the structural integrity of the component and should be considered in the multilayer design, as it has been attempted by several authors [23, 86, 88].

A combined action to avoid both tunneling and edge cracks may be undertaken by selecting the level of residual stresses in the design so that the strength of the layers is not overcome. Such a limit for the magnitude of the residual stresses depends on each particular design and can be selected according to the properties of the employed materials. An example of such approach has been derived for alumina—zirconia-based multilayer ceramics [77]. It is assumed for both A-layers and B-layers the same elastic constants, fracture toughness, and strength, i.e. $E_A = E_B$, $v_A = v_B$, $K_{IC,A} = K_{IC,B}$, and $\sigma_{c,A} = \sigma_{c,B} = \sigma_c$. The residual stress difference between adjacent layers is defined as $\sigma_0 = \sigma_A - \sigma_B$, with $\sigma_A$ and $\sigma_B$ being the residual stress in A-layers (compressive) and B-layers (tensile), respectively, which can be expressed as

$$\sigma_A = \sigma_0 \frac{V_B}{V_A + V_B} \quad \text{(6a)}$$

$$\sigma_B = -\sigma_0 \frac{V_A}{V_A + V_B} \quad \text{(6b)}$$

In order to avoid edge cracks the strength of the layer, $\sigma_c$, seems to be a reasonable limit for $|\sigma_A|$, i.e. $|\sigma_A| \leq \sigma_c$. If tunneling cracks are to be avoided $|\sigma_B| \leq \sigma_c/2$ is a reasonable limit, as derived in Ref. [77]. Using Eqns (6a) and (6b) a limiting $\sigma_0$ to design laminates free of cracks can be derived:

$$|\sigma_0| \leq \sigma_c \left(1 + \frac{1}{V_B/V_A}\right) \quad \text{(7)}$$

$$|\sigma_0| \leq \frac{\sigma_c}{2} \left(1 + \frac{V_B}{V_A}\right) \quad \text{(8)}$$

These two limits (Eqns (7) and (8)) give restrictions on the magnitude of $|\sigma_0|$ to avoid edge cracks and tunneling cracks, respectively, and should be considered for design purposes. In Figure 4 both conditions are plotted as a function of $V_B/V_A$ for a typical characteristic strength value of $\sigma_c = 400$ MPa. Therefore, a region free of surface cracks can be estimated by tailoring the volume ratio between material A and B.

2.3. Experimental Determination of Residual Stresses

The magnitude of residual stresses within a material or component can be experimentally determined by either destructive or nondestructive methods. Mechanical techniques based on the use of a strain gauge are destructive in the sense that they require the component to be drilled or cut. Another method which is less invasive but still regarded as destructive is the use of micro-indentations (in general, Vickers imprints) [89, 90]. This procedure is based on the measurement of the crack lengths arising from the tip of the imprint in both stressed and nonstressed materials. The magnitude and profile of the residual stresses are directly associated with the measured crack length difference. In Figure 5a a scheme of indentation profiles in inner and outer layers of an alumina—zirconia laminate is reported, whereas the results are displayed in Figure 5b. It becomes clear that the level of stresses involved in the outer layers is considerably lower. This method proves to be a quick and low-cost alternative to more complicated analytical methods (see below). However, since the indentations are performed at the surface of the specimen, the use of this technique is limited only to the evaluation of surface stresses. This effect is displayed in Figure 6 as the result of FE calculations of the average residual stresses near surface and at the center of the sample. As could clearly be seen, notable differences are present between stresses in the outer and inner layers.

Nondestructive techniques for the measurement of residual stress in ceramic laminates generally are noncontact methods, namely they make use of an incident radiation (being X-rays, neutrons, or light) to probe the material surface in a position-resolved fashion. The penetration depth and the spatial resolution of the technique depend on the radiation wavelength and energy. In addition, the physical mechanism (especially the associated ease or difficulty in data interpretation) and the availability of the radiation source are very important aspects for the choice of the analytical technique.
The most common methods for residual stress measurements in multilayer ceramics are X-ray and neutron diffraction. Both methods are based on Bragg’s equation, which allows the accurate determination of atomic spacing and, consequently, elastic strains in crystalline materials. The main difference between these two methods (apart from the radiation source) is the fact that laboratory X-rays have a much shallower penetration depth (a few microns) and therefore allow only surface measurements, whereas neutron diffraction enables in-depth measurements up to some millimeters. Both techniques present significant challenges in case a position-resolved analysis is needed. In fact, generally, a neutron beam cannot be focused to a very small size (highest possible spatial resolution: ~200 μm on the material surface); on the other hand, traditional X-ray diffraction (XRD) methods involve rotation or tilt of the specimen, leading to a cumbersome experimental procedure. As a result, only macroscopic residual stresses are accessible by these techniques.

In the recent years, synchrotron XRD has emerged as a powerful analytical method for the study of microscopic residual stresses in ceramic laminates. Due to the high energy of synchrotron X-rays and the advanced collimation and focusing equipment available at common beamlines, this technique allows noncontact measurements with a very narrow spot size (micron to submicron), whereas the wavelength tuneability of radiation allows choosing the penetration depth inside the material. 2D and 3D macroscale residual stress measurements are made possible using special procedures such as the 3D diffraction microscopy (3DXRD), based on bidimensional detectors and sample rotation, or Laue microdiffraction, based on wavelength tuning.

An example of residual stress analysis in layered ceramics is constituted by the work of Leoni et al. on alumina–zirconia laminates. Figure 7 shows a through-thickness residual stress profile obtained with synchrotron XRD and the Laue microdiffraction method on two alumina and one alumina–zirconia layer. In the central...
alumina-zirconia layer, stress values belonging to both phases (calculated with isotropic elastic constants) are reported. The measurement was performed by varying the depth of the diffraction gauge volume. Examples of residual stress analyses in nanometer-sized layers also appeared recently in the literature [104], which prove the high suitability of synchrotron XRD also for the new trends in microelectronics and materials science.

The need for a synchrotron X-ray or a neutron source, however, makes these techniques seldom available and not always cost-effective. These difficulties could be overcome by another method capable of position-resolved residual stress measurement on the microscale: Raman spectroscopy. This technique is based on the Raman effect, namely the inelastic scattering of monochromatic light (a laser) by a crystal. Raman spectroscopy is noncontact and nondestructive, and common spectrometers are laboratory-size equipment [105]. The spatial resolution (spot size and penetration depth) depends on the wavelength of the incident laser, on the choice of the objective lens, and on the absorption properties of the investigated material [106–109]. In common ceramics such as alumina, a spot size of ~1 μm and a penetration depth of ~15 μm can be achieved [108], which makes the technique surface-sensitive in comparison to the aforementioned XRD and neutron diffraction. The interpretation of the relationship between the Raman spectral shift and stress (piezo-spectroscopy) is nontrivial [110,111], but has been well established in the past for a wide range of materials (silicon [112,113], zirconia [114], and silicon nitride [115–118], among others). In the case of alumina, the influence of residual stress on the luminescence signal of intrinsic Cr$^{3+}$ impurities in the ceramic body, induced by the laser and detected in the Raman spectrometer (photoluminescence, PL, henceforth), is used [87,111,119–122]. The physical mechanism in this case is different, in the sense that it is not based on the modification of the force constants of lattice vibrations (Raman piezo-spectroscopy), but rather on the shift in the position of the energy levels of Cr$^{3+}$ impurities under the influence of a macroscopic stress [111]. Generally, a convolution of the calculated stress values by a suitable stress model and the probe response function is done within the probe size, which leads then to a good approximation of the measured stress values [121]. An example of residual stress measurement performed with this method is provided in Figure 8 [122]; here the PL probe was scanned through alumina and alumina—zirconia layers, allowing to obtain different values of measured stresses for different layer configurations and thicknesses. It has to be noted that the stress values measured with the PL technique are generally lower than those predicted inside the layers [87]. This is associated with the free surface effect and also with the fact that the PL probe gives information only on the sum of principal stresses. Therefore, only mean stresses (trace of the stress tensor) can be measured at the surface.

In the recent years, alternative techniques emerged in the scientific community, which could be useful in the future for analyses in ceramic laminates. They are electron back-scattered diffraction (EBSD) [123] and positron annihilation lifetime spectroscopy (PALS) [124]. The first technique can be carried out in an electron microscope, and provides very local (nanometer-scale) strain measurements. It is however only a surface-based technique and the methodology to avoid charging effects in ceramic materials is currently under development. PALS has already been used together with nanoindentation in alumina—zirconia laminates to study the influence of stress on localized defects. The technique has a relatively high penetration depth (~100 μm), and thus could be a valid alternative to neutron diffraction measurements.

3. MECHANICAL BEHAVIOR

The response of a monolithic ceramic material to an external applied load can be characterized by its mechanical strength and its crack growth resistance. Both strength and toughness have proved to be enhanced using layered architectures with tailored properties (i.e. interface toughness, strength and toughness of each layer, composition and disposition of the layers, residual stresses, etc). Following the two main approaches in the design of ceramic
multilayers, systems with weak or with strong interfaces can show a different response to fracture.

3.1. Laminates with Weak Interfaces

Multilayer systems designed with weak interfaces or interlayers are aimed to prevent the component from catastrophic failure by guiding the crack along the interface or propagating the crack within the interlayer, and thus dissipating energy during the fracture process. The fracture process is based on the failure of the (stiffer) layers, and the strength of the system is associated with the mechanical resistance of such layers. Key works on this kind of multilayer material are those from Clegg et al. [10,14,125]. As an example, the fracture toughness of a silicon carbide monolithic ceramic is compared with that of a SiC multilayer containing graphite interlayers [125]. The so-called “graceful failure” of the material can be seen in the load—deflection curve in Figure 9. The failure of one layer does not imply the catastrophic failure of the entire component, raising the apparent fracture toughness from 3.6 to 17.7 Mpa m$^{1/2}$.

A special case is that of porous interlayers, where crack deflection within the layer can occur under bending if the layer has a minimum volume fraction of porosity [14,63,65,126]. The crack path (i.e., deflection within the interlayer or along the interface) is dependent on the composition of the interlayer [59]. Models to predict the deflection of cracks within weak interlayers of laminates loaded under bending have been developed based on energy criteria. This has enabled optimizing the fracture energy of such multilayer systems as a function of interface toughness, strength of the layers, and Young’s modulus [127—130].

The particular case of a crack deflecting along a weak interface between two materials (having different elastic and/or mechanical properties) was first studied by He, Hutchinson, and Evans based on the type of loading and on the properties of the layers and interfaces (i.e., fracture energy of interface and layers, elastic constants of the individual layers, etc.) [131—133]. The tendency of a crack to deflect along or penetrate into the next layer depends on the relations between the involved fracture energies (of the material layers A and B, $G_{layer}$, and of the interface $G_i$) and the relevant energy release rates (of deflecting and penetrating cracks, $G_d$ and $G_p$, respectively). This is also influenced by the combination of elastic properties of layers A and B, as described by the Dundurs parameters [134]. In Figure 10 a diagram showing the regions prone to crack deflection and to crack penetration is represented depending on the angle of crack propagation.

The curves given by the ratio $G_d/G_p$ limit both regions. When a crack propagates normal to the interface, it tends to

**FIGURE 9** Typical load—deflection diagram of a layered structure with weak interfaces. The steps show the failure of individual layers and the crack propagation along the interface. A schematic side view of a broken specimen showing crack deflection into the weak layers is shown in the inset. The failure of a monolithic is also shown for comparison.

**FIGURE 10** Diagram to predict crack penetration into or deflection along the interface in a multilayer structure. When the crack propagates straight toward the interface (left) crack penetration occurs. If crack bifurcation takes place, the low angle with which the crack approaches the interface favors interface delamination (right).
penetrate, whereas cracks propagating with an inclined angle may favor interface delamination, as shown in Figure 10. It has been derived that the type of applied load (i.e. tensile or bending), the presence of residual stresses, and/or the angle of the approaching crack are key features conditioning the crack path. This has also been validated experimentally in different systems [14,25,60,62,68,135], even for multilayers with relatively strong interfaces [136,137].

3.2. Laminates with Strong Interfaces

Multilayer architectures designed with strong interfaces can be subdivided into those with and without residual stresses in the layers. The mechanical behavior of the latter is enhanced by the different microstructures in the adjacent layers (i.e. composition, fine or large grain size, etc.), yielding crack branching or crack bridging as energy-dissipating mechanisms [50,138]. More significant, however, is the mechanical response of laminates with residual stresses. In this case, two multilayered designs, regarding the location of the compressive stresses either at the surface or within the bulk material, have been extensively investigated in order to tailor particular structural applications. Laminates with compressive stresses on the surface have proved to be useful for improving fracture strength as well as increasing wear resistance to contact damage [19,46,49,55,73,139–143]. On the other hand, if the compressive layers are internal, mechanical behavior enhancement is rather achieved in terms of flaw tolerance and energy-dissipation mechanisms occurring at fracture [11,15,20,21,24,70,78,79,144–149]. From this perspective, attainment of a threshold strength, i.e. a failure stress that is independent of the original processing or machining flaw size, is a reliable evidence of the potential effectiveness of this approach [15,20,24]. An example of the mechanical response of both configurations is shown in Figures 11 and 12, respectively. In Figure 11a the bending strength distribution of an alumina—zirconia multilayer with external compressive stresses is plotted in a Weibull diagram and compared with the strength corresponding to a monolithic alumina-based ceramic. Due to the compressive stresses in the outer layer (this layer is subjected to tensile stresses under bending), the strength is enhanced with respect to the monolithic material (without compressive stresses). A slight improvement in the Weibull modulus can also be observed. The fracture path in these specimens is shown in Figure 11b. The presence of compressive stresses slightly deviates the crack from propagating straight. However, the propagation takes place under unstable conditions. In Figure 12a indentation strength results of similar alumina—zirconia multilayers are plotted together with an alumina monolithic material. For the latter the typical strength dependence with critical flaw size (indentation crack) can be observed as for brittle materials; the larger is the indentation crack introduced, the lower is the failure stress. On the other hand, the strength on the multilayer remains constant regardless of the initial indentation flaw size. A minimum failure stress is obtained for the different crack lengths introduced (threshold

![Figure 11](image-url)
strength). This is associated with the initial growth of cracks during bending until they reach the compressive (stopper) layer as can be seen in Figure 12b. The failure of the structure occurs when the crack propagates through the compressive layer. In such a case, the fracture takes place through deflection and/or bifurcation of the crack (see Figure 12c), which enhances significantly the fracture energy of the system.

The design and use of multilayer systems with either external or internal compressive stresses should be motivated by the end application, as described in the Introduction. The design optimization of such structures in terms of strength and fracture resistance should also be regarded according to the property which is to be improved, as we will see below.

4. DESIGN GUIDELINES TO OPTIMIZE STRENGTH AND TOUGHNESS

Layered ceramics designed with strong interfaces have proved to increase the strength and toughness of monolithic ceramics by tailoring residual stresses in the layers. Much effort has been focused on the optimization of symmetrical laminates with a given layer thickness ratio between adjacent layers [26]. Recent research has shown that the distribution of layers of different materials can enhance significantly the fracture response of the laminate while maintaining a constant level of residual stresses [77]. Some design guidelines based on fracture mechanics criteria and experiments will be addressed in this section.

4.1. Fracture Mechanics Approach

The fracture criterion of brittle materials is described by the linear elastic fracture mechanics (LEFM) based on the well-known Griffith/Irwin equation [1]:

$$K \geq K_{lc}$$

where $K$ is the stress intensity factor and $K_{lc}$ is the fracture toughness of the monolithic material which can be experimentally determined using the standardized single-edge V-notch beam (SEVNB) method [150]. For an external applied stress, $\sigma_{appl}$, an applied stress intensity factor, $K_{appl}$, can be defined as

$$K_{appl}(a) = \sigma_{appl}Y\sqrt{\pi a}$$

where $Y$ is a geometric factor depending on the crack shape and loading configuration and $a$ is the crack length. The
crack propagation is possible if the stress intensity factor at the crack tip, $K_{\text{tip}}$, equals or exceeds the intrinsic material toughness. The resistance to crack propagation in monolithic materials is measured normal to the direction of applied stress (mode I). However, in multilayer ceramics, especially those designed with residual stresses, the stress intensity factor at the crack tip as a function of the crack length, $K_{\text{tip}}(a)$, can be given as the externally applied stress intensity factor $K_{\text{appl}}(a)$ plus the contribution of the residual stresses, given by $K_{\text{res}}(a)$. In these cases, an “apparent fracture toughness” as a function of the position of the crack within the layered structure is evaluated, which requires alternative approaches. In general, the effect caused by the residual stresses can be described by considering these stresses to be an additional external stress, thus adding a term $K_{\text{res}}$ to the stress intensity factor at the crack tip, $K_{\text{tip}}$, which reads

$$K_{\text{tip}}(a) = K_{\text{appl}}(a) + K_{\text{res}}(a)$$

(11)

Thus, solving Eqn (11) for $K_{\text{appl}}$, the Griffith/Irwin criterion described by Eqn (9) now becomes

$$K_{\text{appl}}(a) \geq K_{\text{lc}} - K_{\text{res}}(a) = K_R(a)$$

(12)

where $K_R(a)$ is the “apparent fracture toughness”. For $K_{\text{res}} < 0$, as it holds for the action of compressive stresses, $K_R(a) \geq K_{\text{lc}}$, what is called an increasing crack growth resistance curve (R-curve). This describes the “shielding” effect associated with the compressive stresses. If tensile residual stresses are acting, “anti-shielding” occurs and $K_R$ decreases with increasing crack extension.

The evaluation of the fracture toughness (or fracture energy) of layered structures has been attempted by theoretical means using different approaches. One general procedure for the determination of $K_{\text{res}}(a)$ is the finite element (FE) method. An FE model has to be created in order to compute the stress intensity factor at the crack tip under applied external stress for every position of the crack within the multilayer. Although there is, in general, a singularity problem at the interface, methods such as the J-integral can approximate the solution to the problem up to certain distance from the interface [151]. In order to describe the propagation of the crack along or through the interface, methods based on the LEFM have been developed such as the “finite fracture mechanics” approach, where the crack path near an interface between two different materials can be described [152]. Such methods can be used to describe the crack propagation resistance in linear elastic monolithic materials. However, the application to composites or layered structures combining different classes of materials (e.g. metal–ceramic and ceramic–polymer) with different elastic constants should be carried out with care. In this regard, a new approach is the configurational forces (CF), which considers a material inhomogeneity (e.g. second phase and layer) as an additional defect in the material (besides the crack) inducing an additional contribution to the crack driving force. This contribution is called the material inhomogeneity term $C_{\text{inh}}$. The thermodynamic force at the crack tip, denominated as the local, near-tip crack driving force $J_{\text{tip}}$, is the sum of the nominally applied far-field crack driving force $J_{\text{far}}$ (classical J-integral of fracture mechanics) and the material inhomogeneity term $C_{\text{inh}}$ [153–155]. This method allows, for instance, taking into account the contribution of the different compliance of the layers to the crack propagation resistance. For the case of ceramic laminates with layers having relatively similar elastic constants, an alternative method to account for the term $K_{\text{res}}(a)$ is the weight function (WF) approach. This allows us to calculate the apparent toughness $K_R(a)$ for an edge crack of length $a$ for an arbitrary stress distribution acting normal to the prospective fracture path as [156]

$$K_R(a) = K_0 - \int_0^a h(x, a)\sigma_{\text{res}}(x)dx$$

(13)

where $K_0$ is the intrinsic fracture toughness of each individual layer, $x$ is the distance along the crack length measured from the surface, $a$ is the crack length, and $h(a, x)$ is the weight function, which has been developed for an edge crack in a bar under different loading configurations [157].

The methods described above give very similar results and correlate with experimental findings for layered ceramics with residual stresses [77] (see Figure 13). Therefore, any of them can be used for $K$-value or $J$-integral calculations. Through comparison of the computational effort factors of the FE method, the CF method, and the WF method, it can be stated that the WF approach is much less time-consuming and thus is recommended when optimization processes of laminates are to be performed.

![FIGURE 13](image-url) Comparison of the apparent J-integral, $J_{\text{appl}}$, as a function of the crack length $a$, calculated using experimental as well as analytical and numerical methods, for an alumina–zirconia laminate with compressive residual stresses in the external layers [26,77,158,159].
This will be used below to compare and optimize the mechanical properties of layered ceramics with external or internal compressive stresses.

4.2. Optimal Laminates with External Compressive Stresses

In this section, the design optimization in terms of toughness and strength of an alumina-zirconia laminate with external compressive stress (ECS) is illustrated. Material properties used in this study are based on typical values for alumina-zirconia periodic laminates (i.e. all layers of material A have the same thickness; the same holds for material B) [23]. Hence, elastic properties are chosen as $E = 390$ GPa, $\nu = 0.22$, and coefficients of thermal expansion between layers as $\Delta \alpha = 1 \times 10^{-6}$ K$^{-1}$. In Figure 14 the qualitative behavior of ECS laminates is shown (the data correspond to laminates with $\sigma_0 = -590$ MPa).

For the determination of the apparent toughness, the WF method has been employed. Plotted are the R-curves versus the crack length parameter $Y \sqrt{\pi a}$ for the first two layers of two laminates having a different material volume ratio $V_B/V_A$ (i.e. for ECS1: $V_B/V_A = 7/3$ and for ECS2: $V_B/V_A = 9/1$). The geometric factor $Y$ is taken as 1.12 for an edge crack. Therefore, the stress magnitudes in the layers of the two laminates are also different (i.e. for ECS1: $\sigma_A = -413$ MPa, $\sigma_B = +177$ MPa and for ECS2: $\sigma_A = -531$ MPa, $\sigma_B = +59$ MPa). It is assumed that the thickness of the first two layers is constant ($t_A + t_B = 1$ mm) for both laminates. In the laminate ECS1, the compressive outer A-layer is thicker ($t_A$) than in laminate ECS2 but the magnitude of the compressive stresses is lower. The second (tensile) B-layer is thinner ($t_B$) in ECS1 but the stress magnitude is higher than in ECS2. The shielding effect of the compressive residual stresses causing a rise of the R-curve within the first layer can clearly be recognized. In the analyzed region (the first two layers) maximum shielding is always achieved at the first A/B interface of the external compressive A-layer, i.e. the length of the crack having a maximum shielding is $a = t_A$.

For this case, the influence of the compressive residual stresses in the outer A layer can easily be determined. It simply holds: $K_R(a)|_{a=t_A} = 1.12 \cdot \sigma_A \sqrt{\pi a}$ with $\sigma_A = \sigma_0 \cdot V_B/(V_A + V_B) \approx \sigma_0 \cdot t_B/(t_A + t_B)$. Maximum shielding is reached if $a = t_A$. Therefore, the peak toughness is

$$K_{R,\text{peak}} = K_0 - 1.12 \sqrt{\pi} \cdot \sigma_0 \cdot t_B \sqrt{t_A/(t_A + t_B)}$$

Note that $\sigma_0$ is a negative number. Therefore, for a given $\sigma_0$, the peak toughness is maximum, if $t_B \sqrt{t_A/(t_A + t_B)}$ is maximum. For $(t_A + t_B) = \text{constant}$, this value has its maximum for $t_B/t_A = 2$.

Figure 15 shows the peak toughness $K_{R,\text{peak}}$ versus $t_A/(t_A + t_B)$ for laminates with different values of $(t_A + t_B)$. It can be recognized that the maximum always occurs for $t_B/t_A = 2$, i.e. $t_A/(t_A + t_B) = 1/3$, with the toughness increasing with $t_A + t_B$.

Additionally to the optimization of the apparent toughness, the strength of the laminate must also be regarded. The strength depends on the size of the fracture origins (the flaws where fracture starts), considered as the depth of the through-thickness edge crack. The typical range of flaw sizes occurring in a ceramic material depends on the processing conditions. For technical state-of-the-art materials, the typical size of a volume flaw is about 30 $\mu$m (about 10 $\mu$m for surface flaws) [8]. A typical range of flaw origins is indicated in Figure 14 (shaded bar). However, larger flaws may also occur, causing the strength scatter in that type of ceramics [3,160]. Indeed, in a laminate system, no processing flaws larger than the layer thickness ($t_A$) can occur. In order to interpret the

---

**FIGURE 14** Apparent fracture toughness of laminates designed with external compressive stresses. The peak toughness, $K_{R,\text{peak}}$, depends on the thickness of the first layer. The range of typical defect sizes is represented as a shaded bar.

**FIGURE 15** Peak toughness as a function of the thickness ratio $t_A/t_A + t_B$ for different total thicknesses of the first two layers $t_A + t_B$. A maximum toughness value is obtained independent of the $t_A + t_B$ selected.
implications of the design in strength, the dependence of
\(K_{\text{appl}}(a)\) on the crack length parameter 1.12\(\sqrt{\pi a}\) is represented in Figure 14 as lines through the origin. Lines with
different slopes refer to different values of applied stress
(lines 1–4). Hence, for an applied stress intensity factor
corresponding to lines 1 and 2 and for cracks having
a length in the shaded bar, it holds \(K_{\text{appl}} < K_R\) for ECS1 as
well as for ECS2. This indicates that the driving force is too
small to extend these cracks. At higher stresses (line 3) and
for the largest flaw in the laminate ECS1, \(K_{\text{appl}} = K_R\) holds.
This crack (and larger cracks, if they exist) may propagate
but the smaller cracks may not. In the laminate ECS2
cracks having a length corresponding to the shaded bar are
still too small to propagate. If we increase the applied stress
to the slope of line 4, the Griffith criterion is also fulfilled
for the largest flaw in the shaded area for ECS2. In that case
the right end border of the shaded bar corresponds to the
A/B interface, and the corresponding crack length will be
t_A. This case defines a lower limit (a threshold value) for the
strength of ECS2. It is interesting to note that although
laminates ECS1 has higher peak toughness than ECS2, the
strength of ECS2 is higher than that of ECS1.

A comparison of the strength of the monolithic ceramic
(material A) and the laminates is also possible. Since
the monolithic ceramic has no increasing R-curve, its fracture
toughness \(K_c\) corresponds to the horizontal dashed line in
Figure 14. For the flaws in the shaded area, the lowest
strength value of the monolithic material is given by line 1.
It is obvious that the strength of both laminates is signifi-
cantly higher than that of the monolith.

In general, the size of the largest possible processing flaw
in a laminate is limited to the thickness of the outer A-layer.
Then the threshold for the strength depends on the height of
the peak toughness and on the layer thickness, as given below:

\[
\sigma_{\text{th}} = \frac{K_{R,\text{peak}}}{1.12\sqrt{\pi t_A}} = \frac{K_c}{1.12\sqrt{\pi t_A}} - \sigma_0 \frac{t_B}{t_A + t_B} \tag{15}
\]

The first term of Eqn (15) increases as \(t_A\) is reduced. It is
obvious that this threshold stress can be increased by
decreasing the layer thickness, but, for technological
reasons, laminates with layers thinner than 5–10 \(\mu\)m can,
today, hardly be processed. Then the contribution of the first
term to the threshold can reach the characteristic strength of
the material if the cracks are as large as the thickness of the
first layer. The second term is positive and, for \(t_A << t_B\), it
reaches \(\sigma_0\), which can be of the order of magnitude of the
characteristic strength of material A. Compared to mono-
lithic ceramics where failure may occur at any stress, the
occurrence of a threshold stress causes a significant increase
in mechanical reliability (at stresses lower than the
threshold) and can thus be used for safe design.

In ECS laminates the key parameter is the thickness of
the first layer (with compressive stresses), which should be
as thin as possible but thick enough to contain the largest
processing flaws. This can be achieved by designing non-
periodic laminates where the thickness of the layers in the
center of the architecture corresponding to material A can
compensate the small thickness of the very first layer in
order to tailor the residual stresses (associated with \(V_B/V_A\)),
which will maximize the peak toughness predicted by Eqn
(14) (see Ref. [77] for more details).

4.3. Optimal Laminates with Internal Compressive Stresses

For ICS laminates (internal compressive stress, the outer
layer is under tension) simple analytical solutions as for
ECS laminates do not apply and the apparent toughness
(the R-curve) has to be determined using approximation
methods such as the WF method. The apparent toughness
of a periodic alumina–zirconia ICS laminate according to
Eqn 13 is shown in Figure 16. The R-curve is represented as
a function of the crack length parameter 1.12\(\sqrt{\pi a}\) in
the region of maximum shielding (peak toughness), which is
achieved for a crack length of \(a = t_A + t_B\). Due to the
tensile residual stresses the R-curve decreases in the outer
A-layer and increases again in the following B-layer, which
contains compressive residual stresses. In this kind of
lamine, two regions can be distinguished as a function of
the flaw (crack) size. For very small flaws having a size in
region I unstable crack propagation occurs, which causes
catastrophic failure. This happens at relatively high applied
stresses (see the slope of line 1). On the other hand, larger
cracks in region II first pop in to an even greater size if
stress is applied (section of line 3 with the decreasing
R-curve) and then stop (intersection point of line 3 with the
increasing part of the R-curve in layer B). This is caused by
the shielding effect of the compressive layer. If the stress is

![FIGURE 16](image-url)
further increased, stable crack growth up to the crack length of \( a = t_A + t_B \) occurs. Line 2 intersects the R-curve of the laminate at a stress level at which these cracks become unstable again and where catastrophic failure occurs. This determines a threshold for the strength of the laminate, which is given by: \( \sigma_{th} = K_{R, peak}/1.12 \sqrt{\pi (t_A + t_B)} \). Note that only the first two layers of ICS laminates are important for the assessment of the crack resistance of such a laminate. Considering \( t_A + t_B = \) constant an optimal design for maximal shielding can be found. Figure 17 shows the peak toughness \( K_{R, peak} \) versus \( t_A/(t_A + t_B) \) for laminates with different values of \( t_A + t_B \). It can be seen that the maximum always occurs for \( t_A/(t_A + t_B) \approx 0.8 \) (which corresponds to \( t_B/t_A \approx 0.25 \)) and the toughness values also increase with \( t_A + t_B \). Similar behavior can be observed for the threshold strength, which reaches its maximum for the same ratio \( t_B/t_A \approx 0.25 \) and increases with \( t_A + t_B \). In order to enhance the threshold strength one could also increase the number of layers. However, the peak toughness would be significantly reduced. In this regard, improvements may still be possible for nonperiodic laminates trying to optimize the first two layers and then tailoring the internal structure in order to accommodate the residual stresses.

In Figure 18 the apparent fracture toughness of a periodic ICS laminate (ICS1) is compared to that of a nonperiodic one (ICS2), which has been designed with a thin first (tensile) layer and a thicker second (compression) layer. Both laminates have the same tensile and compressive residual stresses in the layers (the same \( V_B/V_A \) ratio). It can be observed that by decreasing the thickness of the first A layer the threshold strength can be improved and by increasing the thickness of the next layer (layer B) the peak toughness is also enhanced.

Summarizing, the optimal design for ICS laminates with nonperiodic layers lies in the combination of both concepts: (i) first make the tensile layer (A) as thin as possible and (ii) the compressive layer (B) as thick as possible. This design ensures a low decrease of \( K_R \) in the first thin tensile layer and a high increase of \( K_R \) within the second thick compressive layer. When both steps are adopted in a unique nonperiodic design, the mechanical properties of the system can be significantly enhanced with respect to monolithic materials.

4.4. Search for New Design Concepts

The potential of layered designs to improve the mechanical properties of ceramics is based on the capability of the layers to arrest or deflect potential cracks in the material as well as to increase the strength reliability of the system through tailored residual stresses. The role of the interface in the fracture process must also be taken into account. For instance, the combined effect of laminates with residual stresses and at the same time relatively weak interfaces can significantly enhance the resistance to crack propagation in the material [161]. This has been recently shown in an alumina—zirconia layered system designed with high compressive residual stresses. The magnitude of such stresses and the thickness of the compressive layer induced crack bifurcation during bending, thus increasing the fracture energy of the system. Additionally, the low bifurcation angle, the different elastic constants of the layers, and the interface fracture energy could also promote interface delamination (see Figure 10). The combination of layers with high compressive residual stresses and weak interfaces can be a promising strategy to combine the benefits of both approaches.

5. OUTLOOK

The production and implementation of layered ceramic structures in advanced material science are either motivated by the need for a composite material whose mechanical
properties surpass those of its constituents or driven by the material’s special applications. The concept of layered ceramics and functionally graded materials allows in fact tailoring of the surface and bulk properties of advanced engineering components with the purpose of enhancing their structural integrity as well as adding multifunctionality, which would translate into higher efficiency and better performance of these components.

Vast industrial sectors involving, for instance, aeronautics, biomedicine, communications, and automotive (among others) could benefit from the development of such advanced materials. Specific examples of applications are thermal barrier coatings, which provide thermal protection for gas turbine blades and combustion chambers of aeroplanes, dental crowns and hip prostheses for biomedical replacements, multilayer stacks in actuators for a reduction of emissions and the better performance of diesel engines, and hard layers in cutting tools. Also in the communication industry the development of reliable layered composites suitable for ultra-THz filters for wireless communications (e.g. mobile phones) is becoming necessary because the increasing demand for bandwidth is pushing the wireless short-range network into the THz frequency. Another example is the miniaturization of SOFCs (e.g. to be used as hybrid batteries in laptops or PDAs), where the performance of the system relies on the structural integrity of the multilayer structure.

The replication of architectural features that are found in nature, at the micro- and nanoscales, into real macroscale structural and functional engineering materials at a reasonable cost is expected to be the future leading concept for layered ceramics. This is generally due to the fact that production facilities for such materials are already available or could be easily scaled up (i.e. the tape-casting process). The main challenge regarding these structures resides in the careful tailoring of all materials parameters in order to induce crack deflection and graceful failure. In addition, it is necessary to build these high toughness architectural features by using more durable materials since they should possess additional functional and/or structural properties as, for example, high-temperature stability and specific functionality (i.e. metal–ceramic composites in the case of piezoelectric actuators). Another challenge is the integration of materials with different structures and phases, such as the multilayer packages for microelectronics, which include single crystalline and polycrystalline silicon, amorphous polymer layers, and metals.

The use of advanced ceramic-based multilayer architectures in engineering systems will be feasible when the mechanisms responsible for the outstanding mechanical behavior of natural systems would be completely understood. The use of bio-inspired multilayer structures on systems of different kinds (e.g. ceramic–ceramic and ceramic–metal) should contribute to material innovation designs within the field of material science and engineering for the next decades.

REFERENCES


Tochino S, Pezzotti G. Micromechanical analysis of silicon nitride.


Chapter | 9.7 Layered Ceramics


Chapter 9.8

Environmentally Enhanced Fracture of Glasses and Ceramics

Stephen W. Freiman
Freiman Consulting, 8212 Inverness, Hollow Terrace, Potomac, Maryland 20854, United States

1. INTRODUCTION

This chapter describes the phenomenon of environmentally enhanced crack growth in brittle, inorganic materials, i.e., glasses, single crystals, and polycrystalline ceramics. Otherwise known as stress corrosion, environmentally enhanced crack growth causes cracks to grow slowly over long periods of time, and complicates the selection and design of components built from such materials because it can lead to time-delayed failure. The chapter focuses on the mechanisms behind this phenomenon. The use of this information in the design of brittle structures is detailed in Chapter X in this volume.

The existence of delayed failure in glass was first reported in 1899 by Grenet [1], who fractured glass plates in three-point flexure, applying the load by hanging a bucket from their center-line and gradually filling the bucket with water at varying rates. He observed that by decreasing the loading rate from 0.25 kg/min to 0.25 kg/h a 50% decrease in breaking stress was obtained. Grenet also observed that specimens left under a constant load survived for as long as 5 days before fracturing.

In this chapter are presented some of the early studies of delayed failure, including attempts to devise a model. However, the major portion of the chapter is focused on describing the mechanisms of environmentally enhanced crack growth based on fracture mechanics tests on glass, including effects of environmental and compositional variations. The effect of longer range order and crystal bonding on environmentally enhanced crack growth behavior is discussed based on studies performed on single crystals. The final portion of the chapter is devoted to attempts to understand the fundamental mechanism through quantum mechanical calculations.

2. EARLY STUDIES

Before the advent of fracture mechanics, measuring the time to failure of a flexural specimen as a function of the applied load was the only method by which the phenomenon of environmentally enhanced crack growth could be approached. It was observed that the stress required to cause a specimen to break decreased with length of time under load (Figure 1).

Investigations of time-to-failure were carried out on glass because it is both homogeneous and isotropic. The forms of the expressions used to fit the time-to-failure data were strictly empirical. The lack of an ability to directly observe the moving crack coupled with the statistical nature of brittle fracture severely limited the development of an understanding of the process. Nonetheless, there were several important findings.

The first indication that water is a key to delayed failure was reported by Milligan [3] who observed that the strength of glass in dry environments was greater than that tested in the presence of water. Black [4] discussed the fact that if glass is held under load, and the load subsequently removed, the

Chapter Outline

1. Introduction 753
2. Early Studies 753
3. Direct Crack Growth Studies 754
4. Mathematical Modeling 761
5. Polycrystalline Materials 761
6. Summary 762
7. What is yet Unknown 762
Acknowledgments 762
References 763
strength after loading is not measurably degraded. We now know that this occurs because flaws grow only slightly until very near the point of catastrophic failure [5].

In 1942, Preston [6] made the observation that the rate of strength decrease during static fatigue is temperature dependent, being small at temperatures less than 0 °C, maximizing at ≈ 200 °C, and then decreasing at temperatures of ≈ 300 °C and greater. Preston interpreted these data in terms of the thickness and chemical activity of the water layer absorbed on the glass surface.

Shortly thereafter Orowan [7] made a significant contribution to the understanding of delayed failure. He used data on the cleavage of mica in both vacuum and air to show that the strength measured in vacuum will be 3.5—4.5 times that tested in moist air. Further, he suggested that the rate of strength decrease with time is due to crack deepening, with the rate of growth governed by the diffusion rate of the absorbed water into the crack.

In 1947, Gurney [8] presented thermodynamic concepts to explain moisture-enhanced fracture. He suggested that the rate of crack growth is controlled by the rate at which thermal motions overcome energy barriers. Even more significantly, Gurney suggested that, “Due to concentration of strain energy, the material at the end of the crack has a much higher free energy than normal unstressed glass, and is therefore much more chemically active. Atmospheric attack will result in the formation of a complex of glass and atmospheric constituents. The crack will extend continually if the strength of this complex, during or after its formation, is less than the load imposed on it.” The molecular model of crack growth described later indicates that this hypothesis was correct.

Chemical reaction rate theory was applied to crack growth in glass by Gibbs and Cutler [9] in 1951, and two years later by Stuart and Anderson [10]. In these papers, it was assumed that crack motion is due to the sequential rupture of bonds at crack tips. It was assumed that the free energy for the rupture process was enhanced by the stress at the crack tip. Consequently, crack velocity increases with crack-tip stress. Stuart and Anderson used the equations for crack velocity as a function of applied stress to calculate the time to failure. Their prediction agreed with much of the experimental data available at that time.

In the 1960s Charles and Hillig [11] published work expanding on the concept that corrosive processes at the crack tip are responsible for its slow growth under stress. The essential hypothesis in the Charles and Hillig model is that the contour of the flaw changes under the combined effect of stress and suitable chemical environment. At sufficiently high stresses, Charles and Hillig suggested that flaw sharpening occurs, leading to an increase in the stress concentration at its tip and crack growth. Under steady-state flaw growth, the sharpening at the tip is balanced by rounding due to corrosion. When the stress decreases to a value termed the stress corrosion limit, flaw rounding occurs in preference to sharpening, leading to crack arrest. There has been no confirming experimental data regarding this change in crack-tip shape.

3. DIRECT CRACK GROWTH STUDIES

The emergence of fracture mechanics specimens containing macroscopic, observable cracks permitted investigators to collect crack growth data directly by measuring growth rates extending under a known stress field. The first systematic study was carried out by Wiederhorn [12], who used double cantilever beam specimens to measure crack extension as a function of applied force in soda-lime-silica glass, using an optical microscope to follow the crack motion. The principal output from this study is shown in Figure 2 in which crack velocity, $V$, is plotted as a function of the applied stress intensity factor, $K_I$, for atmospheres in which the relative humidity was varied from 0.017% to 100%.

Wiederhorn identified multiple mechanisms of crack growth, as evidenced by the change in slope of the $V-K_I$ curves. These mechanisms are discussed in the following sections.

3.1. Region I

Region I is the regime of greatest significance both from a mechanistic viewpoint as well as from an engineering standpoint, because growing cracks spend most of their lifetime in this regime. Wiederhorn [12] assumed that in this regime the crack velocity, $V$, is proportional to the rate of reaction of the bonds at the crack tip. Based upon chemical rate theory approach described by Gibbs and Cutler [9] and Stuart and Anderson [10], he derived the following expression relating crack velocity to the crack-tip stress$^*$:

$$V = V_0a_{H_2O}\exp\left[\frac{-E^* + bK_I}{RT}\right]$$

$^*$ The $K_I$ dependence in Eqn 1 is not necessarily unique. Equations of a form containing $(K_I)^2$ or $1/K_I$ could also be made to fit the experimental data.
where \( V_0 \) is a constant, \( a_{H_2O} \) is the activity, or partial pressure, of water, and \( E^* \) contains all of the nonstress terms:

\[
b = \left( \frac{\pi d}{C_0} \right)^{1/2} \Delta V^*
\]

(2)

\( d \) has units of (length) and \( \Delta V^* \) is the activation volume for the reaction, namely the difference in volume of the reacting species between the initial state and the activated complex. The physical basis for the length term is unknown, but it is possible that \( d \) is related to the size of a zone at the crack tip within which the stress-enhanced reaction takes place. \( d \) could presumably vary from one material to another because of changes in bonding and structure.

Crack growth in nonaqueous liquids such as straight chain alcohols also depends on the partial pressure of water in the liquid, i.e., its relative humidity (Figure 3) [13]. In fact, no liquid can be considered to be completely dry because the partial pressure of water in the liquid will be in equilibrium with water in the air.

Wiederhorn and Bolz [14] showed that different glass composition gives rise to shifts in the crack growth curves in water (Figure 4). Note that in some glasses, e.g., soda lime and borosilicate, there is a downturn in the curves suggesting a limit to crack growth. Figure 5 showing crack
growth in a range of alcohols [13] for the glasses listed in Table 1 further illustrates the dependence of crack growth on the chemical composition of the material; the shifts in composition are apparent in the other regions of crack growth as well. These shifts in V–KI curves do not occur in any systematic way and are not predictable from knowledge of the glass chemistry.

Crack growth exponents for a number of other glasses are given in Table 2.

More crack growth data for glasses in various environments are summarized in Ref. [25].

The pH of aqueous solutions also plays a significant role in determining the slope of the V–KI curves [26] (Figure 6). The smaller slopes of the high pH, basic solutions, indicate that crack growth occurs more readily under these conditions than low pH, acidic, solutions. Later work [27] showed that, not only the pH, but also the accompanying counter-ions, especially Li⁺, could play a role in determining both the slope and position of the Kᵢ–V curve in basic solutions. The role of Li⁺ is especially noticeable, since in solutions of high pH, the Li⁺ ions form a strong chemical association with hydroxide, thereby limiting its effectiveness in promoting crack growth.

Eqn 1 suggests that there should be a temperature dependence to crack growth rates at a given stress intensity. This dependence has been demonstrated for a number of glasses in vacuum [28] and in straight chain alcohols [29]. Suratwala and Steele [30], in an excellent review of the topic, showed that crack growth

![FIGURE 5 Effect of changes in glass composition on crack growth in octanol. (Freiman [13])](image)

**TABLE 1 Glass Compositions Used in Study of Straight Chain Alcohols [13]**

<table>
<thead>
<tr>
<th>Glass No.</th>
<th>Major components (Wt %)*</th>
<th>SiO₂</th>
<th>Na₂O</th>
<th>CaO</th>
<th>BaO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (commercial)</td>
<td>72.0</td>
<td>14.0</td>
<td>7.0</td>
<td>2</td>
<td>69.7</td>
</tr>
<tr>
<td>4</td>
<td>59.0</td>
<td>12.6</td>
<td>28.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>38.8</td>
<td>60.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2 Crack Growth Exponents, N, for Glasses and Glass-Ceramics**

<table>
<thead>
<tr>
<th>Material</th>
<th>Environment</th>
<th>Technique</th>
<th>N</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO₂</td>
<td>air</td>
<td>DF</td>
<td>38</td>
<td>[15]</td>
</tr>
<tr>
<td>H₂O</td>
<td>DCB</td>
<td>DF</td>
<td>30</td>
<td>[17]</td>
</tr>
<tr>
<td>SiO₂-TiO₂</td>
<td>air; H₂O</td>
<td>DCB</td>
<td>25</td>
<td>[17]</td>
</tr>
<tr>
<td>(CGW 7971)</td>
<td>H₂O</td>
<td>DF</td>
<td>27</td>
<td>[18]</td>
</tr>
<tr>
<td>Na₂O-CaO-SiO₂</td>
<td>air; H₂O</td>
<td>DCB 16-21</td>
<td>[18]</td>
<td></td>
</tr>
<tr>
<td>Al₂O₃-B₂O₃-SiO₂</td>
<td>H₂O</td>
<td>DCB 25</td>
<td>[14]</td>
<td></td>
</tr>
<tr>
<td>(CGW 7740)</td>
<td>H₂O</td>
<td>DF 33</td>
<td>[19]</td>
<td></td>
</tr>
<tr>
<td>Na₂O-2SiO₂</td>
<td>H₂O</td>
<td>DCB 21</td>
<td>[20]</td>
<td></td>
</tr>
<tr>
<td>glass-Ceramics</td>
<td>air</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluoride (e.g. ZBLAN)</td>
<td>air</td>
<td>DF 75</td>
<td>[22]</td>
<td></td>
</tr>
<tr>
<td>Li₂O-2SiO₂</td>
<td>H₂O</td>
<td>DCB 11</td>
<td>[21]</td>
<td></td>
</tr>
<tr>
<td>As₂S₃</td>
<td>air</td>
<td>DCB 24-28</td>
<td>[24]</td>
<td></td>
</tr>
<tr>
<td>Li₂O-Al₂O₃-SiO₂</td>
<td>H₂O</td>
<td>DF 24-28</td>
<td>[24]</td>
<td></td>
</tr>
</tbody>
</table>

DCB = double cantilever beam; DF = dynamic fatigue; ULE = ultra-low expansion; CGW = Corning Glass Works

*Much of this data was plotted based on the expression V = AKᵢᴺ, so the slope reported is N.*
in vitreous silica, in contrast to other glasses, has a negative temperature dependence. That is, crack growth rates decrease with increasing temperature. The exact reason for this anomalous behavior is still unclear.

### 3.2. Region II

In region II, crack velocity is nearly independent of $K_I$. Wiederhorn [12] proposed that as the crack increases in velocity, a water-depleted zone forms at the crack tip through which a water molecule must diffuse to react with the crack-tip bonds. Thus, the diffusion rate becomes the limiting step in the process.

This model was quantified by Wiederhorn et al. [31], who derived an expression governing this behavior based on the Stokes–Einstein relation.

$$V = \frac{0.0275kT}{(6\pi r\delta)(X_0/\eta)} \quad (3)$$

where $r$ is the radius of water molecule, $\delta$ is the thickness of the boundary layer, and $X_0/\eta$ is the ratio of concentration of water to solution viscosity. As seen in Figure 7, this expression fits the experimental data reasonably well.

### 3.3. Region III

In his original paper, Wiederhorn [12] showed that the steep portion of the curve (region III) was independent of the partial pressure of water. Later work [13] demonstrated that this regime was affected by other liquids (Figure 8).
Wiederhorn et al. [31] provided a theoretical underpinning to these observations. As noted earlier, the chemical interaction model for environmentally enhanced crack growth; the slopes of the crack growth curves in regions I and III are governed by the activation volume, \( \Delta V^* \), of the crack tip chemical reaction. Wiederhorn et al. argued that in the absence of water, there is an electrostatic contribution to the activation volume. They showed data to indicate that the slope of the crack growth curve in region III is dependent on the dielectric constant of the test media, lending credence to this hypothesis.

3.3.1. Low-Velocity Regime

As noted earlier, the crack growth curves for soda-lime and borosilicate glass are quite steep at small \( K_I \) (Figure 4). At some lower bound on \( K_I \), crack arrest takes place. Michalske [32] was able to demonstrate actual arrest below a \( K_I \) of 0.25 MPa\( \text{m}^{-1/2} \) in soda-lime glass by showing that if the specimen was held at a \( K_I < 0.25 \), and the load then raised, a period of time was required for the crack to reach the velocity associated with that value of \( K_I \). Gehrke et al. [33] clarified this behavior by showing that glasses containing mobile cations (\( \text{Na}^+, \text{K}^+, \text{Li}^+ \)) exhibit thresholds, whereas glasses containing no mobile ions do not. They attributed this behavior to shielding stresses resulting from ion exchange close to the crack tip, as described by Michalske and Bunker [34].

Simmons and Freiman [20] investigated crack growth in much less durable binary alkali silicate glasses (Figure 9) and observed an entirely different phenomenon, namely a plateau in crack velocity at \( 10^{-8} - 10^{-9} \) m/sec. These crack growth curves were strongly affected by the presence of the particular alkali ions in solution. They also showed that once the crack was growing at these lower velocities, and the \( K_I \) was raised, it took close to one hour before the crack stabilized at the velocity typically associated with that load. Similar observations were made by Gehrke et al [33] and by Michalske and Bunker [34]. It appears that stresses generated by ion exchange are also responsible for this behavior.

There are data to suggest that water in the form of an \( \text{H}_2\text{O} \) molecule can diffuse into a glass under the influence of the large crack-tip tensile stress [35]. How this water affects the crack growth behavior is still being debated. Recent work by Wiederhorn et al. [36] suggests that this water can cause swelling at the crack tip, and so retard crack motion.

3.4. Molecular Models of Crack Growth

While it was known for many years that water is the key molecular constituent that leads to crack growth, why water should behave in this way was not understood. It was Michalske and Freiman [16] who demonstrated that it is the electronic structure of the water molecule which enables it to react with the strained silica structure. The reaction pathway is shown schematically in Figure 10, which depicts a strained silicon–oxygen bond.

The key is the fact that a water molecule can donate both a hydrogen ion and two lone-pair electrons from the unbonded atomic orbitals on the oxygen atom to the reaction. The three steps in the bond rupture process are:

1. A water molecule attaches to a bridging Si–O–Si bond at the crack tip. The water molecule is aligned by
hydrogen bonding with the $O_{\text{bridging}}$ and interaction of the lone-pair orbitals from $O_{\text{water}}$ with Si.

2. A simultaneous reaction occurs in which a proton transfers to the $O_{\text{br}}$ and an electron transfers from the $O_{\text{w}}$ to the Si. During this step of the reaction the original bridging bond between $O_{\text{br}}$ and Si is destroyed.

3. Rupture of the hydrogen bond between $O_{\text{w}}$ and transferred hydrogen occurs to yield Si–O–H groups on each fracture surface.

This model has some interesting implications. First, there is no requirement for prior dissociation of the water molecule, nor must any reaction products be formed. Second, the model suggests that environmental species other than water should also enhance crack growth if the species possesses structural and bonding features similar to water, i.e., proton donor sites at one end and lone-pair orbitals at the other.

Michalske and Freiman demonstrated that molecules such as ammonia, hydrazine, and formamide also cause subcritical crack growth in vitreous silica, Figure 11. All possess proton donor sites and lone-pair orbitals in juxtaposition. A key point in the experimental verification of the effectiveness of these molecules is the demonstration of the absence of a region II plateau. A plateau would have indicated that dissolved water, rather than the liquid itself, was causing crack growth. Using this criterion, Michalske and Freiman showed that molecules such as carbon monoxide and acetonitrile, which do not have lone-pair orbitals opposite proton donor sites, do not enhance crack growth in glass.

Michalske and Bunker [37] later proposed that the enhanced reactivity of an Si–O bond was a result of deforming the ring structure in silica. In this same paper, they suggest that the reaction does not occur uniformly across a crack front, but takes place at higher energy kinks, which then spread laterally across the front. In a subsequent paper, Michalske and Bunker [38] demonstrated that to be effective an environmental molecule must also be of a sufficiently small size (< 0.50 nm) to reach and react with the strained bonds at the crack tip.

Modifier ions in glasses, such as sodium or calcium, do not change the fundamental mechanism of crack growth. Crack growth curves obtained for both soda-lime-silica glass [18] and binary alkali silicate glasses [39] have the same fundamental characteristics as those in vitreous silica. Although there are interesting and unexplained changes in the shape of some portions of the $V-K_I$ curve, especially for the binary glasses (Figure 12), the same environments as for vitreous silica caused crack growth.

Other glass systems such as those based on phosphates [40] also exhibit environmentally enhanced crack growth, as do metallic glass systems such as the heavy metal fluorides [41]. Whether the above model applies to these glasses has not been determined.

3.5. Crack Growth in Crystals

All of the experimental data to this point were taken on glasses of one kind or another. We ask the question of whether the existence of long range crystalline order will alter the crack growth mechanism expressed in the Michalske–Freiman model. A number of single crystals, e.g., ZnSe [42], alkaline earth halides [43], GaAs [44], and InP [45], have been shown to be subject to environmentally enhanced crack growth. It is likely that there are many others which also demonstrate this phenomenon.

Michalske et al. [46] demonstrated that the type of bonding in a crystal plays a major factor in determining the crack growth mechanism.

The crack growth data in Al$_2$O$_3$ (sapphire) on the $\{1012\}$ planes are presented in Figure 13. Al$_2$O$_3$ has mixed ionic/covalent bonding quite similar to that in SiO$_2$. The same environments that produced crack growth in silica also do so in Al$_2$O$_3$. One difference is that while in vitreous silica the crack growth curves for each environment had relatively the same slope (Figure 11), the slopes in Al$_2$O$_3$ are different. The authors argued that this difference is due to the nature of the adsorption mechanism on Si–O compared to Al–O bonds.

As seen in Figure 14, the behavior of MgF$_2$ is different from that in either SiO$_2$ or Al$_2$O$_3$. The major difference is
that acetonitrile promotes crack growth in MgF$_2$, while ammonia does not. Michalske et al. [46] hypothesized that this difference was due to the completely ionic bonding in this crystal, leading to electrostatic interactions which reduced the strength of the bonding at the crack tip. The dielectric constant of the liquid becomes the dominant parameter in determining the extent of environmentally enhanced crack growth.
Finally, silicon is the only currently known material in which no environmentally enhanced crack growth has been documented [47]. While crack growth in films of silicon is observed [48], this phenomenon is not equivalent to the growth of cracks in a macroscopic crystal.

4. MATHEMATICAL MODELING

Despite the power of fracture mechanics in allowing an investigator to study crack growth behavior in some detail, there is still no method by which we can directly observe the reaction at the crack tip. We can only hypothesize as to what is happening in this zone. It may be possible with the advent of techniques to create single atomic strands by nanotechnology processing to develop measurement methods to directly measure such reactions. Until then, however, we will have to rely on models of crack-tip behavior.

In principle, it should be possible to use a first-principles, i.e. quantum mechanical, approach to assess the vulnerability of a given material to attack. However, one of the difficulties is the large number of atoms that can be involved in the process. Nevertheless, a number of these type calculations have been carried out.

In one such study [49], ab initio, molecular orbital calculations were used to identify the effect of stress and environment on the electron distribution in a pyrosilicic acid (H$_6$Si$_2$O$_7$) molecule using a Hartree–Fock calculation. It was assumed that this molecule represented the primary structure active at a crack tip. In the first portion of the study, the effect of both bond stretching and bond angle distortion on the charge redistribution in the molecule was analyzed in the absence of external environment. It was observed that the density of bonding electrons associated with the bridging Si and O atoms decreased with increasing strain up to 20%. It was also found that the density of partitioned electrons associated with oxygen increased, indicating a buildup of charge and a polarization of the Si–O group as a consequence of applied strain. Clearly, more electrons will be available to interact with an external molecule.

In a second part of the study [50] similar strains were placed on the H$_6$Si$_2$O$_7$ molecule in the presence of molecules of water, ammonia, formamide, argon, and nitrogen. On approaching the silica molecule, interaction energies were lower for those environments that enhance crack growth, and the interaction energy decreased with increasing strain. These results suggest that the effectiveness of a substance in enhancing crack growth in silica is controlled, in part, by the magnitude of reaction between that substance and the Si–O bond. The increase in electrostatic charges due to strain, as mentioned above, could assist in attracting an environmental molecule to the reaction site. It was also observed that the application of strain reduces steric hindrance.

In a similar molecular orbital study on silicon [51], it was observed that there is no change in electron distribution on either of the crack-tip Si atoms during bond strain. This lack of response of the charges suggests why silicon does not appear to be susceptible to environmentally enhanced crack growth. In addition, this model suggested that the crack-tip shape in silicon presented severe restrictions to any environmental molecule.

More recently, Del Bene et al. [52] carried out a study using second-order perturbation theory that included electron correlation effects. Their basis set was also larger than that of Lindsay et al. [50]. Their principal conclusion was that water reacted as a dimer (H$_2$O)$_2$ with a Si–O bond rather than as a single molecule. However, this study was conducted on a molecule rather than material at a crack tip, so that steric effects may alter this picture.

West and Hench [53] used a semiempirical quantum mechanics method to study the interaction of a water molecule with three-, four-, five-, and six-member silica rings as part of a glass structure. They concluded that rings reoriented prior to fracture of a Si–O bond.

The reaction of silica with water has also been explored by calculating the stress-dependent potential energy surface for the reaction of water with a silica nanorod [54].

As the power of computing grows larger, so that more atoms can be included in a model, the use of quantum mechanics to understand the detailed mechanism leading to environmentally enhanced crack growth will grow as well.

5. POLYCRYSTALLINE MATERIALS

To understand the behavior of polycrystalline materials, one must take into account both the major phase present as the grains as well as the grain boundary phase(s). There is no evidence of a different mechanism at work in these materials than was described in the previous sections. Grain size and shape as well as the grain boundary chemistry and structure will determine the extent of environmentally enhanced crack growth as well as the path of the crack, i.e., transgranular or intergranular. In some cases, e.g., polycrystalline MgF$_2$, there is a transition from intergranular to transgranular crack growth when the crack velocity exceeds the boundary for region I [55].

There are data for a number of other polycrystalline ceramics, indicating that they undergo environmentally enhanced crack growth. These include BeO [56], silicon nitride [57], silicon carbide [58], and ferrites [59]. Crack
growth exponents for a number of single and polycrystalline ceramics are given in Table 3.

6. SUMMARY

Following is a summary of how I view our current state of knowledge on environmentally enhanced fracture:

- In partially covalent materials such as oxides, crack growth occurs because of a stress enhanced chemical reaction between an environmental molecule and strained crack tip bonds. All bonds need not be strained equally; the formation of a "kink" at a reaction site will drag the rest of the crack forward.
- The strain on the crack tip bond increases bond polarization, reduces steric hindrance, and lowers the activation barrier for bond failure.
- Environments that cause crack growth are those that can donate both electrons and protons to a reaction, e.g., water, hydrazine, etc., and are small enough to reach the crack tip.
- The presence of modifier ions such as Na, Ca, etc., in an oxide glass does not change the fundamental mechanism.
- The pH of an aqueous solution is important. High pH, i.e., large concentrations of OH⁻, is much more effective than acidic conditions in causing crack growth in oxide glasses.
- In contrast, mostly ionic crystals, e.g., MgF₂, GaAs, undergo crack growth due to a dielectric shielding of the bonds on the fracture planes.
- Purely covalent crystals, e.g., Si, are not susceptible to either of these mechanisms, and no direct observation of such growth has been reported.

7. WHAT IS YET UNKNOWN

- The exact nature of the crack tip structure is unknown, i.e., Is there a zone of non-elastic strain formed at the crack tip?
- We have no a-priori way of predicting the susceptibility of a given material to environmentally enhanced fracture — We are unable to predict either the position or the slope of V-K₁ curves.
- The nature of the reaction between an environmental molecule and a strained crack-tip bond has been hypothesized, but there is no current method for directly observing the strain-enhanced reaction.
- The exponential form of the crack growth equation has a firm basis in chemical rate theory, but the exact stress dependence is not completely clear.
- Crack arrest has been observed in a few glasses, but whether such a phenomenon occurs in many other materials is unknown.

ACKNOWLEDGMENTS

The 33 years of collaborations and profitable discussions with Shelley Wiederhorn are gratefully acknowledged.
REFERENCES


[23] Freiman SW. Unpublished data.


Development of Superplastic Ceramics

Fumihiro Wakai
Secure Materials Center, Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

1. INTRODUCTION

Superplasticity is phenomenologically defined as an ability of a polycrystalline material to exhibit extraordinarily large elongations in tension at elevated temperatures. The fine structure superplasticity is a property commonly found in many metals, alloys, intermetallics, and ceramics at temperatures $>2/3T_m$ (where $T_m$ is the absolute melting point), when the grain size is very small (less than several micrometers for metals and less than 1 μm for ceramics) and stable during deformation [1,2]. Micrograins apparently move past one another like sand particles flowing, as long as the fracture is suppressed by accommodation processes such as diffusion and dislocation motion. The tensile specimens can be elongated uniformly without necking at relatively low stresses (Figure 1). This property has been applied to manufacture net-shaped components of some superplastic alloys.

While ceramics are brittle materials that show almost no plastic deformation at ambient temperatures, a very wide range of superplastic ceramics had been developed since the superplasticity of zirconia ($ZrO_2$) and its composite was found in the mid-1980s [3,4]. The ceramic superplasticity is common to fine-grained ceramics that are produced by controlling the microstructures to sub-micrometer scales, for example, oxides ($ZrO_2$, alumina, and their composites), nonoxides (silicon nitride [5], SiC), bioceramics, and superconductors, as has been reviewed extensively [6–8]. The ceramic superplasticity is of industrial interest, as it forms the basis of a fabrication method that can be used to produce components having complex shapes from materials that are hard to machine. The superplasticity of $ZrO_2$ has been applied to superplastic forging, sinter forging, sheet forming, gas-pressure forming, extrusion, deep drawing, stretch forming, and superplastic joining. The superplastic sinter forging is also used to produce a textured silicon nitride with improved strength and fracture toughness.

The use of superplastic forming may become even more widespread if large deformation can be achieved at higher strain rates, lower flow stresses, and lower temperatures. The better superplastic ceramics have been developed by improving processing methods and by controlling the grain boundary structure and its chemistry by adding dopant atoms [9]. Especially, high-strain-rate superplasticity in $ZrO_2$-based composite is promising for an efficient shape-forming technology [10,11].

2. MECHANICAL PROPERTIES AND MECHANISM OF SUPERPLASTICITY

In superplastic alloys, the flow stress is particularly sensitive to the rate of deformation

$$\sigma = K\dot{\varepsilon}^m$$

(1)
where \( \sigma \) is the stress, \( \dot{\varepsilon} \) is the strain rate, \( m \) is the strain-rate sensitivity index, and \( K \) is a constant. The very large elongation occurs at \( m \geq 0.3 \), because large \( m \) is a necessary condition for the stability of uniform elongation without local necking. Alternatively, the constitutive equation is described as [12]

\[
\dot{\varepsilon} = \frac{ADGb}{kT} \left( \frac{b}{\sigma} \right)^p \left( \frac{\sigma - \sigma_0}{G} \right)^n
\]

(2)

where \( D \) is the diffusion coefficient \(( = D_0 \exp(-Q/kT)) \), \( Q \) is the activation energy, \( G \) is the shear modulus, \( b \) is the Burgers vector, \( k \) is the Boltzmann constant, \( T \) is the absolute temperature, \( n \) is the stress exponent, \( p \) is an exponent of the inverse grain size, \( \sigma_0 \) is the threshold stress [13], and \( A \) is a dimensionless constant. Equation (2) is equivalent to Eqn (1) at \( \sigma_0 = 0 \), and \( n \) is the inverse of \( m \).

The grain structure remains almost equiaxed after the elongation of hundred percent in superplasticity. This behavior is quite different from that observed in diffusional creep (Figure 2(a)) where grains distort and elongate. Ashby and Verral [14] proposed the grain switching model for superplasticity (Figure 2(b)) from the observation of deformation of oil grains or soap froth. The four grains exchange their neighbors at the final state so that the equiaxed shape is restored. Gifkins [15] considered that the boundary “mantle” of the grain behaved differently from the central “core” of the grain. Grains change their shape by deformation of the “mantle” in his core-mantle model (Figure 2(c)). The shift in the marker line shows grain boundary sliding. The grain compatibility during grain boundary sliding is maintained by a concurrent accommodation process, which involves diffusion, dislocation motion, and grain boundary migration. The grain switching event in the two-dimensional model consists of two independent processes, formation and disappearance of a grain boundary, in three dimensions [16].

When grain boundary sliding is accommodated by grain boundary diffusion and bulk diffusion, the exponents of inverse grain size are \( p = 3 \) and \( p = 2 \), respectively. With reference to Eqn (2), at constant temperature, decreasing the value of the grain size \( d \) will increase the strain rate. But if the grain size is very small and the grain boundary diffusion is very fast, the strain rate is controlled by interface reaction, which is the creation and annihilation of vacancy, rather than by the kinetics of long-range diffusion [17].

At elevated temperatures, failure of ceramics commonly occurs intergranularly by the cavities growing to coalesce to form cracks [18]. Thus, the suppression of cavity formation is important to allow superplastic elongation. A fast grain boundary diffusion, a small grain size, and a small ratio of grain boundary energy to surface energy are necessary for suppressing the nucleation of cavities [19]. To ensure the reliability of superplastically formed components, it is necessary to characterize the cavitation damage induced by the superplastic deformation. The cavity size increases exponentially with strain in superplasticity of \( \text{ZrO}_2 \) [20]. The cavity formation and the maximum elongation to failure are influenced by processing-dependent microstructural factors such as the size distribution of matrix grains, second-phase particles, and residual defects [21].

Superplastic deformation is often accompanied by grain growth, the rate of which depends on either strain [22] or stress [23] and is usually well in excess of that found in the absence of deformation (static grain growth). The enhanced grain growth in superplasticity is termed the dynamic grain growth.

3. SUPERPLASTIC CERAMICS

3.1. Zirconia

The fracture toughness of \( \text{ZrO}_2 \)-based ceramics is enhanced by the stress-activated tetragonal to monoclinic transformation [24,25], in other words, by the transformation toughening. The fine grain size, which was required to retain the metastable tetragonal phase, facilitated also the superplasticity of \( \text{Y}_2\text{O}_3 \)-stabilized tetragonal \( \text{ZrO}_2 \) polycrystals (Y-TZP) [3]. The extensive studies on superplasticity of Y-TZP have been summarized in a review [7]. The relation between the strain rate and stress of high purity Y-TZP is schematically plotted as a solid line in Figure 3 on logarithmic scales. The apparent stress exponent \( n \), which is given by the slope of the curve, is approximately \( n = 2 \) at the high stress region, but it increases to \( n > 3 \) at the intermediate stress region. Several mechanisms have been proposed to explain this transition in stress exponent: 1)

![Figure 2](image_url)

**FIGURE 2** Shear deformation in a regular array of grains [18]. (a) Diffusional creep; (b) Soap froth model; (c) Core-mantle model. (The gray circles show the “core”.)
Transition of the accommodation process of grain boundary sliding from diffusion controlled one at high stresses to interface controlled one at the intermediate stresses, 2) Threshold stress. It has been also reported that the curve shows $n = 1$ at stresses lower than the threshold stress. The deformation of Y-TZP is significantly affected by small amounts of impurities (broken line in Figure 3). The strain rate at the low stress region is enhanced by the addition of 0.12 wt% Al$_2$O$_3$, and $n > 2$ is observed in a wide range of stresses. It has been also reported that the curve shows $n = 1$ at stresses lower than the threshold stress.

The deformation of Y-TZP is significantly affected by small amounts of impurities (broken line in Figure 3). The strain rate at the low stress region is enhanced by the addition of 0.12 wt% Al$_2$O$_3$, and $n > 2$ is observed in a wide range of stresses. A similar effect is also observed in a material containing <0.3 wt% SiO$_2$. The small amount of impurities segregate at grain boundaries, and in some cases, impurity atoms dissolve into ZrO$_2$ grains. The addition of impurity atoms and solute atoms has a wide variety of effects on the grain boundary diffusivities and the lattice diffusivities. It is believed that the grain boundary diffusion and/or bulk diffusion are promoted by these impurity atoms. The large ductility of Y-TZP was achieved by the intentional addition of a large amount of SiO$_2$.

The superplasticity of Y-TZP could be improved by doping transition metals, alkaline earths, and rare earths also. Especially, the codoping of TiO$_2$ and GeO$_2$ is most effective in decreasing flow stress and improving the ductility, because the codoping enhances grain boundary diffusion of Zr cations and decreases the grain boundary energy.

### 3.2. Composites

The fracture toughness of ceramic composite increases with the dispersion of fine tetragonal ZrO$_2$ particles. In two-phase composites, the dispersed second-phase particles often impede the grain boundary migration so that the grain growth of the matrix phase is suppressed. Because the fine microstructure of ZrO$_2$-toughened ceramics is stable at elevated temperature, many fine-grained two-phase and multiphase composites exhibit superplasticity, for example, ZrO$_2$–Al$_2$O$_3$ and ZrO$_2$-mullite. The addition of the second-phase affects the superplasticity and creep of composites in two ways: First, it modifies the continuum deformation mechanics. The effect of volume fraction of the second phase on strain rate can be predicted by a rheology model or by a composite theory. Second, the constituent atoms of the second phase affect interface related deformation characteristics, for example, grain boundary diffusion.

When superplasticity occurs at strain rates considerably $>10^{-2}$ s$^{-1}$, it is conveniently called high-strain-rate superplasticity. High-strain-rate superplasticity of the ZrO$_2$–Al$_2$O$_3$–spinel composite was achieved at a temperature 200 °C higher than the deformation temperature for Y-TZP, mainly because the grain growth was suppressed by dispersion of the second- and the third-phase particles. The grain boundary sliding was partly accommodated by dislocation motion in spinel grains, and the threshold stress was observed in the deformation of the composite.

### 3.3. Silicon Nitride

Silicon nitride (Si$_3$N$_4$) is a light, hard, and strong engineering ceramic that has been developed mainly as a structural material for high-temperature applications. Although creep resistance and superplasticity are incompatible functions, superplastic forming of silicon nitride can be applied to make wear-resistant components that are used at intermediate temperatures. The liquid-phase sintered Si$_3$N$_4$ has residual glass phase pockets and a thin glass film with a thickness of approximately 1 nm at grain boundaries. The superplasticity and creep of Si$_3$N$_4$ are affected by properties of the intergranular glass phase: chemical composition, quantity, glass transition temperature, viscosity, and solubility of Si$_3$N$_4$ to the liquid.

A wide variety of superplastic silicon nitrides have been developed during 1982–2012: (1) Si$_3$N$_4$/SiC nanocomposite, (2) β-silicon nitride, (3) β-SiAlON, (4) equiaxed α/β-silicon nitride, (5) equiaxed α/β-SiAlON, and (6) equiaxed β-silicon nitride.

The superplasticity of silicon nitride was found for the first time in the deformation of fine-grained Si$_3$N$_4$-based composite (material (1)), which was developed by sintering...
amorphous Si–C–N powder [5,45]. Self-reinforced β-silicon nitrides are analogs to whisker-reinforced ceramics relying on the formation of elongated rodlike β-grains in sintering of α-powder. When rodlike grains are very small, silicon nitride (material 2) can still exhibit superplastically large elongations despite this peculiar microstructure [46,47]. SiAlON is a solid solution of Si3N4, in its structure, some Si and N atoms are replaced by Al and O atoms. The strain rate of β-SiAlON (material 3) is enhanced significantly by the liquid phase transiently formed during sintering [48,49]. Because equiaxed grains are preferable for grain boundary sliding and thus, for superplasticity, materials consisting of equiaxed α/β-grains have been developed (materials 4 [50], 5 [51,52]). However, the microstructure of these materials is unstable due to the α- to β-phase transition and grain growth during deformation. This problem was solved by using β-grains as the starting powder in sintering, to get stable fine microstructures with equiaxed grains (material 6) [53–55].

When superplastic forming is conducted on silicon nitride with an equiaxed microstructure, the fracture toughness and the strength at elevated temperatures can be improved by a heat treatment to form elongated rodlike grains for the self-reinforcement. On the other hand, in the superplastic deformation of silicon nitride with rodlike grains, the grains align parallely to the tensile direction in tension tests [54], and in a plane vertical to the compression direction in compression tests [47]. The formation of anisotropic texture can improve the fracture toughness and strength of components in a specific direction [47]. The superplastic forging of silicon nitride has been achieved at high-strain rates by using a spark plasma sintering (SPS) furnace [56].

The superplasticity of silicon nitride takes place by grain boundary sliding accommodated by viscous flow of intergranular glass phase and solution-precipitation creep [57]. In the superplasticity of equiaxed silicon nitride [52], the solution-precipitation process is controlled by diffusion at high stress region [58], and by interface reaction at low stress region [59]. A peculiar phenomenon of shear thickening is also observed in the compressive deformation of SiAlON [51].

3.4. Silicon Carbide

Silicon carbide is a covalent material that adopts network structures featuring vertex sharing of [SiC4]. Although the diffusion coefficient in SiC is very slow, the superplasticity can be achieved by doping with a small amount of boron [60]. The doped boron segregates at grain boundaries and takes the place of silicon, forming bonds in a local environment that is similar to that in the B4C structure [61]. It is supposed that the segregated boron enhanced the grain boundary diffusion, because the diffusion of Si is faster by three orders of magnitude in B4C than in SiC. The superplasticity of SiC is also affected by oxygen impurity atoms [62] and the doping of Al atoms [63].

3.5. Functional Materials

Hydroxyapatite (Ca10(PO4)6(OH)2) is biocompatible with bone and can be used for orthopedic and dental implants. When hydroxyapatite is used as a bone replacement, near-net-shape forming is required to fit for each patients. Hydroxyapatite, with a grain size of sub-micrometer, is translucent, and it can be superplastically deformed at a relatively low temperature of 1000 °C due to its fast diffusivity [64–66]. Carbonate apatite (CAP) is resorbed by osteoclasts and is biocompatible similar to hydroxyapatite. Fine-grained CAP can also be superplastically formed [67].

Oxide superconductor is brittle, and it is usually hard to make wires and ribbons by plastic deformation. A fine-grained YBa2Cu3O7–δ could undergo very large deformation at elevated temperatures, and superplastic elongation was demonstrated recently [68].

4. APPLICATION OF SUPERPLASTICITY

In the ceramic industry, complex-shaped components with accurate dimensions have been usually fabricated by sintering the shaped powder compact. Therefore, in comparison to the sintering, two factors are important for the practical application of superplastic forming: 1) reliability and 2) efficiency. First, it is necessary to ensure the reliability of the products. The source of variability in strength is related to flaws, particularly for ceramics due to their inherent brittleness. The cavitation must be suppressed by selecting appropriate forming conditions. From the viewpoint of efficiency, the combination of sintering and forming, for example, sinter forging [69], is most promising, because both densification and net shaping are achieved simultaneously. Low temperature and high-strain-rate superplastic sinter forging of alumina/spinel composites is actually successfully demonstrated. Porous preforms of nanoceramic composites that were partially densified at low temperatures were superplastically deformed by SPS at 1000–1050 °C at 10−2 s−1 [70]. Spark plasma sintering is similar to conventional hot pressing, and the powder compact (green ceramics) is heated by a direct-pulsed DC current in the graphite die. The compressive deformation of SiAlON also exceeded 10−2 s−1 at 1500 °C in the SPS apparatus [56]. This result suggests a favorable relationship between high-strain-rate superplasticity and ultrafast sintering. Superplastic forming of ZrO2/alumina/spinel
composite could be performed at 1,150 °C by using SPS [71].

Superplastic forming has been applied mainly to ZrO2, for example, sheet forming [72], gas-pressure forming [73], extrusion [74], deep drawing [75], stretch forming [76].

The superplastic forming concurrent with the diffusion bonding process, which is used in the production of titanium alloys in the aerospace industry, is also applicable to ceramics. Superplasticity has been used for the bonding or joining of dissimilar ceramic materials, for example, ZrO2/alumina composites [77–79], ZrO2/hydroxyapatite composites [80], and functionally gradient materials [81]. The superplastic joining of ceramics and ceramic composites has been summarized in a review [82].

REFERENCES


1. INTRODUCTION

In many applications of ceramics in electronics and structural engineering fields, ceramics are required to be joined to dissimilar materials, especially metals. For the electronics applications, for instance, alumina (Al₂O₃) ceramics have been widely used as insulating materials for many decades, represented by printed circuit boards, on which metallic wirings and electrodes are patterned with tight interfaces. Nowadays, functional substrates, such as aluminum nitride (AlN) and low-temperature co-fired ceramics (LTCC), have been also used widely. On the other hand, a variety of ceramic sensors for gas, pressure, magnetic force, temperature, fire, and others have been applied connected with metallic electrodes. As structural materials, oxide ceramics such as partially stabilized zirconia (psz, for instance, ZrO₂−Y₂O₃) and nonoxide ceramics such as Si₃N₄ have been successfully applied to cutting tools and to automotive engine components replacing the conventional metallic materials since the beginning of the 1980s. Figure 1 shows typical ceramic parts established in the commercial fields. Although many methods have been already established for joining ceramics and metals, one has to understand the fact that most of ceramic/metal joint structures are unstable because of the large gaps both in chemical and in physical nature between two materials. There is always some sort of defect structures in an interfacial region. They are, for example, a brittle reaction layer, voids, unjoined area, microcracks from thermal stress, and thermal residual stress. Therefore, in a practical application of ceramic/metal system using joining technology, one has to take the influence of all defect structures in a joint into account and should design an optimized joint structure which can compensate for the negative influence from these defects.

Ceramic/metal joints usually fracture in a brittle manner similar to ceramics and, then, the scatter in strength can be treated by the Weibull statistics. The statistic treatment of strength data can provide us useful information when one tries to grasp suitable joining parameters. There are many defect categories that may cause scatter in strength. From a microscopic view, interface contact formed by wetting, and chemical and physical reaction at interfaces should be of concern in the first place. The nature of atomic bonding, which is on a nanometer scale, may not influence scatter in strength directly but does on the absolute strength because the defect size determining the scatter in strength seems to be beyond a few μm in practical cases. From the macroscopic view, when a reaction layer grows thick, cracking in the layer frequently reduces joint strength. Thermal or residual stress in a joint becomes the other important factor. Large thermal stress both in joining process and in services induces flaws into joints. The final goal for joining research will be in establishing a technique that produces a reliable strong interface by eliminating these defects and by accommodating thermal stress.

The present chapter focuses on reviewing the factors affecting the structural integrity of a ceramic/metal joint. Especially, processing parameters influencing structural reliability are discussed. Processing parameters such as interface reaction are considered initially. The next topic is
the physical contact and the bond face damage effect. Then the effect of thermal stress on strength is discussed. In the last section, practical joining methods, from the conventional to relatively new, are introduced.

2. INTERFACE CHEMISTRIES

Interface formation of a ceramic/metal system has been discussed for many decades on the basis of reaction chemistry [1]. Interface characterization has been carried out with various microstructural observation methods and, especially, with the wetting experiment with the assistance of thermodynamic considerations. Successful results have been achieved from such semi-empirical works and the active metal brazing method is one of the fruitful establishments. In recent years, the research has directed toward the nanometer scale phenomena, i.e., lattice structure observed by high-resolution transmission microscopy [2], atomic binding at dissimilar materials interface evaluation with ESCA/Auger analysis assisted with the first principle/molecular dynamics simulation [3], and even wetting phenomena at ceramic/metal interfaces [4]. These works provide the scientific understanding of ceramic/metal interfaces and the comprehensive explanations for controlling process parameters.

The interface chemistry has a great influence on the reliability of a ceramic/metal joint by changing the uniformity of microstructure resulting in increasing scatter in strength. If homogeneous bonding is achieved over the entire bond face, the scatter in strength is expected to be small. Then, from the viewpoint of the scatter in strength, one should think about the homogeneity of interface bonding and this is influenced primarily by reaction and wetting behavior.

In the metal/metal brazing systems, it is relatively easy to get homogeneous wetting between metal substrates and braze metals if materials and atmosphere are selected suitably. In most of the ceramic/metal brazing systems, it is possible to obtain good wetting by using the active metal brazes [5]. One of the typical active metal brazes is Ag–Cu eutectic braze metal with a small amount of Ti. The addition of Ti is between 1 and 4 wt%. Even though an increased amount of Ti makes good wetting on ceramics, the brazing metals become very brittle resulting in degradation of joint strength. The brazing atmosphere is also important to prevent severe oxidation of the active metals. The vacuum under $10^{-2}$ Pa order is required to achieve good wetting. Figure 2 shows a typical interface microstructure of $\text{Si}_3\text{N}_4$/Ag–Cu–Ti. At the interface, thin TiN/M(Si, Cu, Ti)$_6$N reaction layers are formed [6]. If ceramics are oxides or carbides, they are TiO/M(Si, Cu, Ti)$_6$O or TiC/M(Si, Cu, Ti)$_6$C, respectively.

The Al-based alloys are also regarded as one of the active metals since Al itself is quite reactive on most of ceramics. However, since Al can form a stable oxide skin, it sometimes prevents brazes from achieving good wetting on substrates. Figure 3 shows Weibull plots of the $\text{Si}_3\text{N}_4$ joints brazed with pure Al and Al-X (X = Si, Mg) binary dilute alloys [7]. The scatter in strength of the $\text{Si}_3\text{N}_4$/Al/$\text{Si}_3\text{N}_4$
joint is small and the addition of a small amount of Si does not influence it so much. However, a similar amount of Mg addition to the Al braze makes the scatter very large. These changes in strength are closely related to the change in fracture mode. The addition of Mg increases interface fracture. From TEM observation, it was found that the Mg addition changes the interface microstructure. Figure 4 shows a schematic illustration of the Si$_3$N$_4$/Al alloy interface and the fracture path. When Mg is not in the braze alloy, a nano-crystalline β'-sialon layer and an Al–Si–O
amorphous layer are uniformly formed at the interface. In the case of the Si$_3$N$_4$/Al–Mg interface, on the other hand, Mg concentrates at the interface and oxide products, i.e., $\alpha$-Al$_2$O$_3$ and $\gamma$-Al$_2$O$_3$ with Si, are discretely formed along the interface. The fracture path primarily runs along the oxide layer. Thus, the addition of a small amount of Mg to the Al braze promotes the formation of crystalline Al$_2$O$_3$. This microstructural change increases the scatter in strength of the joint drastically. The reason for the increased scatter in the presence of the stable crystalline oxide products can be ascribed primarily to the weak bonding between the Al oxides and Al/Si$_3$N$_4$.

The presence of sintering additives in ceramics may also influence the interface chemistry and then the reliability. For instance, Si$_3$N$_4$ with or without the sintering additives Al$_2$O$_3$ and Y$_2$O$_3$ was brazed with pure Al and the strength of the joint was found to depend on the presence of the additives [8]. TEM observation revealed that the Si$_3$N$_4$ with the additives formed a thick Al–Si–O layer reaching 1000 nm, while that without any additive formed a 400-nm-thick layer. Figure 5 shows the Weibull plots of strength. The joint of Si$_3$N$_4$ with the additive has higher strength and smaller scatter in strength than the joint without any additive. XPS analysis showed that fracture of the additive-free Si$_3$N$_4$ joint occurred at the Al/reaction layer interface, while the joint with the additive fractured in the Al layer. These results imply that the thick Al–Si–O layer can cover the interface between the substrate Si$_3$N$_4$ and the Al braze layer much more uniformly than the thin formation of the layer. This microstructural difference seems to result in changing the scatter in strength between the two cases. A similar result has been obtained by a pre-oxidation treatment of Si$_3$N$_4$ before brazing [9]. The oxidized treatment of Si$_3$N$_4$ promotes wetting between Si$_3$N$_4$ and Al resulting in decreasing the unjoined area. The oxidized interface has a thick Al–Si–O layer as one of the two reaction layers [10].

It has been known that the ion mixing of a ceramic/metal interface can provide good interface adhesion [11,12]. Peteves reported the influence of ion beam mixing on scatter in strength of the Cr-coated Si$_3$N$_4$/Ni–Cr alloy joint [12]. By Xe ion irradiation on the Cr-coated surface, a two-fold improvement in Weibull modulus was achieved with the improvement of the absolute strength. The improvement of the adhesion of Cr film to Si$_3$N$_4$ can be attributed to the chemistry change due to the enhanced diffusion, and the annealing out of some of the residual surface damage of Si$_3$N$_4$. 

**FIGURE 4** Schematic of Si$_3$N$_4$ joint interfaces brazed with Al dilute alloys [7].

**FIGURE 5** Weibull plots of bending strength of Si$_3$N$_4$ joints with or without additive brazed with Al [8].
3. PHYSICAL CONTACT AT INTERFACE

In the practical joining sequences, a perfect interface adhesion over the entire interface is hardly achieved within a limited joining period and temperature. Then the initial surface roughness and the applied pressure have two of the critical parameters which have great influences on achieving interfacial contact not only in solid-state bonding but also in brazing. In solid-state bonding, interfacial contact is promoted by plastic deformation in the early stage followed by creep deformation and diffusion in the later stage. Pressure across the joint interfaces primarily influences achieving contact by plastic deformation in the first stage. Unjoined islands are inevitably formed on the interface under a limited pressure. It is dependent on the amplitude of pressure, period, temperature, and various material factors such as flow/creep stress. Figure 6 shows the relationship between the fracture stress and the unjoined area of the solid-state bonded Al2O3/Nb joint [13]. Apparently, the increase in unjoined area decreases the joining strength. The importance of the contact pressure during joining on the scatter in strength has been reported by Gottslig et al. [14]. They showed that the high-pressure joining promotes the substantial reduction of the scatter in strength.

If interface reaction releases gas as the reaction product, the pores filled with the gas may be left on the interface resulting in the inhibition of contact. The Si3N4/Ni interface is one of the cases. This interface is weak due to the presence of pores along the interface [15]. When Ni contains nitride-forming elements such as Cr, no pore is formed at an interface and the strength is improved [12,16].

Unjoined area is frequently formed at the edge of a joint. This edge defect weakens the joint extremely as it works as a notch induced on the interface and one must control the formation of edge unjoined band. The inhomogeneity in deformation of the metal layer will also reflect the strength. In the case of the reaction gas-releasing system, the reaction in the outer region may be promoted by continuous evacuation [16]. This will cause excess thinning of the ceramic at the edge region.

In brazing, pressure has a great effect on strength and its scatter. The changes in the joint layer thickness and its uniformity seem to be responsible for the influence. Figure 7 shows the effect of brazing pressure on the strength of the Si3N4 joint with an Al braze [17]. Only slight applied pressure is enough to produce a sound joint with little scatter in strength. Johnson reported the influence of the braze layer thickness effect on the strength of the Si3N4 joint brazed with the Ag–Cu–Ti braze [18]. The strength of the joint was proportional to the inverse of the square root of joint thickness when the thickness was below 50 μm. This indicates that the joint layer becomes one of the defects of which size is taken as the thickness.

Thus, from the processing viewpoint, an appropriate pressure both for solid-state bonding and for brazing should be determined together with the other processing and material parameters. To remove the edge unjoined region, hydrostatic pressing such as HIPing becomes one of the powerful methods. The rough bond face of a metal also has some effect.
4. SURFACE ROUGHNESS AND DAMAGE OF BOND FACE OF CERAMICS

Surface roughness has three basic effects on the strength of a ceramic/metal joint. Rough bond face will prevent completing contact at an interface under a limited pressure. Rough bond face may have the damaged surface layer of ceramics which has deep scratches and residual stress. On the other hand, an irregular bond face may have an anchoring effect which promotes joining by mechanical interlocking. In the actual joining cases, these effects influence on the mechanical properties of a joint in a competing way.

Figure 8 shows the Weibull plots of the strength of the Si₃N₄ joint brazed with pure Al varying the surface grinding condition [19]. Clearly, the rougher bond face made the joint weaker. This degradation can be attributed to the fact that the roughly ground bond face has a damage layer remaining in the joint even after the joining treatment. The fracture of the joint occurs in the damaged layers along the joining interface. Thus, a roughly ground bond face can weaken the ceramic/metal joint if the damaged layer remains in the joint. The surface finishing methods also have important influence on the roughness effects. To make a certain rough surface by polishing is one of the promising conditions for metal—ceramic interface [20].

5. THERMAL STRESS

The distribution of thermal (or residual) stress is not uniform in a ceramic/metal joint even along the interface. Concentration of thermal stress becomes more severe with proximity to the interface and to the free surface. The most harmful effect on joint strength is caused by tensile stress at an interface or in a ceramic. The maximum tensile stress appears near the edge of an interface and on the free surface. Since this stress acts almost vertical to the interface, the apparent bonding strength measured by tensile or bending tests is substantially reduced. The amplitude of residual stress depends on the shape and dimension of the interface [21]. Figure 9 shows the diameter dependence of the thermal stress of the Si₃N₄/invar alloy joint measured on the surface near the interface. The larger diameter makes the larger residual stress. It is also noteworthy that stress concentration at the corner of the rectangular bond face joint is more serious.
The joint with large thermal expansion mismatch decreases strength. However, it occasionally happens that some specimen is strong but the other is weak even if they are the same kind. This depends on the presence and distribution of internal flaws induced by thermal stress during joining treatment. For example, the strengths of the Si$_3$N$_4$/invar and Si$_3$N$_4$/kovar joints are shown in Figure 10 [21]. The latter joint had larger thermal stress than the former as seen in Figure 9. The Si$_3$N$_4$/invar joint exhibits good strength with small scatter. On the other hand, the Si$_3$N$_4$/kovar joint shows large scatter in strength. The distribution of the strength is divided into two parts, i.e., high strength with small scatter and low strength with large scatter. The latter joint always has a large interfacial flaw formed on cooling from the joining temperature. Thus, large thermal stress occasionally induces serious flaws into a joint, which not only weaken the joint but also may make the scatter in strength large. In other words, it is very important to evaluate the scatter in strength, especially for joints with large expansion mismatch.

Thus, it is important to relax thermal stress effect. A key point is in reducing the tensile part of residual stress in a ceramic and on an interface because ceramics are weak against tensile stress while the tensile stress in a metal could be consumed by elastic and plastic deformation. A wide variety of thermal stress-relaxation methods have been developed by using certain interlayers for many ceramic/metal systems, for example, soft metal interlayers, composite interlayers, laminate interlayers (soft metal/low expansion and hard metal), and fine crack interlayers. Table 1 is the summary of those interlayers [1].

Table 1

<table>
<thead>
<tr>
<th>Interlayers</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft metal</td>
<td>ceramics/(Al, Cu, or Ni)/metals</td>
</tr>
<tr>
<td>Soft metal/hard &amp; low expansion metal laminate layer</td>
<td>Nb/Mo for Al$_2$O$_3$/steel</td>
</tr>
<tr>
<td></td>
<td>Fe/W for Si$_3$N$_4$/steel</td>
</tr>
<tr>
<td></td>
<td>Ni/W/Ni for Si$_3$N$_4$/steel</td>
</tr>
<tr>
<td>Grading layer (FGM)</td>
<td>Al$_2$O$_3$/Al$_2$O$_3$—Fe grading/Fe</td>
</tr>
</tbody>
</table>

The soft metal interlayer such as Al and Cu reduces the residual stress by elastic, plastic, and creep deformation. However, such a single layer cannot remove the residual stress effectively, especially for silicon ceramics/metal systems. In addition, it cannot resist to sudden temperature change or severe heat cycle because plastic/creep deformation cannot follow sudden temperature change. Suganuma et al. first proposed the effectiveness of the laminate interlayer that improves resistance to sudden temperature change [23]. Since the soft metal is restricted with a ceramic and a low expansion/hard metal from both sides, it can easily deform following the shrinkage of the ceramic. In addition, the harmful effect of the large expansion/shrinkage of a base metal is blocked by the hard metal layer. Figure 11 shows the effectiveness of Nb/Mo.

![Figure 11](image-url)
interlayer to thermal cycle resistance of the laminate interlayer applied for the Al$_2$O$_3$/stainless steel joint. A single interlayer cannot maintain initial strength after only one cycle between room temperature and 500 °C. The Nb/Mo laminate layer keeps the initial strength even after 100 cycles. The laminate interlayer method, however, still has a limit in bonding size. A large area bonding is still one of the serious matters because the thermal stress increases as increasing bond area.

A composite interlayer, which has a long history so-called as “the powder-graded seal,” “the composite interlayer,” or “the functionally gradient materials,” is one of the effective methods to compensate thermal stress. However, the main problem is strength and reliability of the interlayer itself. Cermet-type composites are generally very brittle, resulting in poor joining strength.

### 6. JOINING PROCESS

Table 2 summarizes a brief comparison of variety of joining methods for ceramics/metals. Mechanical joining and adhesives with organics and cements have been widely used because of their easiness and inexpensiveness. Mechanical joining has a variety of processes. For example, bolting and clamping are the simplest methods. The shrunk-in inserts were applied for the part of production of the turbo-charger rotor [24]. The rocker arm chip made of Si$_3$N$_4$ was inserted in an Al alloy die-cast arm (Figure 1) [25]. The mechanical joints can have heat resistance up to about 500 °C. For the adoption of these methods, it is important to note that the mechanical joining sometimes accompanies a severe stress concentration and that ceramics are weak especially for tensile stress. Although the conventional organic adhesives can produce low strength below 20 MPa generally, the organic adhesive, which has an adhesive tensile strength of about 80 MPa, is commercially available. Fusion welding can produce stable interfacial structures up to elevated temperatures, i.e., melting temperature. However, the problems of grain growth, formation of pores, and thermal stress, in addition to thermal shock, become serious. Friction welding is also one of the simple processes, but effective [26]. Joints can have a tensile strength beyond 100 MPa. However, because the joining torque is limited to be low in order to prevent damage in a ceramic, only soft metals, such as Al, can be selected as a metallic constituent.

### 6.1. Active Metal Brazing and Mo—Mn Method

Brazing and solid-state joining have excellent potential for structural purposes because they can supply both good strength and heat resistance. Research, however, has been focused on oxide ceramics, especially for
application to electrical and electronic components for many years.

Metallizations, such as the so-called Mo–Mn method only for Al₂O₃, are one of the most common processes in industries and have been widely used in the actual insulator productions. Although the Mo–Mn method consists of somewhat complex steps as shown in Figure 12, it is well established. The bonding strength is high depending on brazing alloys.

Recent structural applications of ceramics require different types of joining techniques not only for oxide ceramics but also for nonoxide ceramics in order to produce the stronger and more reliable ceramic/metal joints, which are sometimes expected to have heat resistance superior than those obtained by the conventional metallization processes. Active metal brazing, which is mentioned in the previous sections, is one of the best choices as a brazing method. There are several process choices for the active metal brazing as shown in Figure 13. In each process, vacuum or reduction joining atmosphere is required to achieve tight interface bonding in order to form a semi-metallic layer such as TiN, TiC, or TiO, which is essential in making an ideal bridge layer between ceramic and metal, at the interface.

6.2. Eutectic Liquid Bonding

The Cu/Al₂O₃ joint structure has been successfully applied as shear-resistant electronic circuits. One of the hottest applications is the direct bonded copper (DBC) substrate, which has been used for power semiconductors. Cu and Al₂O₃ can be bonded in the temperature range just beyond 1063 °C, at which Cu forms a thin eutectic liquid film on its surface in the presence of controlled oxygen atmosphere [27]. The eutectic Cu–O liquid promotes wetting. The Cu₄O–Ag system also provides a good wetting for oxide

---

**FIGURE 12** Mo–Mn metallization process. For color version of this figure, the reader is referred to the online version of this book.
ceramics beyond 930°C resulting in the formation of the intimate interface [28].

Some of transition metals such as Fe, Ni, and Co have eutectic reaction between Si₃N₄ and SiC. Figure 14 shows the interface between Si₃N₄ and Ni joined at 1300 °C [29]. At the interface, Si₃N₄ contacts Fe directly after decomposition of Si₃N₄ into Si and N. Si atoms diffuse into Fe during joining, while N atoms go into atmosphere.

6.3. Room-Temperature Bonding (SAB: Surface-activated bonding)

If fresh surfaces of metals and ceramics mate without any contamination, they can form tight bonding even at room temperature. Room-temperature bonding utilizes ion bombardment to remove surface contamination and oxidation followed by intimate contact in low vacuum. Al, Au, and Cu can be successfully bonded to various kinds of ceramics [30]. In most cases, direct atomic contacts are formed at interfaces and the bonding strength reaches the parent body. Si and glass are joined for MEMS applications by using the SAB method as shown in Figure 15.

6.4. SQ Process

The author has developed a cast bonding process for joining ceramics and metals [31]. This method utilizes a pressure casting of Al and its alloys and is called the SQ process (squeeze casting). The basic concept of the SQ process is schematically shown in Figure 16. In this process, ceramic components are heated up to 500 °C–800 °C in a mold. Liquid Al in the temperature range between 700 °C and 800 °C is poured into the mold and pressure is applied immediately. Al liquid rapidly solidifies under a pressure and tight bonding at the interface is accomplished within a few
seconds. This process has many advantages for the production of ceramic/metal joint structures such as no void formation, no brittle IMC formation, low cost without vacuum atmosphere, short tact time, applicability to irregular interfaces, etc. The successful application is the DBA substrate shown in Figure 17 [32]. The strength of the joint interfaces reaches the strengths of Al and alloys.

6.5. Soldering

For joining glasses and metals, soldering has been widely used with high-Pb solder with the aid of ultrasonic vibration [33]. One of the applications is the hermetic edge seal of a pair of glass substrates for buildings and houses. Glazing with glasses of low-temperature melting points,
i.e., Pb-based glasses, has been also widely used by filling and joining very small gaps with large sealing lengths [34]. In the current movement of lead-free technologies in industries, joining glasses also required to be lead free.

The author has developed a novel process for hermetic seal of paired glass substrates [35]. Ti-doped Sn–Zn eutectic solder was developed for sealing glasses. The sealing process is based on friction-activated capillary action. The sealing machine is equipped with a solder container and a solder outlet with a friction plate as shown in Figure 18. Sealing speed is as high as 10 m/s.

6.6. Field-Assisted Bonding
Field-assisted bonding has a long history for glass or β-Al2O3 to metal bonding [36,37]. Bonding is accomplished
beyond 300 °C in about 1 min by applying a DC voltage in 100 to a few hundred volts between the glass and the metals. Under this condition, cations become quite mobile and can form a tight interface. One of the limiting factors is the flatness of the surfaces to be bonded. They must be polished to a roughness below 100 nm. The field-assisted bonding method has been already applied to MEMS applications.

7. SUMMARY

This chapter has focused on joining ceramics and metals. The influences of several important processing factors on the reliability of joints are summarized. The interface formation methods and the relaxation of stress inside the joints are two of the essential parameters in order to obtain high-quality joints. In practical application, a ceramic/metal joined component is used not only under an external stress but also under the internal stress originating from the elastic and expansion mismatch. Since most of engineering ceramics such as Si₃N₄ and SiC usually suffer from severe thermal stress during operation because of their extremely small thermal expansion coefficients against metal components, thermal stress is one of the most dominant factors that restricts the lifetime of joints. There are other important factors in determining the lifetime, i.e., environmentally induced degradation such as oxidation, corrosion, stress-assisted corrosion, and also fatigue. Further intensive works are required for individual applications of ceramics by understanding of the mechanisms by which joint strength decreases in certain environments and also for establishing

FIGURE 18 Friction-activated capillary soldering [35]. For color version of this figure, the reader is referred to the online version of this book.
the design technology for a ceramic/metal joining system. Such efforts are expected to accompany both the development of new techniques and the refinements of existing techniques.

REFERENCES

1. INTRODUCTION

Because Japan relies on imports from abroad for much of its coal, petroleum oil, natural gas, and uranium, the country’s self-sufficiency in terms of energy supply, including hydroelectric power, is only 4%, which is significantly lower than other countries. On the other hand, considering that energy consumption is increasing worldwide, Japan, having such low-energy self-sufficiency, will be greatly affected by world energy consumption, which may threaten Japan’s independency as a nation.

Figure 1 shows a long-term outlook of energy demand up to the year 2030, provided by the International Energy Agency (IEA). According to Figure 1, the world energy consumption in 2009 temporarily decreased for the first time since 1981, and the energy consumption is on the increase again as the economy recovers. World power consumption will steadily increase at the rate of 2.5% per year up to the year 2030, and non-OECD countries, including China, will account for more than 80% of this increase. Fossil fuel will still serve as the main energy source up to the year 2030. In particular, coal dependency will grow so much that coal-fired power generation will become the largest electrical power source, accounting for 44% of the total. On the other hand, although nuclear power generation will become popular in regions other than Europe, the proportion of nuclear energy will decrease, accounting for only about 20% of the total. Renewable energy other than hydroelectric power (wind power, photovoltaics, geo-thermal generation, wave power, and biomass energy) is expected to increase, with wind power and photovoltaics being the major sources, although it will account for as low as 8.6% of the total power generation [1].

As described above, thermal power-generation systems will continue to be the major power-generating energy facilities. Furthermore, nuclear power-generation systems, serving as base load stations, will be enhanced, and the development of environment-friendly renewable energy and new energy, such as fuel cells, will accelerate up to the year 2030.

Figure 2 shows the relationship between power-generation efficiency and operating temperature for various types of power-generating facilities. As is apparent from the figure, high operating temperatures directly lead to enhancements in power-generation efficiency for any type of facility, and hence, power-generation systems that operate at high-temperature are being studied and developed. Thermal power-generation systems, compared with fuel cells, which are capable of converting chemical energy directly into electrical energy, suffer from much energy loss and therefore require operating temperatures of 1000 °C or more to achieve power-generation efficiency equivalent to that of fuel cells.

In recent years, most thermal power-generation systems using liquefied natural gas fuel are likely to employ a combined cycle (LNG C/C) to achieve efficient energy use, environmental friendliness, and improved economical
efficiency. In more detail, the operating temperature of gas turbines in the 1990s and later has been notably high in order to achieve high-efficiency power-generating plants by combining these gas turbines and steam turbines. Such high operating temperatures have been made possible with the development of heat-resistant superalloys forming turbine hot parts, as well as advances made in heat-resistant coating technology and cooling technology. More specifically, for rotor blades with an inlet gas temperature of about 1100 °C in a gas turbine for power generation, the inner surface of the substrate, formed of polycrystalline Ni-based superalloy, is air-cooled, whereas the outer surface of the substrate is covered with a corrosion-resistant and oxidation-resistant coating. Furthermore, gas turbine rotor blades of first stage with a turbine inlet temperature of about 1300 °C, which are more common at present, are formed of a unidirectionally solidified Ni-based superalloy, and the inner surface of the substrate is air-cooled, whereas the surface of the substrate and the inner surfaces of the cooling holes are covered with a corrosion-resistant and oxidation-resistant coating. In addition, gas turbine rotor blades of first stage with a turbine inlet temperature of about 1500 °C, whose commercial operation has already begun, are formed of single-crystal Ni-based superalloys.

Heat-resistant superalloys and heat-resistant coating technology greatly contributing to high-temperature gas turbines, as described above, were first applied to aircraft engines and then to large-scale gas turbines for power generation. Particularly for metal-based corrosion-resistant and oxidation-resistant coatings, processes such as diffusion coating, vapor deposition, and thermal spraying have been realized in consideration of economic efficiency and the characteristics required for hot parts. The development of these corrosion-resistant and oxidation-resistant coating technologies has greatly influenced the development of various types of facilities exposed to a high-temperature atmosphere, such as boilers and industrial furnaces. On the other hand, ceramic thermal barrier coatings, which have recently attracted attention, are metal-based corrosion-resistant and oxidation-resistant coatings covered with ceramic, having superior heat shielding characteristics, which are realized by processes such as thermal spraying or electron beam-physical vapor deposition (EB-PVD) [2–4]. The development of heat-resistant coating technology will be continuously advanced with the aim of further enhancing heat resistance and durability of gas turbine hot parts.

This paper reviews the trend of development of heat-resistant coating technology for gas turbines. The paper also reviews the trend of development and standardization of heat resistance evaluation test methods for coatings, because such evaluation test methods are indispensable for the development of heat-resistant coating technology.

2. DEVELOPMENT OF HEAT-RESISTANT SUPERALLOYS FOR GAS TURBINES

Figure 3 shows one example of the cross-sectional structure of a gas turbine for power generation. The air taken in from the left side in the figure is compressed to a pressure ratio of 15 to 25 by the compressor and is used as combustion air or cooling air. The combustion air is introduced into the combustor in the middle, where it is mixed with fuel and combusted, thus rapidly increasing in volume. The flow direction of the resultant high-pressure combustion air supplied to the turbine section is controlled by the stator vanes, and the high-pressure combustion air is blown onto
the rotor blades secured to the disk to transmit rotational energy to the rotor. This turbine section is composed of normally 3–5 stages of blade rows and vane rows. Complicated cooling flow passages are machined in these rotor blades and stator vanes and are cooled with compressed air, for instance. However, the air-cooling method has already reached its limits. For gas turbines with an inlet temperature of about 1500 °C, steam, with a thermal conductivity 1.5 times higher than that of air, is also employed as a cooling medium to cool the rotor blades and stator vanes.

The key to high-efficiency gas turbines is high-temperature. For this purpose, it is necessary to increase the heat resistance of turbine hot parts. To enhance the heat resistance of hot parts, it is essential to develop heat-resistant superalloys for the rotor blades and stator vanes of turbine. Figure 4 shows year-by-year changes of inlet gas temperature of aircraft engines and gas turbines for power generation [4]. This figure demonstrates that the increases in the inlet gas temperature in the gas turbines for power generation are about 10–20 years behind those of aircraft engines. The figure also shows that the inlet gas temperature in gas turbines for power generation is increasing at a rate of about 20 °C/year. In fact, Ni-based and Co-based superalloys, serving as heat-resistant superalloys, have been developed for the combustor and the turbine sections exposed to the hottest combustion atmosphere. In particular, attempts have been made to enhance the high-temperature strength of Ni-based superalloys by crystal control technology, such as unidirectional solidification and single-crystal structures, to put Ni-based superalloys into practical use as rotor blade superalloys. Compounding processes, such as eutectic alloying, oxide-dispersion strengthening, and fiber reinforcing, are also being developed to increase the high-temperature strength of these superalloys.

Turbine rotor blades, which rotate at high speed, are exposed to a high-temperature combustion atmosphere and undergo compound damage, such as creep due to centrifugal force, low-cycle fatigue due to thermal stress resulting from repeated start/stop operation, and high-temperature oxidation and high-temperature corrosion due to fuel elements. For gas turbines whose operating temperature has been increased to the limit, it is very difficult to realize both high-temperature strength characteristics and environment resistance with the use of existing heat-resistant superalloys. For this reason, gas turbine hot parts are being developed based on the concept that the substrate is responsible for the high-temperature strength characteristics, whereas the coating is responsible for environment resistance and functions. As a result, a wide variety of high-temperature corrosion-resistant and oxidation-resistant coatings have been developed and put into practical use depending on the combustion atmosphere in gas turbines [5,6].
3. DEVELOPMENT OF HEAT-RESISTANT COATINGS FOR GAS TURBINES

Figure 5 shows the progress in development of corrosion-resistant and oxidation-resistant coatings (environmental barrier coatings) and thermal barrier coatings for gas turbine hot parts. The severity of the operating conditions of gas turbines differs greatly among the combustor, rotor blades, and stator vanes, and the instances where thermal barrier coatings have been applied greatly differ accordingly.

The inner surface of gas turbine combustors has been covered with a ceramic thermal barrier coating from
relatively early times. These coatings are based on atmospheric plasma spraying (APS), and the present two-layer thermal barrier coating has been practically realized through the changes to various layered structures. More specifically, in early phases of development, a three-layer coating was developed, i.e., a three-layer coating including a NiAl bond coat (also referred to as undercoat), a CaO or MgO partially stabilized ZrO2 top coat, and a cermet layer, serving as an intermediate layer, formed by mixing a bond coat and top coat for thermal stress reduction. After that, CoCrAlY bond coat with superior corrosion resistance and oxidation resistance at higher temperature was developed and utilized as a bond coat. To further enhance the thermal stress reduction effect, a graded coating in which the ratio of the CoCrAlY bond coat and the MgO partially stabilized ZrO2 top coat changes gradually between the two coats was developed. However, superior oxidation resistance and thermal cycle performance were demanded, and in response, a NiCrAlY or NiCoCrAlY bond coat and a Y2O3 partially stabilized ZrO2 top coat were developed. At present, the inner surface of combustors is covered with a two-layer thermal barrier coating by the atmospheric plasma-spraying process.

On the other hand, applying a ceramic thermal barrier coating to gas turbine blades and vanes that require high reliability is carefully performed because hot parts may be fatally damaged if TBC is failed. Initially, rotor blades were subjected to diffusion treatment with Al or Cr and compound treatment involving Pt plating and Al diffusion treatment for the purpose of corrosion-resistant and oxidation-resistant coatings at high-temperature. After that, coatings of, for example, CoCrAlY and NiCoCrAlY formed by electron beam-physical vapor deposition (EB-PVD) were employed. Next, NiCoCrAlY coatings formed by low-pressure plasma spraying (LPPS), which is superior for mass production, or compound treatment of Al diffusion treatment applied to a CoCrAlY coating surface were realized. At present, the mainstream is a two-layer thermal barrier coating composed of a NiCoCrAlY bond coat formed by low-pressure plasma spraying and a Y2O3 partially stabilized ZrO2 top coat formed by atmospheric plasma spraying. Figure 6 shows examples of 1300 °C-class gas turbine rotor blades (unidirectionally solidified Ni-based superalloy), 1500 °C-class gas turbine rotor blades (single-crystal Ni-based superalloy substrate), and a two-layer thermal barrier coating applied to single-crystal Ni-based superalloy blades. High-velocity oxygen fuel (HVOF) is also used for the bond coat of the two-layer thermal barrier coating. HVOF is a process involving spraying particles onto the substrate surface at a high velocity of 600 m/s or more with the high-pressure combustion energy of fuel and air (or oxygen) serving as a heat source. Although performed in an atmosphere, HVOF is less affected by oxidation during thermal spraying and is capable of producing a dense coating with superior bonding properties. Thus, it is now employed as an alternative to low-pressure plasma spraying.

In addition, hot parts of some aircraft engines and gas turbines are covered with a two-layer thermal barrier coating formed by electron beam-physical vapor deposition (EB-PVD), i.e., a two-layer thermal barrier coating composed of an MCrAlY bond coat and a Y2O3 partially

FIGURE 6 Progress of gas turbine blade for high-temperature use. For color version of this figure, the reader is referred to the online version of this book.
stabilized ZrO₂ top coat with superior erosion resistance, thermal shock resistance, and thermal cycle resistance.

The progress of development of heat-resistant coating technology for gas turbines is described in detail in published reports [5—17]. Furthermore, recent trends are summarized in the Journal of the Ceramic Society of Japan, Special Issue (2008) “Recent progress of ceramic heat-resistant coatings” [18—26].

4. DEVELOPMENT OF THERMAL BARRIER COATINGS

4.1. Effect of Thermal Barrier Coatings

As shown in Figure 7, thermal barrier coating (TBC) is a technology for applying ceramic having low thermal conductivity and high radiation factor to the front surface of a heat-resistant superalloy substrate whose rear surface is forcibly cooled to decrease the front surface temperature of the substrate. Without a thermal barrier coating, the front surface of the substrate is exposed to high-temperature despite the rear surface being forcibly cooled. On the other hand, with a thermal barrier coating, the temperature decreases in the ceramic layer having lower thermal conductivity than that of the substrate, and the front surface temperature of the heat-resistant alloy can be controlled to a low level. Furthermore, with ceramic having a high radiation factor, the front surface temperature of the substrate can be decreased. This thermal barrier effect can be evaluated quantitatively based on the thermal resistance \( R \), as shown in the expression as follows [15,16]:

\[
R = \frac{1}{\alpha_g} + t_c/\lambda_c + t_b/\lambda_b + t_s/\lambda_s + 1/\alpha_c \quad (1)
\]

where \( \alpha_g \) and \( \alpha_c \) are heat transfer coefficient of the hot-side surface and the cold-side surface, respectively, \( \lambda_s, \lambda_b, \) and \( \lambda_c \) are thermal conductivities of the heat-resistant superalloy, bond coat, and the ceramic top coat, respectively, and \( t_s, t_b, \) and \( t_c \) are the thicknesses of the heat-resistant superalloy, bond coat, and the ceramic top coat, respectively.

Eqn (1) demonstrates that when the heat transfer coefficient on the high-temperature-gas side is low, the ceramic layer is thick, the ceramic thermal conductivity is low, the thermal resistance \( R \) becomes large, and the thermal barrier effect becomes enhanced accordingly. Furthermore, because thermal resistance is the ratio of coating thickness to thermal conductivity, a ceramic layer that is thinner than the heat-resistant superalloy substrate is expected to exhibit a sufficient thermal barrier effect. However, a thermal barrier coating is formed on the heat-resistant superalloy surface and is used under a steep temperature gradient. For this reason, important characteristics of TBC materials include a thermal expansion coefficient close to that of the heat-resistant alloy, superior heat resistance, superior heat cycle resistance for repeated thermal load, thermal shock resistance, and so on.

Figure 8(a) shows thermal conductivities measured by the laser flash method for a Y₂O₃ partially stabilized ZrO₂ (8YSZ) top coat formed by atmospheric plasma spraying, a NiCoCrAlY bond coat formed by low-pressure plasma spraying, and the Ni-based superalloy IN738LC. The thermal conductivity of the top coat formed by atmospheric plasma spraying is 1—2 W/mK, which is an order of magnitude smaller than those of the bond coat and the substrate. In addition, unlike the bond coat and the substrate, the thermal conductivity of the top coat changes only
slightly over the entire temperature range, including the high-temperature region, demonstrating that the higher the temperature, the greater the thermal barrier effect.

Thermal barrier effects calculated using these measurement results are shown in Figure 8(b). The calculation conditions are shown in the figure. The heat transfer coefficients of the hot-side surface and the cold-side surface were set to 2750 W/m²K and 2150 W/m²K, respectively. The temperature distribution was calculated by one-dimensional steady-state thermal conduction analysis. The figure demonstrates that the thicker the top coat, the larger the temperature difference, that is, the thermal barrier effect, between top coat surface A and substrate surface C. For example, with a top coat thickness of 300 µm, a temperature difference of nearly 200 K occurs in the top coat layer, showing the effectiveness of the TBC [2].

4.2. Cross-Sectional View of Thermal Barrier Coatings

Figure 9(a) shows one example of a cross-sectional structure of a thermal barrier coating formed by atmospheric plasma spraying on a Ni-based superalloy. This is a normal TBC applied to the inner surface of a gas turbine combustor and is a two-layer thermal barrier coating composed of a NiCrAlY bond coat and an 8 mass % Y₂O₃-ZrO₂ (8YSZ) top coat.

Oxide formation is observed in the NiCrAlY bond coat. The porosity of a mechanically cut piece measured by the Archimedes method is 3.6%. Irregularities on the surface of bond coat are noticeable, and mechanical bonding with the top coat is superior. Furthermore, the 8YSZ top coat is porous, with a porosity of 15.9%.

Figure 9(b) shows the observed microstructure of the 8YSZ top coat. The images are cross-sectional views of a TBC plated with Ni as observed with a scanning electron microscope. The elemental Ni distribution was observed based on a secondary electron image and an X-ray microanalyzer (EPMA). Microstructural evaluation of the complicated sprayed layer is easily performed as a result of Ni being plated in open pores of the 8YSZ top coat formed by atmospheric plasma spraying. As is apparent from the figure, a large number of crack-like narrow open pores can be observed on the 8YSZ top coat. These crack-like open pores can be classified into horizontal cracks along the stack direction (horizontal direction in the figure) of the sprayed layer and vertical cracks along the direction orthogonal to the horizontal cracks. The horizontal cracks are disbonded regions that are formed when flying molten particles adhere to the formed coating. The vertical cracks are solidification/shrinkage cracks formed when the adhering molten particles are cooled and have smaller length and width than disbonded regions. In this manner, there are two types of microcracks on the 8YSZ top coat formed by atmospheric plasma spraying. The disbondings in the stack direction of the sprayed layer increase the thermal barrier effect, whereas the solidification cracks in the vertical direction are effective for thermal stress reduction [8].

Recently, it has been noticed that segmentation of the top coat of thermal barrier coatings is effective in enhancing the delamination resistance of the coatings. Thus, segmented (vertical-crack-introduced) 8YSZ top coats have been realized for thick TBCs formed by atmospheric plasma spraying.

The segmentation structure of a top coat was also observed on a TBC formed by electron beam-physical vapor deposition; one example of the cross-sectional structure is shown in Figure 10. In general, electron beam-physical vapor deposition is the process of irradiating coating materials in a vacuum container by using a high-output electron beam to ablate the materials, thereby depositing the
materials on a substrate surface heated at high-temperature. A segmented TBC formed of columnar grains with a width of several micrometers, as shown in the figure, is obtained by performing coating under appropriate deposition conditions. There is a gap of 1 μm or less between individual columnar grains, and superior thermal cycle resistance characteristics are exhibited as a result of thermal stress being reduced. Furthermore, it has been demonstrated that columnar grains branch to form feather-like structures, exhibiting superior thermal barrier characteristics [22].

5. DAMAGE MODES OBSERVED IN THERMAL BARRIER COATINGS

Figure 11 summarizes the damage modes of thermal barrier coatings applied to gas turbine hot parts [12]. First, chemical damage includes high-temperature corrosion due to high-temperature oxidation or corrosion factors in combustion gas. Evaluation of the high-temperature oxidation characteristics of thermal barrier coatings is carried out, in many cases, by continuous heating tests at a constant temperature in a static atmosphere. For this evaluation, a combustion flame heating test and a corrosive ash deposition heating test in a burner rig have been carried out. A large amount of data have been accumulated from these tests [10,19]. Pure ZrO₂ has a certain degree of resistance to sulfuric acid corrosion and vanadium attack. However, ZrO₂ degrades when Y₂O₃, MgO, CaO, and so on added to stabilize the ZrO₂ crystal structure react with SO₂, SO₃, Na₂SO₄, V₂O₅, PbSO₄, and so on acting as corrosion components in the combustion gas. More specifically, ZrO₂ stabilized in the form of a tetragonal or cubic structure becomes unstable due to a corrosive reaction, causing cracks or delamination to occur on the top coat due to a change in volume at the time of a phase change into monoclinic crystal.
In addition, the delamination life characteristics of thermal barrier coatings prominently decrease through thermal cycle tests, compared with high-temperature continuous oxidization tests. From this point of view, many studies on delamination life evaluation of thermal barrier coatings using thermal cycle tests have been carried out [10,19]. As a result, it has been clarified that 6–8 mass % Y₂O₃ partially stabilized ZrO₂ top coats exhibit significantly superior thermal cycle resistance compared with other ceramic top coats.

Next, when thermal barrier coatings are exposed to a high-temperature atmosphere, thermally grown oxide (TGO) with Al₂O₃ as a main component is formed at the interface between the 8YSZ top coat and the MCrAlY bond coat, as shown in Figure 12. This is because oxygen passing through the top coat causes the MCrAlY bond coat surface to be oxidized. Thus, the thermally grown oxide increases in volume and produces cracks or delamination on the top coat layer. The thickness \( \omega \) of this TGO layer, which is given by the expression below, becomes larger as the heating temperature becomes higher and the heating time becomes longer [27]:

\[
\omega = \kappa \cdot t^n (n = 0.45)
\]

In short, the thickness \( \omega \) of the TGO layer follows an \( n \)-th power law of the heating time \( t \), where the velocity constant \( \kappa \) represents the Arrhenius temperature dependency. As a simplified method, an attempt has been made to define the limit thickness of the TGO layer at which the 8YSZ top coat starts to be delaminated and to predict the delamination service life of the TBC based on quantification of the growth behavior of the thickness of the TGO layer. The growth of an embrittled layer that is formed on the substrate side due to reaction diffusion at the interface between the bond coat and the substrate is also important as a damage mode [28].

On the other hand, mechanical damage includes cracks and delamination on the 8YSZ top coat due to repeated thermal stress. If compressive stress acts upon the TBC, interface cracks occur due to irregularities at the interface between the 8YSZ top coat and the MCrAlY bond coat. These interface cracks develop through thermal cycling.
and, at their limit size or larger, lead to TBC delamination. In addition, when very high compressive stress occurs, the TBC undergoes buckling, which then leads to spalling (delamination resulting from buckling). In contrast, when tensile stress acts upon the TBC, the TBC exhibits vertical cracks, that is, cracks orthogonal to the coating. When the TBC undergoes high tensile stress, vertical cracks initially generated on the TBC develop, resulting in delamination in the TBC layer or delamination at the interface between the 8YSZ top coat and the MCrAlY bond coat. However, many experiments have confirmed that delamination damage observed in thermal cycle tests of TBCs occurs in the course of cooling, and major delamination is assumed to occur in the compressive stress field. While it is true that delamination is greatly affected by the thermal load acting on the TBC, residual stress occurring in the TBC at the fabrication stage, as well as oxidation/inflation of the MCrAlY bond coat, that is, residual stress due to growth of the TGO layer, also contributes to delamination.

In addition, sites where delamination occurs differ depending on the type of TBC. More specifically, for a TBC formed of an MgO partially stabilized ZrO$_2$ top coat, delamination occurs in the non-stabilized ZrO$_2$ layer. For both a 7 mass % Y$_2$O$_3$–ZrO$_2$ top coat and NiCoCrAlY bond coat formed by atmospheric plasma spraying, delamination occurs in the bond coat. If only the NiCoCrAlY bond coat is formed by low-pressure plasma spraying, delamination occurs in the top coat close to the interface with the bond coat. If the 7 mass % Y$_2$O$_3$–ZrO$_2$ top coat is formed by EB-PVD, and the NiCoCrAlY bond coat is formed by low-pressure plasma spraying, delamination occurs in the TGO layer formed at the interface between the top coat and the bond coat or occurs at the interface between the TGO layer and the bond coat. These delamination sites are assumed to be the weakest points of the respective TBCs.

Recently, sintering/shrinkage cracks of the top coat due to the high-temperature in gas turbines and resultant delamination have been observed [29]. Figure 13 shows a TBC surface “as-sprayed”, as well as TBC surfaces heated for one hour at 1373 K, 1473 K, and 1573 K, respectively. Cracks on the 8YSZ top coat surface are seen in the as-sprayed state, and the 8YSZ top coat can be seen to start sintering at a heating temperature of 1473 K or more. This sintering shrinkage of the 8YSZ top coat causes not only a decrease in the thermal barrier effect due to densification of the coating but also delamination. More specifically, if the top coat surface layer undergoes delamination damage due to sintering shrinkage, the underlying surface is exposed to high-temperature and is subjected to another sintering.
shrinkage that leads to secondary delamination damage, thus decreasing the thickness. The sintering shrinkage rate \((\Delta L/L_0)\) of this 8YSZ top coat is given by

\[
\Delta L/L_0 = k' \cdot t^n (n = 0.4)
\]  

(3)

More specifically, the sintering shrinkage rate of the 8YSZ top coat follows an \(n\)-th power law of the heating time \(t\), where the sintering rate constant \(k'\) represents the Arrhenius temperature dependency. Based on this sintering shrinkage behavior of the 8YSZ top coat, the occurrence of TBC vertical cracks can be evaluated under a temperature gradient. Under a temperature gradient, vertical cracks generated on the 8YSZ top coat are bowed, causing delamination on the surface layer.

Last, although depending on the type of fuel used in the gas turbine, erosion damage of the 8YSZ top coat due to solid particles in the combustion gas is also a critical damage mode.

As described above, thermal barrier coatings include various damage modes, and damage phenomena appear in diverse patterns depending on the TBC operating environment and type.

Figure 14 schematically illustrates systematization of coating engineering for heat-resistant coatings of gas turbines. In summary, coating technology is composed of (1) coating materials and process technology, (2) coating design technology, and (3) coating evaluation technology for service life and characteristics. Development of conventional heat-resistant coating technology has been aimed at enhancing characteristics, with emphasis on materials and process technology. To achieve higher functions, performance and higher reliability, however, development of coating design technology is essential, and enrichment of databases will be indispensable in doing so. Furthermore, standardization of test methods is important for evaluating service life and characteristics, which has so far been carried out in various ways.

6. HEAT-RESISTANT EVALUATION TECHNOLOGY

6.1. Trends in Development of Coating Evaluation Tests

In the history of development for practical realization of gas turbine hot parts, development of test methods for evaluating the service life and characteristics of heat-resistant coatings has long played an important role. Initially, tests originally designed for heat-resistant superalloys were adapted to heat-resistant coatings. At present, however, damage evaluation tests exclusively designed for heat-resistant coatings are employed.

Table 1 summarizes heat resistance evaluation test methods used for gas turbine hot parts, classified into several types, as well as the development process of each type. As is apparent from the figure, heat resistance evaluation tests can be classified into (1) burner heating tests,
(2) furnace + burner heating tests, and (3) furnace heating tests, based on the heating method.

First, the burner heating test has been conducted as a high-flow-rate combustion test (burner rig test) to evaluate heat resistance under the operating conditions used in actual facilities. This is a test method primarily for verifying the heat resistance of hot parts by approximating the combustion conditions and high-temperature corrosion environment in the test to those in actual facilities. The burner heating test includes an atmospheric pressure combustion test and an actual-pressure combustion pressure test [9].

Next, the furnace + burner heating test is a hybrid test in which furnace heating is combined with high-temperature corrosive gas generation by burner combustion. Compared with the burner heating test, this hybrid test can control the test temperature with high precision. Furthermore, because of its ability to simulate a high-temperature corrosion environment with combustion gas, the hybrid test is an effective method for carrying out long-term high-temperature corrosion evaluation [9].

The furnace heating test is the most basic heat resistance evaluation test method and has been used at the material development stage. Because uniform heating is easy in a furnace, many actual examples of high-temperature continuous oxidation tests, high-temperature cyclic oxidation tests, and high-temperature corrosion tests for heat-resistant superalloy and heat-resistant coatings have been reported [10].

All the heating tests described above were initially developed for evaluating the heat resistance of noncooled blades (1980s and before). As shown in Table 1, however, these tests have evolved into heat resistance evaluation test methods taking into consideration the mechanical load exerted on the rotor blades and stator vanes, and further into heat resistance evaluation test methods with a temperature gradient, in consideration of the wide adoption of cooled blades. The evolution of such heat resistance evaluation test methods has enabled the evaluation of gas turbine hot parts under conditions more closely approximating the actual conditions and has accelerated the development of hot parts. In more detail, the gas turbine rotor blades rotate at high speed in a high-temperature combustion gas atmosphere and are subjected to centrifugal force (tensile stress) load. On the other hand, the stator vanes are exposed to high-temperature combustion gas with constrained thermal expansion and are subjected to compressive stress load.

First, in the burner heating test, heat resistance evaluation tests reflecting the above-described load conditions are carried out. In short, this is a test method in which a test specimen is subjected to both mechanical load and a temperature gradient together. In fact, a high-temperature axial-strain control test device with a temperature gradient has been developed. With this test, both the temperature and stress history at the time of startup/stopping of the gas...
turbine were simulated to investigate the effect on TBC delamination damage [15–17].

Also for the hybrid (furnace + burner) heating test, on the other hand, a test composed of a thermal mechanical fatigue test plus the influence of high-temperature corrosion is possible, although no actual examples have been reported at present. Such a test will be carried out if long-term and high-temperature corrosion evaluation becomes necessary in future due to the diversity of fuels.

6.2. Standardization of Evaluation Test Methods

Normally, heat resistance evaluation for gas turbine hot parts can be classified into four steps I–IV, as shown in Figure 15. More specifically, step I corresponds to coating design, step II to the furnace heating test, step III to the burner heating test, and step IV corresponds to a rainbow rotor test (Verification test) with hot parts installed in a gas turbine. Based on this classification, a standardization road map, as shown in Table 2, has been formulated for test methods for evaluating the characteristics of gas turbine heat-resistant coatings [18]. Based on this road map, world-leading standardization activities are being carried out in the New Material Center affiliated with Osaka Science & Technology Center. The progress of this standardization is described below.

First, “Testing Methods for Heat Resistance Under Temperature Gradient (JIS H7851)”, involving burner heating, was formalized as a Japanese Industrial Standard in 2005. More specifically, as shown in Figure 16, this test is a combustion test method in which the surface of a disk-shaped test specimen brazed on a copper holder is heated with a burner, and the rear surface of the copper holder is cooled to produce one-dimensional heat flow, thus carrying out heat resistance evaluation. This is an evaluation test method mainly intended to elucidate the damage mechanism of heat-resistant coatings, and the thermal flow rate, equivalent thermal conductivity, and so on can be calculated using the expressions shown in the figure based on temperature measurements. At present, this testing method is under discussion for international standardization (ISO 13123) by ISO/TC107.

FIGURE 15 Four steps for practical use of gas turbine hot parts. For color version of this figure, the reader is referred to the online version of this book.
TABLE 2 Roadmap for Standardization of Heat-Resistant Coating Tests

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1500°C C/C</td>
<td>□Advanced light water reactor ABVR/APWR</td>
<td>□1300°C IGCC</td>
<td>□700°C ST</td>
<td></td>
</tr>
<tr>
<td>Aircraft engine/High-temperature technology</td>
<td>□Power generation GT/High temperature technology developed 10 yrs. late</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel cell stack/improved durability technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automotive PEFC: Durability of 5000 hrs.; output density of 2.4kW/L; vehicle efficiency of 64% LHV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-temperature SOFC: Durability of 40000 hrs.; output density of 0.5kW/L; vehicle efficiency of 50% LHV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental 3 R(Reuse, Reduce, Recycle)/High-temperature reactor technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longer product life &amp; heat resistance of combustion chamber</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, “Testing Methods for Thermal Cycle and Thermal Shock Resistance of Thermal Barrier Coatings (JIS H8451)”, involving furnace heating, and “Testing Methods for Thermal Cycle and Thermal Shock Resistance of Oxidation-Resistant Metal Coatings (JIS H7852)” were formalized as Japanese Industrial Standards in 2008. JIS H8451 specifies a thermal cycle test method in which a TBC test specimen is repeatedly subjected to thermal cycling by furnace heating and air cooling, as well as a thermal shock test method in which a TBC test specimen is subjected to thermal shock by furnace heating and then water cooling, as shown in Figure 17. Figure 18 shows one example of an 8 mass % Y2O3–ZrO2 thermal barrier coating compared with an Al2O3 coating formed by atmospheric plasma spraying in order to elucidate the thermal shock resistance characteristics of the 8 mass % Y2O3–ZrO2 coating. It should be noted, however, that SUS304 is used for the substrate. The 8 mass % Y2O3–ZrO2 thermal barrier coating exhibits a thermal shock temperature difference, $\Delta T$, of 900°C or more, whereas the Al2O3 coating exhibits a thermal shock temperature difference, $\Delta T$, of 500°C, when $\Delta T$ is measured just before the residual bonding strengths of the thermally shocked coatings (JIS H8402) decrease by 30%. At present, JIS H8451 is also under discussion for international standardization (ISO 14188) by ISO/TC107.

For heat-resistant coatings, on the other hand, it will be indispensable to introduce coating design technology, as shown in Figure 15. For this purpose, it is important to construct a database necessary for coating design. From this standpoint, a project sponsored by the New Energy and Industrial Technology Development Organization (NEDO), “Standardization Investigation Project on Testing Method for Characteristic Evaluation of Heat-Resistant Coating”, was conducted in 2006–2008, and a JIS draft regarding evaluation of mechanical characteristics and thermal characteristics of thermal barrier coatings was prepared.

Figure 19 shows an example of residual stress measurements for thermal barrier coatings. The figure shows calculated results of residual stress distribution in the thickness direction of the thermal barrier coatings, based on the substrate deformation when a strip-shaped test specimen is

![FIGURE 17 Testing methods for thermal cycle of thermal barrier coatings (JIS H8451/ISO 14188:2012).](image-url)
sprayed or the substrate deformation resulting from coating removal. The CoNiCrAlY coating exhibited higher tensile residual stress on the surface than the 8 mass % Y2O3–ZrO2 thermal barrier coating, showing relatively good agreement with the X-ray surface residual stress measurements.

Figure 20 shows thermal diffusivity obtained by measuring the temperature history curve of coatings by the laser flash method and then calculating based on the differences from values measured with the substrate alone. The thermal diffusivity of the CoNiCrAlY coating is in good agreement with the results measured with a single CoNiCrAlY layer if the thickness of the coating is 0.25 mm or more. On the other hand, the thermal diffusivity of the 8 mass % Y2O3–ZrO2 thermal barrier coating is slightly smaller than the measurements of a single 8 mass % Y2O3–ZrO2 layer but does not differ depending on the top coat thickness. At present, a JIS draft related to methods for measuring the elastic constant, linear expansion coefficient,
and thermal conductivity of thermal barrier coatings is being produced.

On the other hand, one important technology related to heat resistance evaluation tests for heat-resistant coatings is a technology for nondestructive detecting delamination of coatings. The ultrasound method or holography method has already been investigated, and in addition, technology for detecting delamination of coatings by infrared image analysis is being used in practice [30]. In short, this technology is an approach for detecting delamination sites by infrared thermography, by drawing upon the phenomenon that only the delamination sites are overheated when the TBC surface is heated, as shown in Figure 21. With this technology, not only delamination of heat-resistant coatings can be detected, but also changes in thermal barrier characteristics of coatings due to erosion or local sintering phenomena can be measured. Future improvements in this technology include enhancement in detection accuracy, early application to complicated structures such as actual facilities, and standardization.

7. CONCLUSIONS

For 1500 °C-class gas turbines, the adoption of single-crystal Ni-based superalloy blades and ceramic thermal barrier coatings is indispensable, and additionally, steam-cooled technology should be employed. In particular, thermal barrier coating (TBC) technology is recognized as important, and research papers on this technology have drastically increased in number. The reliability of this technology is also increasing accordingly. Among others, studies on characterization of thermal barrier coatings and delamination life evaluation have advanced remarkably. This is because researchers at universities and research institutes have actively taken part in development of technology for evaluating the service life and characteristics of thermal barrier coatings, while manufacturers of gas turbines have primarily been involved in process development.

In the near future, standardization of heat-resistant coating technology will become increasingly important to improve the reliability of gas turbine hot parts. For this purpose, acceleration of standardization through industry—academia collaboration is essential. From these efforts, we expect rapid advancement of heat-resistant coating technology in the future.

The names of the products referred to in this paper may or may not be trade names of respective corporations.

REFERENCES

1. INTRODUCTION

Promotion of high-level recycling of waste and reduction in environmental load such as CO₂ and dioxins are required for waste treatment worldwide. The concepts of higher efficiency waste-to-energy (WTE) plants and waste gasification and ash melting power generation plants are positioned in the center of thermal recycling in Japan. The high-efficiency WTE boilers have been progressed recently with steam conditions of more than 400°C/3.9 MPa. On the other hand, the fluidized bed boilers burning biomass such as woods, straw, scrap tires, RDF, RPF etc. have recently been operated in 440–540°C and 6–30 MPa electric power generation plants. Nowadays a high total cost performance is required in the storm surrounding the society such as the globalization of economy, the strict pollution regulations, and carbon neutral power plants.

Therefore, the development of low-cost and highly durable materials and application processes has become important issues that are essential for a high thermal efficiency and economy of WTE plants. In addition, in order to improve the durability of high-temperature, high-pressure boiler materials, it is considered necessary to apply optimum boiler designs that prevent high temperature corrosion (HTC) and erosion—corrosion of superheater tubes (SHTs) and waterwall tubes (WWTs). There are high expectations for the development of the high-temperature corrosion-resistant materials (CRMs) and the corrosion-resistant coatings (CRCs) that are highly durable and workable against aggressive corrosion environments.

Advanced refractory materials (RMs) that realized excellent furnace performance have been developed and applied for combustion furnace to prevent high-temperature damages and slugging of ash. Also, CRMs have been progressively used for SHTs, and the application of CRCs such as ceramic tile, metal spray coatings, and weld overlay of Ni-base alloys has been widespread for the WWTs in order to reduce maintenance cost and keep stable operation for long duration. There is a strong need for low-cost materials to improve the durability in severe erosion—corrosion damages caused by the soot blower attack.

This report describes 1) the formation of severe corrosion environment of WTE boilers, 2) advanced RMs used in combustion furnace as porous ceramics mainly manufactured by using artificial raw materials, 3) application trends...
and field experiences of advanced ceramic and cermet coatings that have good durability in WWTs and SHTs, and 4) corrosion and deterioration mechanisms of alloys, cermet, and ceramic coatings formed by thermal spraying.

2. CORROSION ENVIRONMENTS OF WASTE-TO-ENERGY BOILERS

In the recent trends of most WTE plants, the steam condition of 300—450 °C/2.9—5.8 MPa shown in Figure 1 is adopted to avoid the corrosion damages of boilers. Due to a large change in life styles of people and in production styles by industries brought about by technological and economic development, the properties of municipal and industrial waste and the WTE plant operations have changed as follows during recent forty years [1]:

1. Operational change for strict pollution regulation: low O₂ operation, high-temperature combustion, rise of plant operating rate, etc. to prevent generation of NOx and dioxins.
2. Energy-saving and effective utilization of heat recovery: severe corrosion environment has been formed due to improvement of electric power-generation efficiency such as high-temperature and high-pressure steam condition (400—500 °C/3.9—9.8 MPa) of boilers.
3. Needs to reduce volume of combustion ash and produce nonpolluting ash have led to the development of new incineration processes, such as waste pyrolysis gasification and ash melting plants and oxygen-enriched combustion [2].
4. Improvement of cost performance: Reducing of operational cost by advances in maintenance-free operation, such as combustion control and monitoring technologies.

Various substances including incombustibles as well as combustibles such as woods, paper, and plastics are inhomogeneously mixed in the waste. Therefore, the fluctuation of gas temperature and composition containing much HCl, SOₓ, O₂, H₂O etc. is larger than that in fossil fuel boilers; also, many low-melting-point deposits containing high concentration of chlorides is generated in the combustion gas. Corrosive constituents influenced to HTC in different fuel boilers are listed in Table 1 [3]. Also, Figure 2 shows the explanatory drawing of the corrosion factors formed by combustion gas in WTE boilers [4]. Chlorides and sulfate dusts containing high concentration of alkaline metals (Na, K, etc.), heavy metals (Pb, Zn, etc.), deposit on the material surfaces exposed in the combustion gas, and severe HTC is caused in boiler components such as WWTs and SHTs by the deposits melting at 300—550 °C. In SHTs, the metal temperature is as high as 300—550 °C, then deposits are easily molten by eutectic reaction and corrosion rate sometimes shows as high as several mm/year or more. Figure 3 shows the increase in corrosion rate of SHT materials with increase in the combustion gas temperature at 450 and 550 °C [5]. Furthermore, on the WWT surfaces where the gas temperature is very high (850 °C or higher), high concentration of chlorides with lower melting point is promoted, and the corrosion damage is caused mainly by chlorination reactions. The corrosion form of materials is commonly general corrosion, although intergranular and/or localized corrosion occur by the existence of molten deposits and the breakdown of the protective oxide layer on even highly CRMs [6]. Both the adherence of strongly corrosive dust constituents, such as chlorides and sulfates, and the gas temperature fluctuation can be reduced by reducing the combustion gas temperature flowed into the SHTs to less than ~650 °C. Then the reduction in gas
### TABLE 1 Corrosive Matter and Deposits Condition Influencing to High Temperature Corrosion in Typical Different Fuel Boilers

<table>
<thead>
<tr>
<th>Combustor</th>
<th>Fuels</th>
<th>Corrosive matter</th>
<th>Metals</th>
<th>Low melting point deposits</th>
<th>Corrosion phenomena</th>
<th>Molten salt related corrosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wasterto-energy Boiler</td>
<td>Municipal and Industrial waste, Biomass etc.</td>
<td>⬤ ○ ○ ○ ○ ○ Zn, Pb</td>
<td>○</td>
<td>290 - 516</td>
<td>Small ≤ 20%</td>
<td>○</td>
</tr>
<tr>
<td>Sodium Recovery, Biomass Boiler</td>
<td>Black Liquor</td>
<td>◯ ◯ ◯ ◯ ◯ ◯</td>
<td>○</td>
<td>300 - 500</td>
<td>Medium</td>
<td>◯</td>
</tr>
<tr>
<td>Coal Fired Boiler</td>
<td>Coal, Wood etc.</td>
<td>△ ◯ ◯ ◯ ○ ○ Fe</td>
<td>○</td>
<td>550 - 620</td>
<td>Small</td>
<td>◯</td>
</tr>
<tr>
<td>Coal Gasification, Singas Cooler</td>
<td>Decomposition gas etc.</td>
<td>△ ◯ ◯ △ △ Al, Ca Si, C</td>
<td>○ ◯</td>
<td>480 - 725</td>
<td>Medium</td>
<td>◯ ◯ ○</td>
</tr>
<tr>
<td>Heavy Oil Boiler</td>
<td>Heavy oil, Asphalt etc.</td>
<td>△ ◯ ◯ ◯ ○ ○ V</td>
<td>○</td>
<td>480 - 725</td>
<td>Medium</td>
<td>◯ ◯ ○</td>
</tr>
<tr>
<td>Heavy Oil Gas Turbine (G/T)</td>
<td>Heavy oil</td>
<td>△ ◯ ◯ ◯ ○ ○ V</td>
<td>○</td>
<td>480 - 725</td>
<td>Large</td>
<td>◯ ◯ ○</td>
</tr>
<tr>
<td>Remarks</td>
<td>Degree of influence</td>
<td>注: Large ○: Small △: According to condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>○: Main corrosion reaction</td>
<td>○: Existence of reaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>○: According to condition</td>
<td>△: According to condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
temperature is effective for stabilization of protective oxide layer formed on the metal surface. However, the soot blowing used for the purpose of removing deposits is known to result in severe thermal cycling condition and even the breakdown of protective oxides layer.

3. APPLICATION OF ADVANCED REFRACTORY MATERIALS

Progression of combustion technologies in the WTE boiler, the high durability of furnace structure, and RMs become the key issue to maintain high-performance and continuous operation of furnaces. Therefore, raw materials of RMs change from natural materials to artificial materials such as high-purity silicon carbides (SiCs) and alumina (Al₂O₃). RMs are constructed by relatively coarse and inhomogeneous grains compared with metallic materials and fine ceramics, then those properties porous and somewhat scatter. These are the reasons for difficulties of materials evaluation and selection. Influenced factors for damages of RMs used in combustion furnace of boilers are summarized schematically in Figure 4.

Advanced refractory linings (RLs) and RMs mainly applied to the corrosive environment of stoker combustion furnaces are shown in Figure 2 and described in this chapter.

3.1. Firebrick Linings

Furnace walls (FWs) in relatively small incinerators, without boilers are constructed by brick linings. Required properties and mainly applied RMs along the gas flow in stoker and bubbling fluidized bed furnaces are shown in Table 2 [7]. Also, Figure 5 shows typical RMs and installation methods used for waste incinerator and industrial boiler.

High-SiC brick linings that have high corrosion resistance for molten ash (mp: 1200–1300 °C) are used to prevent large slugging of molten ash (so-called “clinker”) in high-temperature zone and to easily remove clinker without damaging RMs. One example of clinker growth in
SiC bricks are also very hard and have high wear resistance, than that installed near and along the stoker grate to prevent wear by hard components including in waste. SiC bricks have high thermal shock resistance due to their high thermal conductivity, but are easily oxidized by H₂O vapor etc. at a high temperature of more than approximately 900°C, as shown in Figure 7 [9]. SiC particles oxidize to SiO₂ ones that have higher volume, and as a result of oxidation, swelling and fracture damages occur as listed in Figure 8 [10].

As high Al₂O₃ bricks are also hard and highly oxidation and corrosion resistant, they are therefore sometimes installed in the high gas temperature zone. However, thermal shock resistance is relatively lower than that of SiC bricks. Common fireclay bricks made by alumina and silica are installed at a relatively low-gas-temperature zone in furnace because of lower heat and corrosion resistances.

Table 3 shows comparison of physical properties in typical RMs used in WTE boilers [11]. Also, Table 4 shows comparison of main damage resistances in typical RMs required for boiler applications [12].
3.2. Cooled Furnace Wall

3.2.1. Water Cooled Lining

Increase in throughput and calorific value of waste and reduction in furnace wall temperature have been required to prevent slugging problem of clinker and damages of RMs. Also in large-scale WTE plants, efficient heat recovery from combustion gas becomes important, subject to increase in boiler thermal efficiency. Therefore, water-cooled FWs with boiler WWTs become major furnace structures at this time. In water-cooled refractory linings (RLs), reduction in furnace wall temperature, strength of RMs, and highly durable support system for linings are important subjects. The following support systems for unshaped RLs shown in Figure 9 are commonly applied to maintain the long lifetime of FWs [13,14]:

1. Studs with SiC ceramic cap system
2. Anchor system made by heat-resistant alloys or ceramics

In castable and plastic refractories, grain structures cannot be sintered completely, because rear-side of RLs is cooled by WWTs that is low surface temperature of approximately
TABLE 2
Required Performance and Applied Refractories for the Furnace of Stoker and Fluidized Bed Combustor in Waste Incineration Boilers

<table>
<thead>
<tr>
<th>After combustion</th>
<th>Stoker type</th>
<th>Fluidized bed type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom Gas Mixing</td>
<td>Boiler</td>
<td>Air dispersion plate</td>
</tr>
<tr>
<td>800 ~ 1000</td>
<td>800 ~ 1000</td>
<td>800 ~ 1000</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

FIGURE 5
Classification of refractory materials used for waste incinerators and industrial boilers.

* mark: mainly applied for waste incinerator
FIGURE 6 Clinker growth problem due to high temperature combustion in waste incinerator.

FIGURE 7 High temperature oxidation rate of SiC refractory at 1000 °C in typical oxidized gases.

FIGURE 8 Classification of main damage modes and factors in high-SiC bricks.
230–330 °C (against steam pressure: 2.9–9.8 MPa). Therefore, deterioration of RMs, also studs and anchors occurs by penetrated corrosive gas components such as HCl, SO2 etc. and then RLs are sometimes peeled off. Corrosion-resistant materials such as Cr–Si–Al–Fe-base alloys and stainless steels are used for studs and anchor.

### 3.2.2. Ceramic Tile System

Ceramic tiles with longer lifetime than conventional castable and plastic RMs have been developed and applied. Tile is manufactured by high-SiC and sometimes high-Al2O3 RMs in many cases [15]. As shown in Figure 10, the tiles are installed by using special hanging hook or fixing nut etc. with refractory mortar onto the outer surface of WWTs. The surface temperature of the tile linings reaches approximately 900 °C or lower by cooling of WWTs. SiC tiles are chemically stable and have a size of 200–300 mm and a thickness of 25–60 mm to prevent damages due to thermal stresses during operation. The RMs, production process, shape, and installation methods are determined based on the evaluation of the deterioration resistance, thermal shock resistance, etc.

### 3.2.3. Air-Cooled Structure

In order to prevent clinker trouble, air-cooled FWs shown in Figure 11 have been developed [16]. Surface temperature of high-SiC brick (with the thickness of 80–150 mm) is controlled by air cooling from rear side of FWs. The following performances for air-cooled FWs are required in furnace design.

1. Thermal shock resistance of FWs is required to prevent fracture damages against rapid temperature change such as start-up/shut-down and gas temperature fluctuation etc. Applied RMs, size, and support system are also determined considering individual operating conditions.
2. Heat- and corrosion-resistant structures are designed to keep a long lifetime for dead weight loading, thermal stress, corrosion reaction, etc.
3. Good workability of FWs is required against easy installation, repair of RMs, and using of refractory mortar.

### 3.3. Unshaped Refractory Linings

Unshaped (monolithic) RMs such as castable and plastic refractories are used mainly for hanging wall, WWTs, and curved parts etc. Application forms of each RLs were described in detail as follows.

#### 3.3.1. Castable Refractories

Commonly, fireclay, high-Al2O3, and high-SiC castable RLS have been used with a thickness of more than 100–150 mm. Casting molds are installed along FWs, and blended castable refractories are poured into moldings. Compressed or vibration casting methods and gunning are sometimes applied to disperse large particles homogeneously. The terms important for designing and operation of furnace are as follows:

1. RLs are divided by adequate cut score lines to prevent cracks and peeling off of RLs due to loading of thermal stress.
2. Anchor systems supporting RLs at high temperature are important to maintain long lifetime. Therefore, heat-resistant alloys and sometimes ceramics shown in Figure 12 [17] are selected for anchor materials.

<table>
<thead>
<tr>
<th>Bricks</th>
<th>Apparent density</th>
<th>Bulk density</th>
<th>Apparent porosity (%)</th>
<th>Crush strength / kg·cm⁻²</th>
<th>Thermal conductivity/ kcal·m⁻¹·hr⁻¹·°C⁻¹ at 1000 °C</th>
<th>Chemical composition/ mass%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fireclay</td>
<td>2.67</td>
<td>2.27</td>
<td>15.1</td>
<td>600</td>
<td>1.5</td>
<td>Al₂O₃ 43</td>
</tr>
<tr>
<td>High-alumina</td>
<td>3.19</td>
<td>2.48</td>
<td>14.2</td>
<td>800</td>
<td>2.0</td>
<td>Al₂O₃ 55</td>
</tr>
<tr>
<td>High-SiC (Silicate bonded)</td>
<td>3.12</td>
<td>2.63</td>
<td>15.8</td>
<td>817</td>
<td>13.5</td>
<td>SiC 90</td>
</tr>
<tr>
<td>High-SiC (Si₃N₄ bonded)</td>
<td>3.10</td>
<td>2.65</td>
<td>14.6</td>
<td>1500</td>
<td>14.1</td>
<td>SiC 73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Si₃N₄ 22</td>
</tr>
</tbody>
</table>
3. Slow drying rate at start-up is needed to prevent initial damages by quick water vaporization. Also, water penetration should be prevented at water cleaning in shut-down.

3.3.2. Plastic Refractories

Plastic RMs can be applied for boiler tube surface, curved parts such as among boiler tubes, narrow parts etc., due to its good workability, and can be installed as thinner linings than castable RMs. The design points important for application of plastic refractories are as follows:

1. Dense microstructure of plastic RMs is favorable for corrosion protection, but thermal shock resistance somewhat reduces. Also, homogeneous installation by using hammer and rammer etc. is required to maintain mechanically strong properties for prevention of erosion/corrosion damage etc.

2. Support systems as same as castables are important to keep high durability of RLs.

Recently, spray lining (gunning) methods have been commonly carried out to minimize working duration and to form homogeneous coated layer. Also, it is effective to give special properties such as corrosion resistance, anti-slagging etc. onto the surface of RLs.
4. ADVANCES IN HIGH-TEMPERATURE CORROSION-RESISTANT MATERIALS AND COATINGS

In order to combat corrosion in above-mentioned severe environments, the development and application of advanced CRMs and CRCs, and also evaluation methods of the corrosion resistance, have progressed as shown in Figure 13. Details are described in the following.

4.1. Corrosion-resistant Metal Coatings for Waterwall Tubes

The metal temperatures of the WWTs are relatively low depending upon the steam pressure as mentioned above. CRCs such as ceramic tile, metal spray coatings, weld overlay, or the cladding are applied on the tube surfaces of carbon steels or Cr–Mo steels as shown in Figure 14. Table 5 shows the examples of application and durability evaluation of various CRCs.

1. Spray Coatings

The metal spray coatings can be applied on site quickly with the thickness of approximately 150–500 μm. The actual application of Al/80Ni20Cr alloy flame spray coating on WWTs has started in 1985 [18]. Many tests in actual plants were conducted for WWTs and SHTs [19–24], and recently NiCrSiB alloy coatings by the high-velocity oxygen fuel (HVOF) process with high durability have been installed in many cases [25–27]. The coating systems (applied both on-shop and on-site) are selected in accordance with the workability, the severity of corrosion environment, the required lifetime, and the cost. Recently fused NiCrSiB alloy spray-coating layer by induction heating has been successfully applied to WWTs in South Asian countries. This coating has a homogeneous, dense, and chemically bonded thick layer of 0.8–1.5 mm. Lifetime of this coating shows five times or more compared to that of above as-sprayed coatings due to no permeability of corrosive gases and completely chemically bonded interface [28].

2. Weld Overlay and Composite Tubing

A dense and thick (2–3 mm) coating layer chemically bonded with the base metal can be obtained by the weld overlay. The durability of alloy 625 weld overlays, that is 10 years or longer, is higher than that of as-sprayed coatings [29]. In many cases for the WWTs, automatic MIG and shielded arc welding are adopted for both on-shop and on-site installations [30]. For the SHTs, also PPW (plasma powder welding) are applied [31], and the tube bending can be done even after weld overlay. For the alloy 625 weld overlay applied in WWTs of 500°C/9.8 MPa high-efficiency boiler, the maximum corrosion rates of approximately 0.1–0.2 mm/year have been observed by the 2-year field tests as shown in Figure 15 [32].

Mainly in Europe, composite tubings with cladding of CRMs (outer tube) to carbon steels and low-alloy steels (inner tube) are used for WWTs and SHTs. There are many application experiences of high Cr–high Mo–Ni base alloys including alloy 625 (Sanicro 63) [33].
4.2. Corrosion-resistant Alloys for Superheaters

The alloy design of seamless tubes for SHTs takes into consideration both corrosion-resistant and heat-resistant characteristics. Furthermore, excellent weldability, plastic workability etc. are required for fabrication of the boilers. Generally, at steam temperatures of approximately 350 °C or lower, carbon steel is used, while in the high-temperature region of 400 °C or higher, CrMo steels stainless steels and Ni-base austenitic materials are used. The addition of alloying elements such as Mo, Nb, and Si to Ni—Cr—Fe alloys is considered to be effective for corrosion resistance from the results of field corrosion
### TABLE 5 Example of Durability for Corrosion Resistant Coatings in WTE Boilers

<table>
<thead>
<tr>
<th>Coating process</th>
<th>Chemical composition (Materials)</th>
<th>Applied Condition (°C)</th>
<th>Parts</th>
<th>Metal temp.</th>
<th>Gas temp.</th>
<th>Durability (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spray Coating</td>
<td>Flame Al/80Ni20Cr</td>
<td>WW</td>
<td>230 − 300</td>
<td>700 − 900</td>
<td>&gt;3</td>
<td></td>
</tr>
<tr>
<td>Flame/ Fused</td>
<td>10Cr-Si, B-Ni base (12C)</td>
<td>SH</td>
<td>370 − 540</td>
<td>—</td>
<td>&gt;1</td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>Plasma/ Fused 15Cr-Si, B, Fe-75Ni base</td>
<td>WW</td>
<td>230</td>
<td>700 − 800</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>HVOF</td>
<td>Hybride Cap 18Cr-5Fe-5Nb-6Mo-Ni base (Diamalloy)</td>
<td>SH</td>
<td>370 − 540</td>
<td>—</td>
<td>&gt;0.4</td>
<td></td>
</tr>
<tr>
<td>DJ-1000</td>
<td>17Cr-4Fe-3.5B-4Si-Ni base (No-Fused)</td>
<td>WW SH</td>
<td>230 − 330</td>
<td>700 − 900</td>
<td>&gt;3</td>
<td></td>
</tr>
<tr>
<td>D-Gun DJ-1000</td>
<td>50TiO2-50 Alloy625 (Cermet)</td>
<td>SH</td>
<td>430 − 460</td>
<td>500</td>
<td>&gt;3</td>
<td></td>
</tr>
<tr>
<td>JP-5000</td>
<td>17Cr-4Fe-3.5B-4Si-Ni base</td>
<td>WW</td>
<td>230</td>
<td>700 − 900</td>
<td>&gt;3</td>
<td></td>
</tr>
<tr>
<td>Ex-4000</td>
<td>Modified Alloy625</td>
<td>WW</td>
<td>—</td>
<td>—</td>
<td>&gt;1</td>
<td></td>
</tr>
<tr>
<td>Plasma Jet</td>
<td>Modified PJ ZrO2 / Alloy 625, NiCrSiB Alloy</td>
<td>SH</td>
<td>430 − 500</td>
<td>510 − 650</td>
<td>&gt;3</td>
<td></td>
</tr>
<tr>
<td>Weld Overlay</td>
<td>MIG 21Cr-9Mo-3.5Nb-Al, Ti-Ni base (Alloy 625)</td>
<td>WW SH</td>
<td>270 − 330</td>
<td>700 − 1000</td>
<td>&gt;10</td>
<td></td>
</tr>
<tr>
<td>PPW</td>
<td>18Cr-14Mo-4W-Ni base (C-276M, 625M)</td>
<td>SH</td>
<td>470 − 510</td>
<td>510 − 650</td>
<td>(≈ 625)</td>
<td></td>
</tr>
<tr>
<td>MIG</td>
<td>23Cr-16Mo-1.6Cu-Ni base (HC-2000)</td>
<td>WW SH</td>
<td>260 − 480</td>
<td>870 − 1730</td>
<td>(0.52mm/y)</td>
<td></td>
</tr>
<tr>
<td>PAW</td>
<td>Alloy 625</td>
<td>SH</td>
<td>470 − 530</td>
<td>510 − 650</td>
<td>&gt;1</td>
<td></td>
</tr>
<tr>
<td>Laser Cladding</td>
<td>625, HC-22, NiCr 309L, 686 etc.</td>
<td>WW (Coal Fired Black Liquor)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

Remarks: WW (Waterwall), SH (Superheater)
The CRMs currently developed and applied are described as follows [35].

1. High Cr—high Ni—Fe Base Alloys

Austenitic stainless steels of 25Cr—14—20Ni steels such as 309S, 310S, and 310HcbN, NF709 which are modified 310S alloys, are mainly used for 400°C/3.9 MPa boilers, because the difference of corrosion resistance in each Ni—Cr—Fe alloy is small under the steam conditions below 400°C in relatively weak corrosion environments, as shown in Figure 16 [5]. Also alloy 825 and Sanicro 28 (composite tube) of 20—30% Cr 30—40% Ni alloys are sometimes used in a severely corrosive environment [36].

2. High Cr—high Mo—Ni Base Alloys

Alloy 625 and Sanicro 63 (composite tube) have good corrosion resistance under high-temperature steam conditions of higher than 450°C as shown in Figure 16. Also, it is known that HC-22 and JHN24 alloys, whose Mo content is higher than alloy 625, show excellent corrosion resistance. For high temperature applications of these high-Mo containing alloys, consideration of age deterioration of microstructures and mechanical properties is required.

3. High Si—high Cr—high Ni Alloys

QSX3 and QSX5 are modified 310S alloys with added Si of approximately 3%. Also, it is known that Nicrofer 45TM demonstrates good corrosion resistance. MAC-F and MAC-N alloys containing 21—28%Cr, 3.2—4.2%Si, Ni, and Fe have been applied for four years in the 500°C/9.8 MPa SHTs [37]. The corrosion resistance varies according to the locations in the SHTs where corrosion environments are different. Therefore, it is necessary to make best use of the materials in accordance with the corrosion environment.

FIGURE 15 Maximum corrosion thickness loss of alloy 625 weld overlay and NiCrSiB alloy spray coatings in 500°C/9.8 MPa boiler.

FIGURE 16 Change in maximum corrosion thickness loss of corrosion resistant superheater alloys with [Cr+Ni+Mo] concentration after 6000 hours exposure in metal temperature of 450°C and 550°C.
5. FIELD EXPERIENCES OF CERMET AND CERAMIC COATINGS FOR SUPERHEATERS

In order to improve the lifetime of SHTs at metal temperature of more than 430 °C, challenging for application of cermet and ceramics spray coatings have been carried out in commercial plant that operates under 500 °C/9.8 MPa steam conditions shown in Figure 17 [38]. TiO$_2$·625 alloy cermet coating by HVOF, Cr$_3$C$_2$·NiCr alloy cermet, and ZrO$_2$/Ni-base alloy dual layer coatings by the plasma jet process were exposed in SHTs that have metal temperature of 432–448 °C for 1.3–2 years. Figure 18 shows three spray processes used in this investigations. In this chapter, the results of lifetime evaluation of coatings in actual boiler are described in detail [39,40].

5.1. TiO$_2$·625 Cermet HVOF Coating

Before field evaluations, 50 hours laboratory corrosion tests were carried out to select the adequate composition of cermet coatings. D-gun and HVOF are both a kind of supersonic frame spray process that forms dense layer. Addition of 30–50% TiO$_2$ ceramic powder to alloy 625 improves largely corrosion resistance of coating in both spray processes as shown in Figure 19 and Figure 20. This effect is considered to be due to increase in corrosion resistance and coating density with less than 0.5% porosity [39].

In order to prevent corrosion in tertiary (alloy 625) and secondary SHTs (310HCbN) affected by soot blowing, 50% TiO$_2$·50%625 cermet coating that is selected in laboratory has been installed. Specimens sprayed on shop and on-site were both installed to the following three positions. In each position, on-site spraying was applied from the upstream side of the gas in the first row of SHTs where accelerated corrosion had been observed due to soot blowing attack [3]:

- Steam inlet and outlet portions of the tertiary SHTs (alloy 625 tube, OD.38.1 × thickness 8 mm)
- Steam outlet portion of the secondary SHTs (310HCbN, OD.38.1 × thickness 8 mm)

Erosion/corrosion resistance of the TiO$_2$·625 cermet exposed for 2 years was clarified to be high compared to typical CRMs as shown in Figure 21. Also, corrosion thickness loss of layer reduced linearly as shown in Figure 22. Therefore, lifetime of coating was estimated to be approximately 4 years (27,000 hours) for a coating thickness of 200 μm. This mechanism was considered to be due to the formation of protective oxide layer, mainly including TiO$_2$ and Cr$_2$O$_3$ from the EDX analysis results shown in Figure 23.
5.2. \( \text{Cr}_2\text{O}_3/\text{NiCr} \) Cermet and \( \text{ZrO}_2/\text{Ni-Base Alloy Plasma-jet Coating} \)

5.2.1 Test Procedure

Figure 24 shows a system flow of a field-applicable ultrasonic plasma-jet spray (UPJ) equipment. This system, the Plazjet-250 (250 kW), is characterized by a higher output (100 kW), higher voltage, and lower current (200 VDC/500 A) compared with general atmospheric thermal spray (APS) equipment (40 kW, 50 VDC/800 A). The system can generate a supersonic \( \text{N}_2/\text{H}_2/\text{Ar} \) plasma jet whose velocity is 2,000 m/s by a water-cooled spray gun [40].

Two types of dual-layer coating with alloy 625 and NiCrSiB alloy as the under-coat and yttria-stabilized zirconia (\( \text{ZrO}_2-8\%\text{Y}_2\text{O}_3 \), hereinafter abbreviated as YSZ) as the top-coat were applied to the tertiary SHTs on site. In addition, a cermet single-layer of \( \text{Cr}_3\text{C}_2\cdot75\%\text{NiCr} \) (CrC·NiCr), were applied to the secondary SHTs. This cermet coating has been successfully used for prevention of erosion damage in commercial power generation boiler SHTs burning coals including much Cl [41].

The existing tube surface has been corroded with large pits mostly on the upstream side from the combustion gas. Then, the surface of those tubes was slightly ground and grid-blasted before field spraying, while a part of the corroded surface was blasted without any grinding in order to examine the effect of surface condition. Superheater tubes were sprayed for a quarter of the circumference (approx. 90°) on the surface facing the gas upstream. Furthermore, new tubes were sprayed on shop for the entire surface of each tube, and were welded onto the front row SHs in each test portion.

An inorganic sealant was only applied to a half area in length of the shop-sprayed tubes to examine the effect of the sealant. Table 6 shows a list of the temperature conditions and sprayed materials. Figure 25 shows typical microstructures of the UPJ coating layers tested. The application specifications for each UPJ layer are shown below:

- **a.** \( \text{Cr}_2\text{O}_3 (150 \mu\text{m})/\text{YSZ} (150 \mu\text{m}) \) dual layer;
- **b.** NiCrSiB (150 \( \mu \text{m} \))/YSZ (150 \( \mu \text{m} \)) dual layer: An inorganic sealant was applied on the YSZ layer; and  
- **c.** CrC·NiCr (200 \( \mu \text{m} \)) single layer.

<table>
<thead>
<tr>
<th>Process</th>
<th>Principal of Guns</th>
<th>Features</th>
</tr>
</thead>
</table>
| D-Gun   | Powder, Spark plug, Barrel | • discontinuous spray process: 3 time/sec  
• particle velocity: 800-900 m/s  
• gas temperature: Approx. 3227°C |
| HVOF    | Powder, Oxygen/Fuel gas | • continuous spray process  
• particle velocity: 500-700 m/s  
• gas temperature: Approx. 3227°C |
| UPJ     | Plasma gas, DC (-), DC (+), Powder, DC (+), Cooling gas | • continuous spray process  
• particle velocity: 700-2000 m/s  
• gas temperature: Approx. 5723°C  
• electric power: 100 kW  
(DC200V, 500A) |

**FIGURE 18** Three kinds of spray coating equipments applied in field tests.
The following results were obtained regarding durability of the coatings in the field tests.

5.2.2. Field Examination Results

Figure 26 shows the external appearance of an on-site sprayed 625/YSZ layer in the steam inlet portion of the tertiary SHT after 1.3 years of exposure. Both 625/YSZ and NiCrSiB/YSZ layers showed an almost sound condition in the upstream side of the gas. In sealant-coated positions, the sticking between the YSZ layer and deposits was weak and the deposits could be easily removed. Therefore, the sealant is thought to have an effect of preventing slugging of ash. Also in the steam outlet portion of the tertiary SHTs, deterioration such as peeling, etc. was not found even after 1.3 years of exposure. The on-site sprayed layer was applied satisfactory even in the presence of many corrosion pits on the older surface, as shown in Figure 27.

On the other hand, peeling occurred in the CrC-NiCr layer in the secondary SHT, particularly in the soot blower-affected position after four months’ exposure as shown in Figure 28. Many hexagonal patterned cracks were found near the area of peeling, while in the same portion, limited erosion/corrosion damage of the YSZ layer occurred in a part of both the 625 and NiCrSiB/YSZ coatings after 8 months’ exposure.

The on-shop sprayed Ni base alloy/YSZ layer in the steam inlet portion of the tertiary SHTs maintains

---

![Figure 26](image_url)  
*Coating layer spalled off by the thermal shock testing*

**< Test Conditions >**

- Temperature: 550°C
- Test Time: 50h
- Synthetic ash: 11%KCl-22%K₂SO₄-12%NaCl-25%Na₂SO₄-20CaSO₄-10% PbO
- Coating quantity: 400 g/m² (coated on the one side of the specimen)
- Synthetic gas: 1000 ppm HCl-50 ppm SO₂-10%O₂-10%CO₂-20%H₂O-bal.N₂
- Test specimens: 10 mm × 20 mm × 2 mm (coating layers were formed on both surfaces of Alloy625)

**FIGURE 19** Corrosion weight loss of various D-gun coatings in laboratory tests.
**FIGURE 20** Corrosion weight loss of D-gun and HVOF coatings in laboratory tests.

**Test Conditions**
- Temperature: 550°C
- Time: 50h
- Synthetic ash: 11%KCl-8%NaCl-11%Na₂SO₄-4%ZnSO₄-20%Al₂O₃-46%CaSO₄
- Coating quantity: 400 g/m² (coated on the one side of the specimen)
- Synthetic gas: 1000 ppm HCl-50ppmSO₂-10%O₂-10%CO₂-20%HO-bal.N₂
- Test specimens: 10 mm × 20 mm × 2 mm (coating layers were formed on both surfaces of Alloy625)

**FIGURE 21** Maximum corrosion thickness loss of D-gun coating and tube materials exposed for 13800 hours in secondary superheater.

* Stem outlet of secondary SH: Tg = 761 K, Tm = 724 K. Stem inlet of secondary SH: Tg = 734 K, Tm = 702 K.
a satisfactory sound condition even after one-year exposure as shown in Figure 29. The coatings showed a lower deterioration tendency in compared with the on-site sprayed layer.

5.2.3. Results of Lifetime Evaluation

After one year and 1.3 years’ exposure, some on-site applied layers were partially cut off, and a detail examination was carried out.

Figure 30 shows typical microstructures of the on-site sprayed 625/YSZ layer in the steam inlet portion of tertiary SHTs. The formation of small pores was found within the coating layer, but no corrosion occurred in the Ni base alloy/YSZ interfaces. On the other hand, in the left and right side surfaces of the tube, porosity and cracking inside the coating layer were formed with coating thickness reduction. By the results of EPMA shown in Figure 31, a slight amount of Cl was found within the under-coated layer and interface in the sound coating layer. However, alkali and heavy metals that form low melting point components of the deposits were not found. It was clarified that the sealant and YSZ layer effectively act as a penetration barrier of corrosive components into the interface. Since the corrosive environments in the steam outlet areas of both tertiary and secondary SHTs were weaker than that in the steam inlet area of tertiary SHTs, the degree of deterioration and penetration of the corrosive components was small.

Figure 32 shows the hardness changes of 625 and NiCrSiB under-coated layers with exposure time. The hardness of the NiCrSiB alloy shows an increase with the metal temperature, but the hardness of the alloy 625 remains almost a constant value. Table 7 shows a list of the microstructural examination results for each part. It is thought that the lifetime of the Ni-based alloy/YSZ in the steam inlet part of the tertiary SHTs where corrosion conditions are most severe is 1.5 years or longer. From the many experiences that the durability of the conventional single NiCrSiB HVOF layers are several months in the same superheaters, it became clear that the durability of the Ni-base alloy/YSZ/sealant dual coating was considered to be remarkably improved. The durability of each spray coating is large in the order shown below including the results of TiO$_2$·625 cermet HVOF coating.

$$625/YSZ > \text{NiCrSiB}/YSZ > \text{TiO}_2\cdot625 \text{ Cermet HVOF} > \text{CrC}$$

Table 8 shows the comprehensive evaluation results of the durability of the coatings after 1.3 years’ exposure. These coatings can prevent largely tube thickness losses in each part of the SHTs. It could be confirmed that the 625, and NiCrSiB/YSZ spray coatings have a good durability of 3 years or longer in an aggressive corrosion environment near a soot blower, which cannot be achieved by single layer of as-sprayed metal spray coating. Practical applications of these spray coatings are expected to contribute to continuous plant operation, and reductions in maintenance costs.

6. DETERIORATION MECHANISMS OF COATINGS

6.1. Metallic Materials Forming Protective Oxides Film

Main corrosion factors influencing the corrosion rate are shown schematically in Figure 33. The gas temperature, especially the temperature gradient $\Delta T$ (gas temperature – metal
temperature), is considered to be the driving force for condensation of the vapor components in the gas. The chlorides concentration in the deposits shows a high level at locations where $\Delta T$ is large, and there is a tendency to form low melting deposits [42]. In addition, it is known that the amounts of Cl, SO$_4$ and alkaline and heavy metals affect the physical properties of the deposits such as, molten phase amount, permeability, etc. [43]. The corrosion peak appears around 20% in the molten phase amount of the NaCl/KCl/Na$_2$SO$_4$ mixture as shown in Figure 34. The penetration of corrosive gas components through the deposits and the presence of oxidizing constituents such as O$_2$ are considered necessary for this kind of corrosion reaction.

Moreover, severe thermal cycling acts on the tube surface due to gas temperature fluctuation and the use of soot blowers in the actual plants. Then, peeling-off and the regeneration of deposits and oxides layer are repeated. Figure 35 shows the EPMA results in the crack generation site of alloy 625 oxides layer at the soot blow-affected zone. The partial pressures of Cl and S
FIGURE 24  Schematic illustration of 100 kW plasma-jet spray coating equipment. For color version of this figure, the reader is referred to the online version of this book.

### TABLE 6 Test Conditions and Installed Positions of Ceramic Coatings in WTE Boiler Superheater

<table>
<thead>
<tr>
<th>Test positions</th>
<th>3rd superheater steam inlet</th>
<th>3rd superheater steam outlet</th>
<th>2nd superheater steam outlet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube Materials</td>
<td>Alloy625</td>
<td>Alloy625</td>
<td>310HCbN</td>
</tr>
<tr>
<td>Steam Temp.</td>
<td>445 °C</td>
<td>500 °C</td>
<td>450 °C</td>
</tr>
<tr>
<td>Gas Temp.</td>
<td>621 °C</td>
<td>510 °C</td>
<td>488 °C</td>
</tr>
<tr>
<td>Under / Top Coating Materials</td>
<td>625 / YSZ</td>
<td>625 / YSZ</td>
<td>625 / YSZ</td>
</tr>
</tbody>
</table>

Test Bundles
- On-site (for used tube) : First tube in No. 2, 3, 4 bundles
- In-shop (for new tube) : First tube in No. 9 bundle

FIGURE 25  Typical microstructures of spray coating layers before corrosion test. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 26  General view of alloy 625/YSZ layer sprayed on site near soot blower in each superheater position after 1.3 years exposure. For color version of this figure, the reader is referred to the online version of this book.
FIGURE 27  Microstructure of alloy 625/YSZ dual layer sprayed to corrosion pit of alloy 625 tube on site in tertiary superheater steam inlet position after 1.3 years exposure. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 28  Deterioration of Cr₃C₂-NiCr layer sprayed on site near soot blower of secondary superheater after one year exposure. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 29  Typical surface condition of dual coating layer sprayed on shop in tertiary superheater steam inlet position after one year exposure. For color version of this figure, the reader is referred to the online version of this book.
FIGURE 30  Typical circumferential microstructures of alloy 625/YSZ layer sprayed on site in tertiary superheater after 1.3 years exposure. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 31  EPMA images of alloy 625/YSZ layer sprayed in tertiary superheater steam inlet position after one year exposure. For color version of this figure, the reader is referred to the online version of this book.
gases on the corrosion interface are considered to rise under the deposits due to penetration of chlorides and sulfates through the cracks. This type of corrosion is so called “molten salt induced corrosion”, because the corrosion reaction becomes active, when the amount of deposits increases and a part of deposits melts [44]. If the protective oxides layer breaks down, the direct reaction with molten chlorides may occur as shown in the schematic drawing of the alloy 625 corrosion scale in Figure 36 [34]. The formation and self-healing of stable protective oxides layer, that can be considered to be like a kind of ceramic coating film, are important properties for CRMs applied to aggressive corrosion environments. From the configuration and properties of corrosion products distributing as chlorides, sulfides, and oxides from the nearest side to the interface, the corrosion in steady state is considered to be kept as a high temperature gaseous reaction such that chlorination/sulfidation/

**TABLE 7 Microstructural Examination and Evaluation Results of Coatings in Each Position**

<table>
<thead>
<tr>
<th>Positions (tube materials)</th>
<th>Bundle no.</th>
<th>On-site sprayed layer under / top</th>
<th>Effect of soot blower</th>
<th>Circumferencial position of tube</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary Superheater Stream Outlet (310HCbN)</td>
<td>4</td>
<td>Cr₂C₂·NiCr Single Layer</td>
<td>Large</td>
<td>Left 35° Gas upstream Right 35°</td>
<td>- Partial peeling-off near soot blower</td>
</tr>
<tr>
<td>Tertiary Superheater Steam Outlet (Alloy 625)</td>
<td>4</td>
<td>NiCrSiB / YSZ</td>
<td>Weak</td>
<td>O</td>
<td>- Many longitudinal cracking</td>
</tr>
<tr>
<td>Tertiary Superheater Steam Inlet (Alloy 625)</td>
<td>3</td>
<td>625 / YSZ</td>
<td>Large</td>
<td>△</td>
<td>- Partial peeling-off of YSZ layer in side of tube</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>625 / YSZ</td>
<td>Weak</td>
<td>△</td>
<td>- No corrosion scale formation of tube</td>
</tr>
<tr>
<td>Tertiary Superheater Steam Inlet (Alloy 625)</td>
<td>4</td>
<td>Weak</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>625 / YSZ</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Weak</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
</tbody>
</table>

**<Basis of evaluation>**

- ○: Totally sound condition
- △: Deterioration of sealant and top coat (YSZ)
- △: Consumption of top coat and deterioration of under coat
- ×: Consumption of under coat
### TABLE 8 Results of Lifetime Prediction for Ni Base Alloys/YSZ Coatings after 1.3 Years Exposure

<table>
<thead>
<tr>
<th>Position (materials)</th>
<th>Gas temp./ steam temp. (°C)</th>
<th>Coating materials</th>
<th>Corrosion environment</th>
<th>Lifetime of coatings</th>
<th>Position affected by SB</th>
<th>Position not affected by SB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ry SH Steam Outlet (310HCbN)</td>
<td>488/450</td>
<td>YSZ/625</td>
<td>Weak</td>
<td>Strong</td>
<td>&gt;3</td>
<td>≥1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YSZ/NiCrSiB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cr3C2-NiCr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tertiary SH Steam Outlet (Alloy625)</td>
<td>510/500</td>
<td>YSZ/625</td>
<td>Medium</td>
<td>Weak</td>
<td>&gt;3</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YSZ/NiCrSiB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tertiary SH Steam Inlet (Alloy625)</td>
<td>621/445</td>
<td>YSZ/625</td>
<td>Strong</td>
<td>Strong</td>
<td>&gt;3</td>
<td>≥1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YSZ/NiCrSiB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 33** Schematic illustration for the role of main corrosion factors influencing to corrosion rate.
FIGURE 34  Change in corrosion weight loss with molten phase content of deposits between coating test and crucible test in laboratory.

FIGURE 35  Breakdown of protective oxide layer and penetration of corrosive species in alloy 625 superheater tube influenced by soot blowing.
oxidation occur simultaneously. It is presumed that corrosion by the direct reaction with molten salt changes gradually to the gas reaction as the oxides layer grows. The influence of major factors such as temperature gradient, temperature fluctuation, molten ash amount, etc. to the corrosion rate was examined quantitatively by recent research [45]. Furthermore, the factors of characteristic intergranular corrosion should be investigated considering both the material (impurities, grain boundary precipitates, etc.) and the environment (stress, molten salt, etc.) [46].

6.2. Metal and Ceramic Spray Coatings

Figure 37 shows the deterioration mechanisms and bonding strength reduction of the metal spray coating layers used for a long period in a boiler. Corrosion of the base material and deterioration of the coating layer proceed due to the corrosive gases penetration (HCl, etc.) onto the base material/coating interface. Then, “swelling” of the coating layer occurs, the reduction of adhesive strength is accelerated, and finally breaking down of the layer occurs [18]. Accordingly, dense coating is indispensable for improvement of lifetime, and HVOF is preferable process. The material factors that govern durability are the corrosion resistance and open porosity of coating materials, the adhesive strength with base materials, the thermal properties such as thermal expansion coefficient, and the residual stress. The physical properties of the coating are largely dependent upon the spraying conditions. Recently, a quantitative evaluation method of lifetime by using small electric resistance methods [47] has been developed and applied, based on these deterioration mechanisms.

In the case of dual ceramic coating of Ni base alloys/YSZ, the same mechanisms are clarified in the field tests. The YSZ top-coated layer acts as a diffusion barrier for penetration of corrosive gas components shown in Figure 38 and is able to improve the lifetime of under coating and also base materials.

7. SUMMARY

Recently, the WTE plants are required to satisfy many additional performances such as the suppression of pollutants, high electric power generation efficiency, material recycling, etc. Various combustion methods and plant systems have been developed, therefore a kind of such HTC environment increased more wide and diversified. Although the main factor influenced to HTC is considered to be almost
FIGURE 37  Deterioration process of spray coating layer and reduce in bonding strength in severe corrosive environment.

FIGURE 38  Schematic illustration of deterioration mechanisms in Ni base alloy/YSZ dual layers.
same for many corrosion environments for many kinds of WTE boilers. It is believed that the development and application of CRMs and CRCs aim strongly at the right material in the right place with a reasonable total cost. The improvement in performances and lifetime of the high temperature parts has been supported by the development of CRMs and CRCs technologies including both metallic and ceramic materials. It is also required to advance the application of ceramics under adequate corporation with metals and plant engineering.

There are many subjects for future research for this kind of aggressive corrosion environment; it is expected that engineers and researchers in the field of CRMs and CRCs will take up the challenge.

REFERENCES


A New Thick Film Coating Technology-Laser Chemical Vapor Deposition

Takashi Goto

Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

1. INTRODUCTION

Nowadays, highly functional materials are used under severe conditions such as high-temperature, high-stress, and corrosive environments. Metallic materials have often been used as structural materials because of their superior properties such as high strength and ductility. Ceramic materials, in contrast, have inferior mechanical properties than metallic materials do. However, they have various functionalities such as high thermal/chemical stability and higher hardness than metallic materials have. Therefore, ceramic-coated metallic materials can function as good high-performance multifunctional materials.

In general, coating technology can be divided into dry (vacuum) and wet (solution) processes [1]. The dry process is further categorized into two subtypes: physical vapor deposition (PVD) and chemical vapor deposition (CVD). In PVD, a source material (target) is transported to a substrate forming film by a physical process, such as ablation and sputtering, without an accompanying chemical reaction. The composition may be readily conserved from the target to the film, where the thickness of the coating can range from nanometers to several millimeters. Because PVD would encompass the sequential processes of collision, adsorption, and deposition of particles and/or clusters, it is fundamentally a “line-of-sight” process. Thus, the backside or inside of holes can hardly be coated by PVD. CVD, in contrast, accompanies the chemical reactions of source gases and takes place where the gases can arrive. CVD is then not a line-of-sight process and has good adherence and conformal coverage. However, it is a thermally activated process; thus, it requires a high temperature. As a result, it is difficult to use CVD for coating on nonrefractory substrates such as low melting point metals and polymers, but it can be used to prepare thermally stable refractory ceramic films even on substrates with complex shapes [2].

CVD has many deposition parameters, such as temperature, total pressure, partial pressure of each source gas, and the geometry of the CVD chamber (how to heat the substrate, etc.). By controlling these parameters, various microstructures, that is, amorphous fine-grained polycrystals and epitaxial single crystals, can be prepared [3]. CVD has wide-ranging applications, typically in semiconductor thin film devices and for coating on cutting tools. The deposition rate of CVD is generally a few to several tens of micrometer per hour. Thus, it is mainly applied to the thin film coatings. However, thick film coatings also have many useful industrial applications, and CVD can be applicable for such coatings. This chapter discusses the
high-speed CVD used in the preparation of thick film coatings.

2. CVD FOR HIGH-SPEED DEPOSITION

The main deposition process in CVD is the diffusion of gases in a gas phase and their chemical reactions on a heated substrate. A schematic diagram of the CVD deposition process is depicted in Figure 1 [4]. The source gases diffuse from the main gas stream to the substrate through the gas boundary layer and are absorbed/reacted on the substrate. Then, the byproduct gases desorb from the substrate and diffuse out to the main gas stream through the gas boundary layer. These sequential steps are all thermally activated processes; therefore, CVD is often termed thermal CVD. Figure 2 depicts the general relationship between the logarithmic deposition rate and the reciprocal deposition temperature [5]. In a low-deposition-temperature region, the rate-controlling step could be a chemical reaction on the substrate, where the deposition rate exponentially increases with increasing deposition temperature. The activation energy in this temperature region is generally several tens to hundreds of kilojoules per mol. In a higher-deposition-temperature region, the deposition rate still increases, while the activation energy becomes rather small, generally less than tens of kilojoules per mol, because the rate-controlling step could be mass transport (diffusion) of gases (source gas or byproduct gas) through the gas boundary layer. Therefore, to increase the deposition rate of the CVD so that a thick film is obtained, as depicted in Figure 2 (broken line), the region in which the chemical reaction occurs should be expanded to a high-temperature region while keeping source gases easily accessible to the substrate.

For the highest deposition rate, CVD can be conducted at a high deposition temperature under the mass transport controlled region. The source gases should be appropriately thermally stable in the high-temperature region without becoming depressed via a homogeneous reaction in a gas phase. Halide gas is commonly used because of its moderate stability at a high temperature. This reaction is called halide CVD. The chemical/thermal stability of substrate materials is also crucial to high-temperature deposition. Graphite substrate is often used for preparing nonoxide films in a nonoxidizing atmosphere. Nonoxide thick films, typically Si₃N₄ and SiC, are prepared using halide precursors (such as SiCl₄ and CH₃SiCl₃) and graphite substrates at high deposition rates of 1–2 mm/h at 1700–1800 K [6,7]. Metal organic (MO) compound precursors can also be used in CVD, termed Metalorganic Chemical Vapor Deposition (MOCVD), but MO precursors are easily decomposed and product films (mainly oxides) are reactive with the substrate material at a high temperature. Thus, the deposition rate of MOCVD is generally lower than that of halide CVD.

The precursor gases are heated before arriving at the substrate, and the deposition reaction occurs in various places inside the CVD chamber—partially in a gas phase to form powder and partially on a CVD chamber wall to form premature deposits. The deposition rate of CVD is therefore significantly affected by the manner in which the precursor gases and the substrate are heated, that is, whether the substrate is indirectly heated by the chamber wall or is directly heated without heating the CVD chamber wall. CVD is also categorized by the type of chamber wall, that is, the cold wall type CVD and the hot wall type CVD. Figure 3 depicts the schematics of cold and hot wall CVD.

**FIGURE 1** A schematic diagram of the deposition process of CVD.
chambers. In the cold wall type CVD, the substrate is directly heated by either radiofrequency or microwave induction heating and electric current joule heating. In these heating methods, the gas phase reaction (homogeneous reaction) can be minimal; only radiation from the substrate may cause the homogeneous reaction. The deposition rate and deposition efficiency are generally high, whereas the uniform temperature zone in the CVD chamber is rather narrow. In the hot wall type CVD, a wide area in the CVD chamber can be uniformly heated. Thus, a large number of substrate pieces or wafers can be simultaneously coated; however, the deposition rate is usually low, around several micrometers per hour.

Lasers are versatile energy sources and have been applied to CVD, termed laser CVD. Laser CVD is a typical cold wall type CVD in which a laser directly irradiates the substrate (or only gas) but not the chamber wall [8]. Because a laser has high-energy photons, it can directly debond molecules photochemically, thereby causing deposition reactions. This process is called photolytic laser CVD. Because the deposition occurs without heating the substrate, CVD can be conducted at almost the room temperature. By using photolytic laser CVD, semiconductor device films have been prepared on n- or p-type Si single crystal substrates without affecting the substrate’s dopant distribution [8]. In contrast, a laser such as a CO2 laser or an Nd:YAG laser (infrared wavelength laser) has high thermal energy as that of a heat source. By focusing a laser beam on a specific area of a substrate, thermally activated chemical reactions and grain growth can occur in a localized area without a homogeneous reaction in a gas phase (lasers usually do not interact with gases), as depicted in Figure 4. In common thermal CVD, the deposition rate would be limited by diffusion in a gas phase in a high-temperature region, as shown in Figure 2. In pyrolytic laser CVD, the deposition zone (i.e. the laser beam size) is usually small; thus, the source gas can easily access the deposition zone. As a result, the deposition rate is not limited by diffusion and can be enormously high, more than several hundreds of meters per hour. Figure 5 demonstrates
the deposition rate of pyrolytic laser CVD as a function of laser power [9]. However, it is believed that pyrolytic laser CVD can be used to prepare only thin rods, whiskers, and nanotubes (i.e. one-dimensional deposits) at a high deposition rate. Therefore, wide-area films or plates cannot be prepared at a high deposition rate by pyrolytic laser CVD.

We have developed a new laser CVD to prepare thick films at deposition rates 100–1000 times greater than that of conventional CVD on wide-area and complex-shaped substrates [10]. The laser is broadly irradiated to a substrate and only the substrate surface is heated, that is, the entire substrate body is not heated. This methodology enables thick ceramic film coatings to be prepared on metal substrates at a high deposition rate. In this chapter, we discuss refractory thick film coatings on metal substrates using laser CVD for applying a thermal barrier yttria-stabilized zirconia (YSZ) coating on an Ni-based superalloy, an antiabrasive α-Al2O3 coating on cutting tools, and a hydroxyapatite (HAp) coating on a Ti implant.

3. YSZ THERMAL BARRIER COATING

YSZ is widely adopted for thermal barrier coatings (TBCs) on gas turbine blades made of an Ni-based superalloy because of its high thermal stability, low thermal conductivity, and a relatively large thermal expansion, which is close to that of the metal substrate. An Ni-based superalloy can withstand a temperature of around 1400 K, while the gas turbine can endure a temperature of around 1800 K. The temperature difference (several hundred degrees Celsius) between the working temperature of the gas turbine and the refractory temperature of the metal substrate is partially attributable to TBC. Atmospheric plasma spray or electron beam physical vapor deposition (EBPVD) has been applied to TBC [11]; however, the working temperature of TBC should be raised, and the lifetime should be extended by improving the quality of the YSZ coating. Thus, a new and more advanced coating process should be developed.

CVD is advantageous for preparing a YSZ coating with good conformal coverage and adherence and is also a promising technique for TBC [12]. However, the precursor and oxygen source gases are generally reactive in gas phase to form powder (homogeneous reaction) at a high gas concentration, and the deposition rate of oxide films is generally too low (usually several tens of micrometers per hour) to prepare a TBC coating with a thickness of around several hundreds of micrometers. Laser CVD can be used for preparing YSZ thick film coatings at high deposition rates without causing the homogeneous reaction in the gas phase because the laser would not generally interact with the source gases.

Figure 6 depicts the cross-section of the YSZ coating prepared by laser CVD [13]. A typical columnar texture, similar to that of EBPVD, can be prepared. Although the thermal expansion coefficient of YSZ is close to that of an Ni-based superalloy, a small difference still exists between the two, which might have caused thermal stress and delamination at the interface. The columnar texture is vertically aligned. Furthermore, the crack formation resulting from severe thermal stress would extend vertically as well. This tendency would prevent catastrophic delamination of the entire coating. Figure 6(b) shows a magnified view of the columnar grains. Many voids exist at the columnar boundary. Similar voids are also observed in the YSZ coatings by EBPVD, and they are known as being effective in the relaxation of thermal stress at the interface.

The CVD of YSZ films has been widely investigated because YSZ is used for oxide ion conductor films and as a buffer layer between a metal tape and a Y–Ba–Cu–O superconductor film. Figure 7 demonstrates the deposition rates of YSZ films as a function of deposition temperature by conventional CVD [14–18]. Attempts have been made for preparing a YSZ coating using CVD so that the coating can be applied as TBC. A deposition rate of 50–100 μm/h has been attained by employing a cold wall type CVD. However, the conventional CVD still does not have a suitable deposition rate for practical application. In contrast, the deposition rate of YSZ film by laser CVD is several to hundred times greater than that of conventional CVD. The chemical reaction around the substrate can be highly activated by laser irradiation. Therefore, the rate-controlling step of the deposition process is not a chemical reaction but rather a diffusion reaction (mass transfer) in a gas phase with low activation energy.

Figure 8 demonstrates the nanostructure of the columnar grains of the YSZ coating by laser CVD [5]. A large number of nanopores of about a few nanometers in diameter can be observed in the grains. Figure 9 depicts the relationship between the deposition rate and the thermal...
The conductivity of the YSZ coating by laser CVD [12]. The number of nanopores (i.e. density) is almost proportional to the deposition rate; in contrast, thermal conductivity is inversely proportional to the deposition rate. The thermal conductivity of the YSZ coating was one-third to one-fourth that of YSZ bulk ceramics. Thus, this coating is a promising candidate for TBC.

4. \(\alpha\)-Al\(_2\)O\(_3\) COATING FOR CUTTING TOOLS

WC–Co cemented carbide has been widely used as a cutting tool because of its high hardness and ductility. It is commonly coated with \(\alpha\)-Al\(_2\)O\(_3\) thick film to protect the substrate from thermal and mechanical degradation. CVD is most commonly used for the \(\alpha\)-Al\(_2\)O\(_3\) coating [19]. Al\(_2\)O\(_3\) has many poly types, mainly \(\alpha\), \(\gamma\), \(\kappa\), \(\theta\), \(\delta\); \(\alpha\)-Al\(_2\)O\(_3\) is the
most stable and hardest among them. The volume change by phase transformation at a high temperature during service will cause delamination of the Al2O3 coating on WC–Co. Therefore, the top coat on WC–Co should comprise the most stable \( \alpha \)-Al2O3 coating [20]. Because \( \alpha \)-Al2O3 is a high-temperature form of Al2O3, the deposition temperature for preparing its coating should be generally high, that is, >1200–1300 K. Such a high deposition temperature will often cause degradation of the WC–Co substrate; thus, the deposition temperature should be lowered. W is also a rare resource, and the substitution of WC–Co into TiN–Ni cermet has been investigated. The deposition temperature should be further lowered to prepare the \( \alpha \)-Al2O3 coating on TiN–Ni cermet because the outward diffusion of Ni and many interface reactions become significant at high temperatures. Therefore, a low-temperature and high-speed \( \alpha \)-Al2O3 coating process should be developed. We have applied laser CVD to prepare the \( \alpha \)-Al2O3 coating for enabling low-temperature and high-speed deposition [21–23].

Figure 10 depicts the deposition rates of the Al2O3 coating as a function of deposition temperature by both conventional CVD and laser CVD [24–28]. The \( \alpha \)-Al2O3 coating has been commercially prepared by halide CVD using mainly AlCl3 as the Al source and CO2 gas as an oxidant. This coating has commonly been prepared by halide CVD at >1300 K. A strong temperature dependence of deposition rate whose activation energy is 90–170 kJ/mol is characteristic of halide CVD, as depicted in Figure 10. The MOCVD has also been applied to prepare the Al2O3 coating, and amorphous and the \( \gamma \)-Al2O3 coating has often been prepared. The deposition rates by MOCVD are higher than those by halide CVD at low temperatures <1000 K.

The \( \alpha \)-Al2O3 coating can be prepared by laser CVD even at a low-deposition temperatures of around 1000 K at a significantly high deposition rate of several hundreds of micrometers per hour. This value is almost several hundred times greater than that of conventional CVD.

Figure 11 presents the relationship among deposition temperature, Al source rate, and microstructure [21]. The microstructures are depicted in Figure 12(a)–(c). The orientation can change from (300) to (006), primarily because of an increase in the Al source rate. In practical applications, (006)-oriented \( \alpha \)-Al2O3 is preferred because of its high hardness and smooth surface. Figure 13
demonstrates the cross-section of $\alpha$-$\text{Al}_2\text{O}_3$ prepared by laser CVD. Nanopores are contained in the $\alpha$-$\text{Al}_2\text{O}_3$ film as well as the YSZ coating, which is as effective as a thermally protective coating on cutting tools.

5. HAp COATING FOR DENTAL IMPLANTS

Ti-based alloys have been widely used for dental implants and artificial bone because of their high strength, relatively low Young’s modulus (which is close to that of the human bone), and their tendency to cause no allergy in living tissues. However, the regeneration of a bone on the Ti-based alloy takes a long time, usually several months in the human body. To accelerate this regeneration of bone (oste conductivity), the Ti-based alloy surface should be coated with a bioceramic film [1]. A typical bioceramic coating consists of bioabsorbable tricalcium phosphate (TCP) and a bioactive HAp film. Plasma spray, sol–gel, and sputtering are common to prepare bioceramic coatings [29]; however, highly oriented and well-adhered coatings have not been prepared by these techniques.

Laser CVD can be used for preparing various calcium phosphate compounds with a well-controlled morphology and orientation. Figure 14 demonstrates the effects of deposition temperature and Ca to P molar ratio in source
FIGURE 14  Effects of deposition temperature and Ca to P molar ratio on the crystal phase of calcium phosphate at $P_L = 30$ W (a) $P_L = 150$ W (b).

FIGURE 15  (002)-Oriented HAp (a) and (400)-oriented $\beta$-TCP films (b) prepared by laser CVD.

FIGURE 16  Regenerated HAp on HAp-coated Ti in Hanks’ solution after 259.2 (a) and 604.8 ks (b).
gases on the crystal phase of calcium phosphate at laser powers of 30 and 150 W [30]. The crystal phase and microstructure are dependent on laser power, deposition temperature, and Ca to P molar ratio. At a laser power of 30 W, HAp, β-TCP, and tetra calcium phosphate (TTCP, Ca₄P₂O₉) are formed depending on conditions. CaO codeposits at a low-deposition temperature. In a single phase, HAp can be prepared at an intermediate Ca to P molar ratio. At a laser power of 150 W, α-TCP and HAp are obtained in a single phase. Calcium pyroclore (Ca₂P₂O₇) and TTCP are also prepared together with α-TCP and HAp.

Laser CVD is advantageous in preparing highly oriented coatings, and it is known that the orientation of the coating significantly affects the regeneration of HAp in a simulated body fluid (SBF) such as Hanks’ solution. Figure 15 presents the (002)-oriented HAp and the (400)-oriented β-TCP films prepared by laser CVD. Cauliflower-like and pyramid-like microstructures are commonly observed at a low and high deposition temperatures, respectively. Figure 16(a) and (b) demonstrate the regeneration of HAp on HAp-coated Ti in Hanks’ solution after 259.2 and 604.8 ks, respectively. The HAp regeneration rate on HAp-coated Ti is higher than that on TCP-coated Ti. HAp demonstrates a hexagonal structure and has anisotropic crystal growth. The regeneration rate also significantly depends on the orientation of the HAp coating. Figure 17 summarizes the HAp regeneration rate on various bioceramic coatings in an SBF. The (002)-oriented HAp coating prepared by laser CVD has the highest regeneration rate among the other HAp coatings [31].

6. SUMMARY

Thin film coating processes by PVD and CVD have been indispensable technologies in many engineering fields. In contrast, thick film processing has not been well developed despite various useful applications, such as thermal barrier, anticorrosion, and antiabrasive coatings. Laser CVD has been traditionally employed to prepare small-scale deposits such as thin film or dots, whiskers, and nanotubes. Further, this technique cannot be applied to prepare thick and wide-area coating films. However, by using a high-power laser and maintaining appropriate conditions, the thick film coating even on a complex-shaped substrate can be prepared with a significant orientation. Laser CVD is also applicable in preparing oxide and nonoxide thick coatings [32,33]. In addition, it is a promising technique for a wide range of engineering applications.
REFERENCES


Aerosol Deposition Method for Room-Temperature Ceramic Coating and Its Applications

Jun Akedo
National Institute of Advanced Industrial Science and Technology, 1-2-1 Namiki, Tsukuba, Ibaraki 305-8564, Japan

1. INTRODUCTION

The manufacturing of ceramic materials usually requires sintering at temperatures higher than 1000 °C, which makes it difficult to compound or integrate ceramics with low-melting-point metals, glass, or plastics. This is a serious hindrance for upgrading electro-ceramic components and optical components. Up to now, a number of research works have been made to reduce the sintering temperature for the purpose of reducing energy consumption and implementing innovative functional components through integration with metal or glass materials. The aerosol deposition (AD) method is a unique coating technology based on the collision of solid-state fine ceramic powder with a substrate. This coating method has many advantages such as high deposition ratio, low process temperature, high adhesion force with a substrate in comparison with those by conventional thin-layer coating technology. During the AD process, submicron ceramics particles are accelerated by gas flow in the nozzle up to a velocity of several hundred m/s and sprayed onto the substrate under vacuuming condition. We found an interesting consolidation phenomenon of ceramic in this method over 10 years ago. During collision of fine particles and interaction with substrate, these ceramic particles, not only for oxide materials but also for nonoxide materials, formed thick, dense, and hard ceramic layers at room temperature. No additional heating for solidification of ceramic powder was required. We named this phenomenon “room temperature impact consolidation (RTIC)”. Same behavior as RTIC for metal fine particles was also observed. Consolidated ceramic powder with RTIC via AD method can be called as a high-density binder less ceramic green. Aerosol deposition is a novel and very attractive coating method for ceramic integration. In this paper, we explain the principle of the AD method and its application to ceramic layer formation for information/communication and energy technology applications.

2. AEROSOL DEPOSITION METHOD

The AD method is based on shock-loading solidification due to the impact of ultrafine ceramic particles with a surface [1,2]. First, particles are mixed with a gas to generate an aerosol. This aerosol is ejected through a nozzle at low pressure and impacted onto a substrate to form a thin layer. During impact with the substrate, part of
the particle’s kinetic energy is converted into thermal energy that raises the temperature of the particle and promotes bonding between the substrate and the particles and also between multiple particles.

Figure 1 shows the deposition apparatus and Table 1 gives the deposition conditions for the AD method. The AD apparatus is made up of two vacuum chambers connected by a tube. The first chamber is an aerosol-generation chamber and the second chamber is a deposition chamber. The deposition chamber is used for the formation and patterning of layers.

The aerosol chamber has a carrier gas system and a vibration system to mix the powder with the carrier gas. The aerosol generated in the aerosol-generation chamber is delivered to the deposition chamber by a pressure difference between the two chambers. The deposition chamber contains a nozzle, substrate with heating system, and a mask alignment system, used for making patterned layers. A rotary vacuum pump coupled to a mechanical booster pump is used to evacuate this chamber to a pressure of about 1–50 kPa during deposition. The particles flow from a micro-orifice nozzle deposits onto the substrate through

---

**Figure 1** Schematic (a) and prototype photograph (b) of aerosol deposition (AD) apparatus.
a patterning mask. Figure 1 (b) shows various size deposition machines. The scale-up of AD apparatus is easy because of a simple principle and low vacuuming condition.

Sintered, ceramic powders with a particle size range of about 0.08–2 μm are typically used as the deposition particles. After suspension in the carrier gas to form an aerosol, the aerosol is accelerated to several hundred m/sec through an orifice with a diameter less than 1 mm. Not all of the particles deposit onto the substrate, and the ratio of deposited to nondeposited particles strongly depends on the size and degree of aggregation of the particles. To form layers with acceptable density and material properties, particles with a particular size and morphology must be preferentially used. To generate a jet of particles with acceptable size and morphology, between the aerosol-generation and deposition chamber a deagglomeration device is used to break the particles apart as much as possible, and then an aerodynamic filter is used to select particles in a prescribed size range.

### 3. ROOM-TEMPERATURE IMPACT CONSOLIDATION (RTIC)

For the AD method, high-speed layer formation of ceramic layers at room temperature with high density and high transparency is possible by optimizing the particle diameter and deposition conditions. The result is a process that yields acceptable solidification without the need for high-temperature heat treatment. We call this process room temperature impact consolidation (RTIC) [2,3].

Figures 2 and 3 show micrographs of typical RTIC ceramic layers, cross-sectional SEM image of an as-deposited γ-Al₂O₃ layer at room temperature, comparison of its XRD profile with the profiles obtained for the raw powder and bulk material, and TEM images of as-deposited layer and starting powder. The formation of thick layers with a thickness over 100 μm is shown in Figure 2 (a). A high-transparency γ-Al₂O₃ layer on polymer film is obtained in Figure 2 (b), because of room-temperature coating. Using role-to-role coating system with a multinozzle spraying, the coating area could be enlarged to 60 cm × 1 m as shown in Figure 2 (c). The aerosol-deposited layers showed a relatively good thickness uniformity of 1.4%. The surface roughness (Ra) was less than 100 nm. The AD-layer shown in Figure 3 has high-density and randomly oriented polycrystalline nanostructures with crystallite size less than 20 nm. TEM and electron diffraction imaging did not show either amorphous layers or heterostructures at the boundary of crystal grains. XRD profiles confirmed that the spectral phases of the γ-Al₂O₃ particles were retained in the deposited layer. However, broadening of the spectra and a slight shifting of the spectral angle were observed. The reason for the change between the spectra of the particles and the deposited layer may be due to the reduction of the layer crystal size or distortion during deposition. The densification mechanism in detail is described in reference [3,24]. Clear lattice images in crystal grains less than 10 nm were observed, as well as uniform microstructures at the boundary between the substrate and the deposited layer. For γ-Al₂O₃ layers deposited at room temperature, layer density was over 95% of theoretical density and Vickers hardness was over 1600 Hv. Such γ-Al₂O₃ layers are acceptable for use as antiwear coatings [4]. The layer hardness increased with increasing particle-impact velocity and sometimes was higher than that of the bulk material, which was sintered at a high temperature. The particle-impact velocity during deposition was estimated by newly developed measurement system based on the combination of cutting a flow and “time-of-flight registration” measurements [5]. Critical particle velocities for acceptable RTIC ranged from 100 to 500 m/sec, and the velocity needed to create layers with acceptable hardness tended to increase with increasing sintering temperature of a particular ceramic material.

Acceptable room-temperature deposition was observed not only for oxide materials such as lead zirconate titanate (PZT: Pb(Zr₀.₅₂,Ti₀.₄₈)O₃) and nickel—zinc ferrite (Ni—Zn—Fe₂O₄) but also for nonoxide materials such as aluminum nitride (AlN) and magnesium diboride (MgB₂). In either case, particles with diameter greater than 80 nm and with single-crystal structure are needed to make layers with acceptable hardness. The crystal grain size of as-deposited layers was less than 50 nm, which was smaller than that of the starting particles. The reason for this is apparently that the starting particles break down

### TABLE 1 Typical Deposition Condition of AD Method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure in deposition chamber</td>
<td>0.05 – 2 kPa</td>
</tr>
<tr>
<td>Pressure in aerosol chamber</td>
<td>10 – 80 kPa</td>
</tr>
<tr>
<td>Size of nozzle orifice</td>
<td>5 × 0.3 mm², 10 × 0.4 mm²</td>
</tr>
<tr>
<td>Accelerating gas</td>
<td>He, N₂, air</td>
</tr>
<tr>
<td>Consumption of accelerating gas</td>
<td>1 – 10 l/min</td>
</tr>
<tr>
<td>Maintained substrate temperature during deposition</td>
<td>300 K</td>
</tr>
<tr>
<td>Scanning area (area of deposition)</td>
<td>40 × 40 mm², 400 × 400 mm²</td>
</tr>
<tr>
<td>Scanning speed of the nozzle motion</td>
<td>0.125 – 10 mm/sec</td>
</tr>
<tr>
<td>Distance between the nozzle and substrate</td>
<td>1 mm – 40 mm</td>
</tr>
</tbody>
</table>

Figures 2 and 3 show micrographs of typical RTIC ceramic layers, cross-sectional SEM image of an as-deposited γ-Al₂O₃ layer at room temperature, comparison of its XRD profile with the profiles obtained for the raw powder and bulk material, and TEM images of as-deposited layer and starting powder. The formation of thick layers with a thickness over 100 μm is shown in Figure 2 (a). A high-transparency γ-Al₂O₃ layer on polymer film is obtained in Figure 2 (b), because of room-temperature coating. Using role-to-role coating system with a multinozzle spraying, the coating area could be enlarged to 60 cm × 1 m as shown in Figure 2 (c). The aerosol-deposited layers showed a relatively good thickness uniformity of 1.4%. The surface roughness (Ra) was less than 100 nm. The AD-layer shown in Figure 3 has high-density and randomly oriented polycrystalline nanostructures with crystallite size less than 20 nm. TEM and electron diffraction imaging did not show either amorphous layers or heterostructures at the boundary of crystal grains. XRD profiles confirmed that the spectral phases of the γ-Al₂O₃ particles were retained in the deposited layer. However, broadening of the spectra and a slight shifting of the spectral angle were observed. The reason for the change between the spectra of the particles and the deposited layer may be due to the reduction of the layer crystal size or distortion during deposition. The densification mechanism in detail is described in reference [3,24]. Clear lattice images in crystal grains less than 10 nm were observed, as well as uniform microstructures at the boundary between the substrate and the deposited layer. For γ-Al₂O₃ layers deposited at room temperature, layer density was over 95% of theoretical density and Vickers hardness was over 1600 Hv. Such γ-Al₂O₃ layers are acceptable for use as antiwear coatings [4]. The layer hardness increased with increasing particle-impact velocity and sometimes was higher than that of the bulk material, which was sintered at a high temperature. The particle-impact velocity during deposition was estimated by newly developed measurement system based on the combination of cutting a flow and “time-of-flight registration” measurements [5]. Critical particle velocities for acceptable RTIC ranged from 100 to 500 m/sec, and the velocity needed to create layers with acceptable hardness tended to increase with increasing sintering temperature of a particular ceramic material.

Acceptable room-temperature deposition was observed not only for oxide materials such as lead zirconate titanate (PZT: Pb(Zr₀.₅₂,Ti₀.₄₈)O₃) and nickel—zinc ferrite (Ni—Zn—Fe₂O₄) but also for nonoxide materials such as aluminum nitride (AlN) and magnesium diboride (MgB₂). In either case, particles with diameter greater than 80 nm and with single-crystal structure are needed to make layers with acceptable hardness. The crystal grain size of as-deposited layers was less than 50 nm, which was smaller than that of the starting particles. The reason for this is apparently that the starting particles break down
during collisions and then each particle bonds together to form a nanocrystal layer.

If the carrier gas and pressure in the deposition chamber are properly selected so as to reduce the electrical discharge during particle deposition, transparent PZT and alumina layers with transmission efficiencies of 60–80% at wavelengths ranging from 450 to 800 nm could be formed at room temperature [3,6]. The reason this can be achieved is probably because during deposition the size of the defects in the starting particles is reduced to crystal sizes that are smaller than those found in sintered, bulk materials, and that are also smaller than the wavelength of visible light. Although distortions were included inside the layers, a rise of substrate temperature due to particle collisions was not observed, and ceramic powders could be solidified at room temperature without the use of binders. Although the formation of thin layers using RTIC does not require sintering, it produces high-density layers.

4. DEPOSITION PROPERTIES AND LAYER PATTERNING

Deposition of PZT at layer-deposition rates of 10–30 μm/min over an area 5 mm² was achieved. The particle
velocity and concentration in the aerosol jet downstream of the nozzle increased with increasing carrier gas flow rate, which increased the deposition efficiency. The deposition efficiency was also strongly affected by properties of the starting particles, such as the average particle size and size distribution, and mechanical and surface properties [7]. The deposition efficiency of the AD method is about 30 times higher than conventional thin-layer-formation methods such as sputtering, which have a deposition rate of only about several tens of nm/min for an area of 5 cm². Therefore, the AD method is attractive for manufacturing processes because it permits high throughput.

In conventional thin-layer and thermal spray-coating methods, surface cleaning before deposition is needed to achieve acceptable adhesion with a substrate. On the other hand, the AD method does not require such pre-cleaning, because during the initial stage of particle impaction with the surface, similar to sand-blasting processes, surface contaminants such as dirt and oils are removed by the initial particle collisions. Deposition automatically begins when the surface becomes sufficiently clean. The layer adhesion strength to glass and metal substrates was in excess of 20 MPa. However, to realize maximum adhesive strength, suitable hardness and elasticity of the substrate are needed. The effect of ceramic particles sprayed onto a substrate change from deposition to erosion [8], depending on the particle diameter, velocity, and angle of incidence of the particle jet to the substrate. These factors also influence the layer density and surface roughness.

Ceramic layer patterning can be achieved by using a stencil mask, which are made from metal and hard polymer materials. A particle jet is impacted onto the substrate through a pattern mask that contains many openings, each of which is on the order of 50 μm. In this

FIGURE 3 Microstructure of α-Al₂O₃ layer deposited at room temperature by AD method. (a) Cross-sectional SEM image of as-deposited layer at RT; (b) Comparison of XRD profiles; and (c) TEM images of starting powder and as-deposited layer at RT. For color version of this figure, the reader is referred to the online version of this book.
case, it is important to consider the aerosol jet flow in the deposition chamber and through the mask orifices. If the pressure in the deposition chamber is not sufficiently low, the particle jet is scattered by the edge of the openings in the mask and the resulting mask pattern is not preserved on the substrate.

To achieve acceptable patterns through a mask with acceptable detail, however, the angle of incidence of the particle jet must be kept within a specified range. This is because the angle of incidence affects the flow patterns on the downstream side of the mask, which in turn affects the deposition efficiency and the degree to which the mask profile is distorted. Figure 4 shows a thick, patterned PZT layer deposited onto Si, SUS, and Pt/Si substrates [9] under optimum deposition conditions. A ceramic microstructure with a 50 µm linewidth and aspect ratio greater than 1 can be patterned by controlling the substrate heating temperature and starting particle properties. To make finer pattern of AD-ceramic layers, the lift-off process using the photoresist was introduced in Figure 5 (a) [10]. PZT fine line pattern with the thickness of 2 µm and the width of 10 µm and α-Al₂O₃ fine line pattern with the thickness of 1.5 µm and the width of 8.5 µm were obtained in Figure 5 (b). Formation of crystallized ceramic layer at room temperature as particular characteristics of AD method provides an ability of ceramic fine patterning. Control of the cross-sectional profile of patterned photoresist was important to improve the making of fine ceramic pattern. The ductile property of photoresist after post-baking is very important to improve erosion resistance during AD process. More investigations are needed to unveil the influence of particle diameter of starting powder on AD process for the patterning properties.

FIGURE 4  Patterning properties of thick ceramic layers on AD method using mask deposition method.
5. ELECTRICAL PROPERTIES AND RECOVERING PROPERTIES BY HEAT TREATMENT

As-deposited AD layers deposited at room temperature generally have thermal insulation and electrical breakdown characteristics that exceed that of the bulk material. For example, the electrical breakdown of α-Al₂O₃ and PZT exceeds 3 MV/cm and 500 kV/cm, respectively. Such electrical characteristics can be useful for developing devices such as electrostatic chucks [11]. Although PZT layers deposited at room temperature exhibit piezoelectric and ferroelectric behavior, their properties are unacceptable for practical applications because of structural defects introduced during deposition. By post-annealing in air at temperatures ranging from 500 to 600 °C, grain growth of fine crystals and defect recovery in AD layers has been observed, dramatically improving ferroelectric properties. The dielectric constant (ε) and piezoelectric constant (d₃₃) of post-annealed layers formed at 600 °C were 800—1200 and —100 pm/V [12], respectively, which is comparable to that achieved with conventional thin-layer-formation methods. Moreover, the electrical breakdown (<1 MV/cm) and Young’s modulus (>80 GPa) of AD layers exceeded those obtained with conventional thick-layer formation technology. By post-annealing at 850 °C, remnant polarization (Pr) of 38 μC/cm² and coercive field strengths (Ec) of 30 kV/cm were obtained in the PZT layers [13]. Even without adding special additives to the feed particles or using special procedures, compared to conventional screen-printing methods, the AD method permits a 300—400 °C reduction of the process temperature. For layers deposited by using both conventional and AD processes, Figure 6 shows the effect on electrical properties of heat-treating layers. For both conventional-bulk and thin-layer processes, heat treatment at more than 600 °C was required to obtain crystallization and densification of the layers. The most important characteristic of the AD method compared to thin film and bulk process is that the as-deposited layer at room temperature has denser crystallized structure.

Recently, several groups presented several papers about the improvement of lead-free piezoelectric layer for KNN, BTO, and BNT system using AD method [14—17].

FIGURE 5 Fine patterning of ceramic layers deposited by AD method using lift-off process with photoresist. (a) Flow of lift-off patterning process of AD-ceramic layer. (b) Optical photography of PZT fine pattern (thickness: 2 m) on glass substrate. For color version of this figure, the reader is referred to the online version of this book.
They reported that the AD process was found to be very effective to prevent the vaporization of potassium (K), sodium (Na), or bismuth (Bi) for the formation of dense lead-free piezoelectric KNN and (Bi0.5Na0.5)TiO3 (BNT) layers. The higher density and slight crystal orientation were suggested as being the main reasons for the improved properties.

6. ANTIPLASMA CORROSION COMPONENTS USING AD-YTTRIUM OXIDE LAYER [18]

Figure 7 shows the cross-sectional SEM and TEM images of the yttrium oxide layers fabricated by the AD method at room temperature. Table 2 shows the electrical and mechanical properties of the layers. Conventionally, to sinter yttrium oxide requires the hot isostatic pressing (HIP) process with more than 1700 °C. However, the AD method attained not only the dense consolidation at room temperature but also that the AD layer has a larger electrical insulating property and mechanical strength than bulk hitter. Figure 8 shows the test result of the antiplasma corrosion. Even before exposed plasma (Figure 8 upper images), the AD-yttrium oxide layer had certainly better surface smoothness than the bulk yttrium oxide and the layer produced by the thermal spray coating; after exposed plasma (Figure 8 lower images), the significant differences were found between the AD-yttrium oxide layers and

| Table 2: The Electrical and Mechanical Properties of AD-yttrium Oxide Films |
|---------------------------------|------------------|
| Volume resistance (R.T.)       | $>10^{14}$ Ω cm  |
| Breakdown voltage              | 150 V/µm         |
| Vickers hardness                | 9.2 GPa          |
| Adhesion force                  | 80 MPa           |

FIGURE 6 Recovering electrical properties by heat treatment. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 7 The fine structure of yttrium oxide films formed by the AD method at room temperature. For color version of this figure, the reader is referred to the online version of this book.
that produced by the other processes, i.e. the AD-yttrium oxide layer was free from pores, although noticeable pores were found in the yttrium oxide samples fabricated by the other processes. This result demonstrates that the AD method considerably improved the antiplasma corrosion and the surface smoothness of yttrium oxide. Owing to the attainment about the AD method, we achieved the antiplasma corrosion, the wear resistance, and the reduction of the adsorption gas and the particle emission that are required for the next-generation semiconductor manufacturing industries. Figure 9 shows the large area ($50 \times 50 \text{ cm}^2$) AD-yttrium oxide coating and the electrostatic chuck produced by TOTO Co. Ltd [18]. The success of the development is due to the distinct feature of the AD method that can form good-performance ceramic layers for the antiplasma corrosion and the wear resistance at low cost.

7. APPLICATION TO MEMS DEVICES

Because piezoelectric materials act both as sensors and as actuators, they have a wide range of potential applications,
such as ink-jet printers, high-speed actuators for nanopositioning, and micro-, ultrasonic devices. To realize integrated microdevices with piezoelectric materials, thin-layer deposition technologies and associated microfabrication methods are being actively studied in the research fields of micro-electro-mechanical systems (MEMSs) and micro-total analysis systems (μ-TASs). The required thickness of piezoelectric layers for these applications is from 1 to about 100 μm (this thickness range is called “thick layers”).

High-speed resonance-type optical microscanners are expected to be a key component in future laser displays and retinal projection-type displays. For such applications the requirements are scanning frequency greater than about 30 kHz, scanning angle greater than 20°, mirrors sufficiently large to prevent distortion, and reduction of drive voltage. It is thought that piezoelectric layers will be desirable for such applications because of their simple structure and high actuation force. However, these devices are not easy to make using conventional thin-layer deposition technologies combined with MEMS fabrication processes, because of low deposition rates and complicated etching processes.

Using the AD process, a high-performance optical microscanner with a scanning speed at a resonance frequency over 30 kHz and a scan angle (peak-to-peak value) over 30° in atmospheric environment was successfully fabricated by the deposition of the piezoelectric materials at a high rate onto the scanner structure fabricated by Si-micromachining, as shown in Figure 10 [19]. Neither mirror bending nor deformation of the scanning laser beam was observed by the thick structure of the microscanner device. This scanner performance was higher than that of scanners made with conventional methods.

However, MEMS process technology has several problems: processes are complicated, products are costly, and many facilities are needed. In addition, Si-based optical microscanners are made from silicon and can be easily broken due to mechanical shock and high stress concentration near the mirror. It is very attractive to use a metal substrate instead of silicon in optical microscanner devices because of a reduction of the device cost and an improvement of the ambient durability by changing brittle silicon to ductile stainless steel. Figure 11 (a), (b) shows a schematic and an optical image of
a metal-based optical microscanner. An optical mirror was attached using two narrow torsion hinges from both sides at the center of gravity of the mirror (symmetrical mirror structure). The other sides of these hinges were connected to the scanner frame and one edge of the scanner frame was cramped and fixed with a heavy block. A piezoelectric layer was directly deposited on the stainless-steel scanner frame beside the mirror. The $d_{31}$ actuation mode of the piezoelectric layer bends the scanner frame and induces the vibration of a “Lamb wave” in the scanner frame [20]. Lamb waves are two-dimensionally guided and concentrate at the narrow torsion hinges holding the mirror through the specially designed metal scanner structure. The waves change with the torsion vibrations of the mirror caused by the vertical difference in the direction between progressing waves and the torsion hinge. The torsion vibration vertically scans the mirror axis. A large optical scanning angle (80°) was achieved at a high resonance frequency (30 kHz) at a driving voltage of 15 V (peak-to-peak) in ambient air using the prototype scanner device. Then low-power consumption laser projection system using this optical microscanner was obtained in Figure 11 (c).

In addition, diaphragm actuators [21] used for micropumps can be made from thin Si membranes deposited with the AD method and micromachined. These have flat frequency response and good symmetrical elastic deformation over a wide frequency range. During fatigue testing (stability of long time performance) under high driving electrical fields, there was neither de-polarization nor peeling of the AD deposited layer from the substrate [13]. The amplitude of this actuator was 25 μm at a resonance frequency of 22.4 kHz and a driving voltage of 8 V. This performance is suitable for applications using micromixers and micropumps.

8. POTENTIAL OF ENERGY APPLICATION

AD method allows room-temperature coating of ceramics materials; therefore, the thermal diffusion at the interface between ceramic and metal/polymer in the material integration can be reduced. The realization of low-cost and high-throughput wide area coating for energy related devices can be expected. Figure 12 (a) shows the all solid-state Li-ion battery with LCO/LATP/LTO structure, which was fabricated by AD method at room temperature. The charge and discharge properties were confirmed [22]. Figure 12 (b) shows the superconductive material MgB$_2$ thick layer deposited at room temperature by AD method. The phase transition of superconductivity was observed at 25–33 K [23]. These results indicate the potential of AD method application to energy-related devices. In the ceramic fuel cell (SOFC), the interconnected electrode AD layers assisted to improve the device lifetime, and the ionic conductive electrolyte AD layer reduces the internal resistance. Now, other applications such as CIGS solar cell, TiO$_2$ dye-sensitized solar cell (DSSC), and thermoelectric conversion devices are going to be fabricated.
9. FUTURE PROSPECTS FOR USING AD METHODS IN MATERIAL INTEGRATION TECHNOLOGY

Figure 13 shows the application load map of AD method toward the future. The application in left-lower (small size and high cost) area is microelectric devices, which have been researched and developed since 2000–2008 in NEDO national project “Nano structure forming for advanced integration ceramic technology” [24]. Now application developments are going on the right-higher area toward the future, which indicates the large-size and low-cost devices such as energy-related devices.

The AD method is a nonthermal equilibrium process, which solidifies feed particles at room temperature to form layers. This differs from thermal spray-coating methods, which use higher temperatures for layer deposition. There are two important features of the AD method. One is that the AD deposition rate is higher than that of conventional thin-layer processes, because the deposition material refers to particles, which carry substantially more mass than the molecules used in conventional methods such as sputter deposition. In addition, it is easy to form complicated oxide thin layers with the AD method, because the crystal structure of the starting particles is preserved during deposition. Therefore, the AD method is useful for developing composite or integrated materials with various ceramics, metals, and polymers. However, because many defects are introduced during deposition, a process such as annealing is needed to achieve acceptable electrical properties of the deposited layers. One of the origins for these defects may be that the crystals in the starting particles are crushed during collision with the substrate. Or it may be that surface defects of the starting particles are introduced into the deposited layer.

Because integration of ceramics with low-melting-point materials is important for a broad range of applications, reduction of process temperature is a requirement of future fabrication techniques. To achieve these goals with the AD method, the deposition mechanism, formation of defects in...
the layer, and recovery of electrical and mechanical properties by post-deposition annealing will be investigated in detail. Establishment of methods for controlling the particle size and degree of aggregation of the starting particles is also important.

REFERENCES


Akedo J, Lebedev M. Effects of annealing and poling conditions on piezoelectric properties of Pb(Zr_{0.52}, Ti_{0.48})O_3 thick films formed by aerosol deposition method. J Cryst Growth 2002;235: 397–402.


Akedo J, Sakata H, Lebedev M. Microstructure and superconductivity of dense MgB_2 thick layer on metal and Si substrate formed at room temperature with novel method: aerosol deposition method; 2003. The 5th International Meeting of Pacific Rim Ceramic Societies Incorporating the 16th Fall Meeting of the Ceramic Society of Japan, Nagoya, Japan.

Ceramic Powders for Advanced Ceramics: What are Ideal Ceramic Powders for Advanced Ceramics?*

Shigeyuki Somiya,† Sridhar Komarneni**,† and Rustum Roy**,†
†Professor Emeritus of Tokyo Institute of Technology, Japan, ‡Materials Research Institute, Materials Research Laboratory building, University Park, Pennsylvania 16802, United States, †Passed away on August 26, 2010

Chapter Outline
1. Introduction 863
2. Characteristics of Powders 864
3. Methods to Produce Fine Ceramic Powders 866
4. Hydrothermal Syntheses and Characteristics of Hydrothermal Powders 871
5. Summary 879
Acknowledgments 879
References 879

1. INTRODUCTION

Powders play a key role in many materials, such as ceramics, catalysts, cosmetics, food, and medicine. Many books and articles have reported on the particles and powders used for sintered ceramic bodies as final products: for example, there are articles described or edited by C. R. Veale [1]; A. Kato and T. Yamaguchi [2]; P. Vincenzini [3]; D. W. Johnson, Jr. [4]; D. L. Segal [5a,5b]; S. Saito [6]; S. Shirasaki and A. Makishima [7]; G. Jimbo [8]; R. J. Brook [9]; S. Somiya, T. Hayashi, and K. Asaga [10]; D. Ganguli and M. Chatterjee [11]; TIC Editorial Office [12]; and P. Vincenzini and J.-F. Baumard [13].

There are many scholars who mentioned important points of ceramic powders for ceramic-sintered materials. According to A. I. Aksay of the University of Washington [10], presently at Princeton University, and who is well known for work on ceramics before firing and mullite ceramics, the important powder characteristics are as follows:
1. Pore distribution.
2. Particle size distribution.
3. Shape of the particles.
4. Size of the particles.

As for H. K. Bowen of the Massachusetts Institute of Technology (MIT) [10], presently at Harvard University, and who is an author of “Introduction to Ceramics, 2nd Edition,” Wiley, New York, the following are important:
1. Use of fine powders.
2. Narrow particle size distribution.
3. No agglomeration.
4. As much sphericity as possible.
5. Uniformity of chemical composition.

According to J. A. Pask of the University of California, Berkeley, California [10], who was one of the pioneers of modern ceramics, the important characteristics are as follows:
1. Absence of coagulation of particles.
2. Particle size.

† A part of this article was originally published in Transactions of the Materials Research Society of Japan, 35(3), 473 (2010). This article is dedicated to Professor Rustum Roy.

Copyright © 2013 Elsevier Inc. All rights reserved.
R. Roy of The Pennsylvania State University [14,10] developed the sol–gel process in 1949–1956, which has been cited formally and informally >30,000 times for making fine and pure powders of virtually any composition in ceramics. He stated the following:

1. Control of chemical composition.
2. Particle size.

According to O. J. Whittemore, Jr. of the University of Washington [10], who studied ceramic products, the important characteristics are as listed below.

1. Size of the pores.
2. Shape of the pores.
3. Distribution of the pores.

A. Krauth who worked at Rosenthal, Germany (a famous company for Christmas decorative dishes) [10], stated the following:

1. Preferable to be able to sinter green bodies <1500 °C.
2. To sell products at a reasonable cost.

Each scholar has mentioned several important characteristics of powders from a research standpoint. We describe the characteristics of real powders in Section 2.1 and ideal powders in Section 2.2.

2. CHARACTERISTICS OF POWDERS
   [8,11,12,15]

In Section 3, we explain the processing of powders based on phases from which the powders are produced: the processing of powders from the gas phase in Section 3.1, the liquid phase in Section 3.2, and the solid phase in Section 3.3. We describe the major characteristics of real powders or powders prepared by using various techniques in Section 2.1 and the characteristics of ideal powders in Section 2.2.

2.1. Characteristics of Real Powders

Characteristics of powders prepared from the gas phase are as follows [8,15]:

1. Ability to produce particles with specific sizes and with a narrow distribution of sizes because it is easy to adjust and control preparation conditions.
2. Ability to produce nonaggregative particles.
3. Powders do not need crushing because particles are fine and are not aggregated.
4. Ease of obtaining high-purity powders because the starting materials are highly pure and controlled processes do not add impurities.
5. Ease of producing powders, such as oxides, carbides, and nitrides because it is easy to control the atmospheric condition for producing powders.

In addition, the following powders may be produced [8]:

1. Both single crystal and polycrystal powders with crystallinity and crystal structure that are easily controlled.
3. Both homogenous powders (i.e. powders with uniform chemical composition) and coated powders (i.e. powders coated with various components).
4. Both nonaggregate and aggregate particles with different porosities including nonporosity.

In order to explain the characteristics of powders prepared from the liquid phase, powders may be grouped into two types: (a) powders prepared by precipitation and (b) powders prepared by spray drying (i.e. a method to remove the solvent). Characteristics of powders are shown below [8,15]:

(a) Powders prepared by precipitation.
   1. It is easy to control small amounts of components because the mixing of small quantities by using liquids is easy.
   2. The rate of the formation of powders through the liquid phase is low, and hence, it is easy to control particle size, the distribution of particle size, crystallinity, the uniformity of composition, etc. Characteristics of powders prepared from both the liquid and the gas phase are able to be controlled, but it is easier to control powders prepared from the liquid phase than those from the gas phase.
   3. In general, washing, filtration, the removal of some chemicals such as precipitants, dry process, etc. are necessary.
   4. In many cases, particles are aggregated.

(b) Powders prepared by spray drying.
   1. In general, particles are spherical and the flowing of powders is good.
   2. Multicomponent powders can be produced because it is easy to control the chemical composition by changing the concentration of the solution.

In addition, the following powders may be produced [8]:

1. Powders of components that are easily controlled.
2. High-purity powders.
3. Amorphous powders.

Characteristics of powders prepared from the solid phase also depend on the preparation methods. Two conventional methods are (a) the thermal decomposition of solid [15] and (b) solid-state reaction [15].
The characteristics of powders are shown in the following:

(a) Powders prepared by thermal decomposition.
1. Particles have defects and strain because atoms of the particles are difficult to move freely.
2. Specific surface areas of particles are fairly large.
3. Powders become hydroxide, carbonate, or both because powders are active at room temperature and absorb water, carbon dioxide, or both from the atmosphere. Thus, it is necessary to consider the storing conditions.

(b) Typical powders prepared by solid-state reactions.
1. Powders tend to become inhomogeneous because conventional solid-state reactions are inhomogeneous.
2. Powders tend to become impure because products generally need to be crushed to obtain powders.

2.2. Characteristics of Ideal Powders

We described the characteristics of powders from the gas, the liquid, and the solid phase in the previous section. What then are the characteristics of ideal powders for advanced ceramics?

In 1984, T. Yamaguchi, Keio University, described the following characteristics of ideal powders to obtain highly dense and uniform sintered bodies sintered at temperatures as low as possible under ideal conditions of forming and sintering, and mixing of ideal amounts of additives with powders if necessary [16]:

1. Particles are highly dense.
2. Particles are fine, and their particle size distribution is narrow.
3. Particles are equiaxed. If possible, particles are spherical.
4. Particles with soft agglomeration are acceptable even though agglomeration occurs.
5. Particles or powders have the same single phase and are highly pure.
6. Particles have uniform properties such as chemical and physical uniformity. Examples of such properties are chemical composition distribution (e.g. impurities and additives), defects distribution, strain distribution, and thermal history. Chemical uniformity would be achieved to mix the particles thoroughly. Meanwhile, physical uniformity would be achieved to decrease differences of thermal history of particles.

In addition, he mentioned that the required values are dependent on the processes that are further used. For example, the requirements of ideal powders for conventional sintering are different from those for hot pressing.

In 1984, as per E. A. Barringer and H. K. Bowen of MIT, the four characteristics of ideal ceramic powders for sintering are as follows [17]:

1. Particle sizes are 0.1—1 μm.
2. Particle size distribution is narrow.
3. Particles have equiaxed shapes.
4. Particles are nonagglomerated.

Characteristics of ideal sinterable powders for advanced ceramics were mentioned by D. L. Segal [5b], who worked at Harwell Laboratory, United Kingdom Atomic Energy Authority, in 1991 and published a book titled “Chemical Synthesis of Advanced Ceramic Materials” in 1989 and an article titled “Powders for Electronic Ceramics” in 1991:

1. Submicron size.
3. Narrow size distribution.
4. High chemical purity.

In addition, he mentioned the following:

5. High compositional homogeneity.

Ideal powders for sintering should have the following factors based on the ideas that appeared in the Handbook of Hydrothermal Technology described by K. Byrappa of the University of Mysore and M. Yoshimura of the Tokyo Institute of Technology in 2001 [18]:

1. Fine particles <1 μm.
2. Soft or no agglomeration.
3. Narrow particle size distribution.
4. Spherical or equiaxed morphology.
5. Controllable chemical composition.
6. Controllable microstructure.
7. Uniformity.
9. High density.
10. Few defects.
11. Low stress.
12. Reactivity; sinterability.
13. Crystallinity.
15. Process control.

Based on our ideas along with some in Refs. [5b,16–18], the characteristics of ideal powders or quasi-ideal powders for sintering or reaction areas follows:

1. Fine powders (within 1 μm).
2. Spherical geometry or geometry close to spheres.
3. No particle size distribution (one particle size for requirements of a user) or a narrow particle size distribution.
4. No or soft agglomeration.
5. Good microstructure of grains (a grain without pores and grains without glassy grain boundaries (i.e. a grain boundary represents the edges of grains)).
7. No or low stress in the grain.
8. Crystallinity of grains suitable for the requirements of a user.
9. Chemical compositions suitable for the requirements of a user or chemical compositions close to the value suitable for the requirements.
10. Uniformity of particle properties: for example, the uniformity of the thermal history, microstructure, and chemical composition of particles.
11. No or low defects for both microstructure and the chemical composition.
12. Good reactivity of grains when powders are used as starting materials.
13. Good sinterability of grains when powders are sintered.
14. Reproducibility of mechanical, physical, and chemical data observed in particles.

3. METHODS TO PRODUCE FINE CERAMIC POWDERS [1,5,11,12,15,19–23]

There are two ways to produce fine ceramic powders: the so-called breaking-down [1,11,15,23,24] and building-up processes [1,5,12,15,19–22]. Breaking-down processes are listed in Table 1 [24]. As for building-up processes, they are divided into two groups, namely, (1) physical methods, such as solvent exclusion, and (2) chemical methods, such as thermal decomposition. These building-up processes are listed in Table 2 [25]. Generally, the starting materials of advanced ceramics are produced by using these building-up processes because it is difficult to make powders <1 μm by using breaking-down processes such as crushing. Therefore, we mainly describe the building-up processes, namely, the processing of artificial powders in the following sections.

Artificial fine ceramic powders are also classified according to the phases from which powders are produced: powders from the gas phase (Section 3.1), the liquid phase (Section 3.2), and the solid phase (Section 3.3). Powders from the gas and the liquid phase are more important than powders from the solid phase because high-purity and homogenous particles with specific sizes are generally able to be produced.

3.1. Ceramic Fine Powders from the Gas (Vapor) Phase [15,21,22,26,27]

Direct synthesis routes in the gas phase have important advantages for the preparation of ceramic powders in order to fabricate advanced ceramic parts. As for using the gas (vapor) phase, there are two ways, namely, physical vapor deposition (PVD) [15,22,26] and chemical vapor deposition (CVD) [15,22,26,27]. The former contains the processes of heating, vaporization, cooling, and condensation of materials in order to produce powders. Meanwhile, the latter contains chemical reactions in the gas state to produce powders.

PVD processes are simple, and involve just heating and cooling of materials. By heating, materials are vaporized. Then, the vapor is condensed by cooling in order to form powders. Heating methods for PVD include many methods, such as (a) the laser beam heating method, (b) the electron beam heating method, methods to use furnaces such as (c) electric resistance furnaces and (d) high-frequency inductance furnaces, and (e) the plasma jet heating method.

As for CVD, there are two ways to produce powders [22]: One is the thermal decomposition of a chemical as...

<table>
<thead>
<tr>
<th>TABLE 1 Methods for Producing Fine Ceramic Powders by Milling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Millling (fine grinding mill)</strong></td>
</tr>
<tr>
<td>Roller mill</td>
</tr>
<tr>
<td>Roller mill</td>
</tr>
<tr>
<td>Frictional grinding mill</td>
</tr>
<tr>
<td>Autogenous mill</td>
</tr>
<tr>
<td>Ball mill</td>
</tr>
<tr>
<td>Attrition</td>
</tr>
<tr>
<td>Centrifugal fluidizing mill (CF mill)</td>
</tr>
<tr>
<td>Ultra fine mill (UF mill)</td>
</tr>
<tr>
<td>Agitating ball mill</td>
</tr>
<tr>
<td>Mixing vessel type</td>
</tr>
<tr>
<td>Tube type</td>
</tr>
<tr>
<td>Annular type</td>
</tr>
<tr>
<td>Annular agitating type</td>
</tr>
<tr>
<td>Jet mill</td>
</tr>
<tr>
<td>Wet grinding mill</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Stone mill</td>
</tr>
</tbody>
</table>

shown in Eqn (1) and the second is to react more than one kind of chemical with another as described in Eqn (2).

\[ A(g) \rightarrow B(s) + C(g) \]  
(e.g. the decomposition of CH\(_3\)SiCl\(_3\) (g) to form SiC (s) [22])

\[ A(g) + B(g) \rightarrow C(s) + D(g) \]  
(e.g. 3SiH\(_4\) (g) + 4NH\(_3\) (g) \rightarrow Si\(_3\)N\(_4\) (s) + 12H\(_2\) (g) [15])

By using the second method, not only oxide powders but also nonoxide powders, such as boride, carbide, nitride powders can be produced. As for heating methods for CVD [22], there are also various ways: (a) electric furnace heating, (b) combustion (chemical flame) heating, (c) plasma heating, and (d) laser heating.

Characteristics of powders produced by CVD [22] are as follows:

1. Powders are highly pure.
2. Powders are fine, and hence, it is not necessary to crush the powders further.
3. Particles are not aggregated.
4. The particle size distribution becomes narrow when the reaction conditions are controlled. In addition, it is easy to obtain particles of the order of angstroms to several micrometers.
5. Oxide, boride, carbide, and nitride can be produced because it is easy to control the environment of the reaction.

### 3.2. Ceramic Fine Powders from the Liquid Phase [15,19,21,22]

In the case of powders produced from the liquid phase, powders in some cases need the removal of precipitates, namely, processes such as filtration, washing, and drying. Moreover, the particles after drying are ordinarily agglomerates.

In the following, we explain the methods to prepare ceramic powders from the liquid phase, such as physical methods like solvent exclusion and heat treatment, and chemical methods such as precipitation, hydrolysis, the sol–gel method, the microcapsule method, the reactive electrode submerged arc method (RESA) [28], and the spark discharge in solution method [29,30].

There are various methods related to solvent exclusion [15,19,21], such as the evaporation to dryness method, the freeze drying method (the cryochemical method), spray drying methods (spray drying methods: e.g. the nozzle type and the plasma flame spray method), and the emulsion drying method (the hot kerosene (kerosine) method or the hot (petroleum or oil) drying method).

A simple method to produce powders is to heat the solution for the removal of solvent or to evaporate the solvent. In addition, the method to evaporate the solution for drying (the evaporation to dryness method) is easy to make powders mix with the binder, etc., in laboratories. For example, a salt of a nitrate is placed in a quartz beaker or on a ceramic evaporating disk or a platinum (Pt)-evaporating disk; the salt is heated in order to evaporate and decompose, and finally, powders are obtained. An advantage of powders produced by this method [21] is as follows:

1. Powders are obtained directly from the solution.
However, the disadvantages are as follows:

1. It takes time to evaporate water.
2. It is difficult to get a large amount of powder, such as oxide powders.

The freeze-drying method [15], one of the solvent exclusion methods, is done to freeze mixed salt solution rapidly in order to separate salt and water, namely, ice, and to obtain anhydrous salts after ice sublimes. Salts are spherical, well mixed, and chemically uniform. After anhydrous salts are calcined, powders such as oxides are obtained. Because these powders have very fine pores, they easily absorb the binder and mix with the binder.

Characteristics of powders produced by the freeze-drying method are as follows:

1. During drying, the coagulation of particles does not occur.
2. During drying, the shrinkage of particles does not usually occur.
3. Powders are highly pure. One may use these powders as highly pure starting powders for sintering, etc.
4. After the powders are dried, it is not required to crush the powders because the powders are fine.
5. It is easy to absorb the binder and to mix with the binder because the powders are porous.
6. The reactivity of particles is high because the surface reactivity is high.
7. It is easy to calcine powders because surface reactivity is high and powders are porous.

The spray drying method [15] is used to spray solution mixed with salt as droplets into a room with high-temperature gas flow and to obtain fine dried powders after drying. Otherwise, droplets freely fall in a high-temperature room and become dry. When a flame is used instead of the room, the method is called the flame spray drying method or the flame spray method.

Spray drying methods are of two types: nozzle and spinning disk types. The former is a method to spray salt solution into a tank by using compressed air and to introduce hot air into the tank in order to obtain dry powders. Particles produced by using this type are 250–350 μm in diameter. The latter is a method to spray or drop salt solution onto the central part of a rotary disk. In order to vaporize the liquid from the salt solution, hot air is introduced onto the rotary disk. Then, powders are decomposed to oxides at 800–1000 °C to flow hot air. Then, the particles produced are 50–150 μm in diameter. In addition, the distribution of particle sizes is narrow.

The plasma flame spray method [19] directly produces powders from liquid. This is a method of spraying salt solution into a plasma flame and obtaining solid powders after the scattering and vaporization of water, and then calcining. Powders produced by this method are spherical and have good surface reactivity.

The emulsion drying method [19] is also used to remove solvent from the solution. It is a method in which one drops mixed salt solution with sulfate or acetate into 170 °C of hot kerosene or oil under stirring and calcines the mixed crystal powders after filtration. This method is easy and simple.

The heat treatment of powder preparation involves three processes: (a) volatilization and cooling, (b) crystallization, and (c) melting and rapid quenching [19,21,22].

(a) The volatilization and cooling method is used to get fine powders after heating the materials in order to vaporize and then cooling down the vapor.

(b) By using the crystallization method (b), fine crystallized powders are obtained in some cases. This method allows amorphous powders containing water to become crystallized under heating with or without high pressure.

(c) The melting method involves the melting and quenching of a mixture in a Pt crucible, and crushing the quenched mixture to powders. The crushing process may be repeated three times in order to obtain homogeneous powders as good starting materials for the preparation of ceramics, etc. Faults of powders produced by using this method are chemical compositions with impurities due to the vaporization of alkali during melting when substances containing alkali are used. In addition, the impurities from tools for crushing would be added. For example, iron would become an impurity when iron mortars are used.

In order to obtain ceramic powders, materials may be heated to decompose them. Thermal decomposition is also used after spray drying. These methods are the spinning disk type of spray drying and the flame spray method [15]. The spinning disk type is used to vaporize the liquid from a salt solution and then the powders produced are decomposed. In the flame spray method, a salt solution is sprayed into a high-temperature environment; spherical powders are obtained simultaneously after thermal decomposition and synthetic reaction. The particle size distribution produced by this method is 50–150 μm, which is narrow. It is said that powders produced by this method are suitable for hot die molding [19].

Precipitation [15,19,21] is also used to obtain ceramic powders. There are various precipitation methods such as the direct precipitation method, the homogenous precipitation method, the coprecipitation method, and the compound precipitation method. The direct precipitation method is done by neutralization and precipitation. In addition, the starting powders for sintering may be produced after the calcination of precipitations separated from the solution. Because precipitates are required to be separated from the liquid by filtration, large-size
NH₄OH needs to be gradually formed in order to avoid local inhomogeneity of chemical composition. One is able to get high-purity powders that are easily filtrated when the rate of the decomposition of urea is controlled.

Concerning the coprecipitation method [15,19,22], a precipitate is obtained by the addition of a precipitant to a mixed-salt solution. Different ions precipitate as ceramic precursors. This method is carried out using fractional precipitations, and hence, the products are mixtures of salts. These mixed-salt solutions are made by using solutions of carbonate, hydroxide, oxalate, sulfate, etc. As a characteristic, powders produced by this method have good reactivity. The coprecipitation method is done using mixed precipitation while the compound precipitation method is done using the precipitation of a stoichiometric compound.

The hydrolysis method [19,21] is a method to decompose starting materials by the addition of water and produce hydroxides or oxides containing water. Then, hydrolyzed powders may be calcined. Sizes of fine oxide particles calcined after the decomposition of powders are 10–100 nm [19]. Precipitates are homogenous. Hydrolyzes of inorganic salts are done using ammonium salts, chlorides, nitrates, or sulfates as starting materials. Sizes of particles produced by the decomposition of inorganic salts are mostly of the order of 20 nm.

By hydrolyzes of metal alkoxides [1,9,22], amorphous metals containing water are obtained as agglomerate particles. Then, they are changed into fine particles or powders by heating. These hydrolyzes of alkoxides are one of the important methods to obtain fine and high-purity powders for advanced ceramics. Characteristics of powders are dependent on the mole ratio of alkoxide to water, reaction temperature, aging time, and so on. In some cases, organometallic compounds produced by the hydrolysis are followed by reduction due to the existence of carbon formed during the hydrolysis.

As for the sol–gel method or the colloidal chemical method [15,19,21], gels appear when a solution of metal ions, such as aluminum (Al), barium (Ba), calcium (Ca), lead (Pb), magnesium (Mg), potassium (K), and sodium (Na), is added to a colloidal sol of silicon dioxide (SiO₂), titanium dioxide (TiO₂), or zirconium dioxide (ZrO₂) with stirring and then nitric acid (HNO₃) is added to the solution in order to make its pH 5.5–6. A gel that consists of an oxide and metal ions is dehydrated at about 70 °C, then the dehydrated gel is calcined at about 400–700 °C. Consequently, powders are obtained. This method is suitable for preparing silicate, titanate, and zirconate. The powders prepared by using this method are very densely packed because these powders are fine. In addition, these powders can be melted at a temperature that is 100–200 °C lower than that for conventional powders and be sintered for a shorter time than for conventional powders. Sintered bodies become highly dense and have a uniform structure.

The microcapsule method [19] requires one to drop the suspension of an inorganic material into an immiscible medium and obtain powders after the reaction that occurs between drying and calcination. Oxide powders produced by this method are almost spherical, and the particle size distribution is narrow.

The hydrothermal synthesis method is explained in detail in Section 4.

The RESA method [28] was invented by R. Roy et al. at The Pennsylvania State University. The RESA method utilizes metal electrodes, such as Al, silicon (Si), and titanium (Ti), submerged in a dielectric fluid, such as water, organic fluid, mineral oil, and silicon oil for producing fine ceramic powders. A spark caused by electrodes provides an extremely high temperature and pressure for a short duration. Then, the high temperature and pressure result in the vapor of electrodes and surrounding dielectric fluid, and the reaction occurs within “bubbles” formed by the vapor. The reaction between various vapor species or between solid and vapor occurs, and the reacted products in the form of fine smoke are quenched in the surrounding dielectric liquid yielding a colloidal solution of spherically shaped powders. After the colloids are dried, oxide, carbide, and nitride powders are obtained depending on the kind of liquid used, such as water, as described before.

The spark discharge in the solution method (the Iwatani–Ishibashi method) [29,30] is similar to the RESA method, which was mentioned above. Because an experimental setup of the spark discharge in solution method involves a pool of water and a series of sparks at many points of metal pellets (e.g. Al, Mg), the setup is different from that of the RESA method. In any case, products obtained by the spark discharge in solution method are hydroxides and hydrated phases. At the point of spark discharge, surfaces of metal are peeled off and decompose in water. Metals, such as Al, react with water to form hydroxides, such as aluminum hydroxide (Al(OH)₃). By using this method, Al or Mg
hydroxides can be produced. In addition, after the heating of hydroxides, oxides can be produced.

3.3. Ceramic Fine Powders from the Solid Phase [15]

Concerning ceramic fine powders from solids, it is difficult to make high-purity fine powders by crushing because typical crushing tools add impurity to powders and limit the particle size to a minimum (e.g. ~1 μm). Some conventional methods are the thermal decomposition of solids [15] and solid-state reactions [15]. We explain these two methods in the following. In addition, we also describe powders syntheses under special conditions, such as high pressure [15,31–34].

Thermal decomposition starts from surfaces even if single crystals thermally decompose. Decomposition products may be fine powders. It is important to control thermal decomposition by the amount of heat and the removal of gas as byproducts that occur during the heating of materials. Thus, it would be better to check gaseous byproducts using thermal analyzes, for example, thermogravimetry or thermogravimetric analysis.

Fine powders are also produced by solid-state reactions. Powders are heated to the reaction temperature and the reaction occurs. There are three types of reactions as shown [15]:

\[
A(s) + B(s) \rightarrow C(s) \\
\text{(e.g. } SiO_2(s) + 3C(s) \rightarrow SiC(s) + 2CO(g)) \\
A(s) + B(g) \rightarrow C(s) \\
\text{(e.g. } 3Si(s) + 2N_2(g) \rightarrow Si_3N_4(s) [35], \\
2Ti(s) + 2NH_3(g) \rightarrow 2TiN(s) + 3H_2(g). \\
3SiO_2(s) + 6C(s) + 2N_2(g) \rightarrow Si_3N_4(s) + 6CO [35]) \\
A(s) + B(l) \rightarrow C(s) \\
\text{(e.g. } 7C(s) + 2B_2O_3(l) \rightarrow B_2C(s) + 6CO(g) [35])
\]

Under certain conditions, solid-state reactions involve processes, namely, calcination and sintering. Calcination is a process in which the starting materials react (to some extent) at temperatures lower than that of sintering, and in sintering, the remaining unreacted materials react completely, if these exist, and form grains in order to produce highly dense products. Because sintering is a process of forming sintered bodies and is not generally suitable for obtaining powders, calcination and pulverization are carried out to produce powders. Then, impurities would be introduced by the crushing, and this is a disadvantage for this method. In addition, it would be better to fire and crush calcined powders more than once in order to obtain homogenous powders.

Before calcining the starting materials or beginning solid-state reactions, one needs to consider the following points:

1. Aims of producing powders because calcination conditions depend on them.
2. Storing conditions of the starting materials because starting materials may change depending on materials and atmosphere (e.g. materials might absorb water).
3. Impurities of the starting materials, for example, water contents.

Solid-state reactions are dependent on the following factors:

1. Particle sizes; fine powders are better.
2. Particle size distribution; narrow distribution is better.
3. Accurate weighing of the starting materials: it is better to heat starting materials to remove water if the starting materials contain water.
4. The mixing of powders: homogenous mixing is better.
5. Packing conditions of powders: highly dense packing is better.
6. Heating rates; for example, rapid heating would produce fine and hollow particles.
7. Reaction temperature; for example, if calcination temperature is not within a suitable temperature range, some starting materials would remain as they are.

Powders prepared by using solid-state reactions with more than one solid starting material tend to become inhomogeneous because of poor contact among different kinds of solid. However, there are solid-state reactions using gases (e.g. oxygen (O2) and nitrogen (N2)), such as the oxidation and nitriding of metal powders in order to form powders or particles, such as oxides and nitrides, respectively. For example, high-purity silicon nitride (Si3N4) powders are produced to heat high-purity silicon (Si) powders under a N2 atmosphere (see the first example following Eqn (5)) [35].

At present, there are syntheses performed under special conditions such as high pressure, high temperature, or both high pressure and temperature [31–34]. Apparatus, such as diamond cells, and methods, such as laser heating syntheses and shock wave syntheses, are used.

Apparatus used for syntheses under high pressure are diamond cells, anvil cells, etc. The maximum pressure is as high as 300 GPa. Diamond is synthesized using the piston–cylinder type under a pressure of up to 7 GPa, the Bell type of up to 13 GPa, and the tetrahedron anvil cell type of up to 8 GPa. Because the space for samples is very small, the weight of samples is of the order of several micrograms.

Laser heating in some apparatuses is used at temperatures of up to 3000 °C. In 1959, the General Electric Company reported the synthesis of diamond at about 5.5 GPa and 1300 °C [32,33].
Shock wave syntheses are carried out to explode explosive powders under a high pressure from shock waves. Materials such as diamond and cubic boron nitride are synthesized. Relating to high oxygen pressure apparatus, experiments under 4 kb and up to 700 °C were reported [34]. In the case of experiments with oxygen gas, special care is necessary to not explode the apparatus.

4. HYDROTHERMAL SYNTHESES AND CHARACTERISTICS OF HYDROTHERMAL POWDERS [4,5,11,25,36–45]

We explain hydrothermal syntheses and powders produced under hydrothermal conditions in detail because characteristics of hydrothermal powders are close to those of ideal powders as listed in Section 2.2.

The term “hydrothermal” as in “hydrothermal synthesis” [25,36,37a,37b] came from earth science and implies a region of water pressure (and water vapor pressure) and high temperature >100 °C. Cement researchers, however, prefer a temperature > 60 °C rather than one >100 °C. Hydrothermal reaction occurs in water (liquid H2O). In addition, there is a possibility that water vapor (gas) reacts with substances by hydrothermal reaction.

For typical hydrothermal research [4,25,36–46], one needs a high-temperature and high-pressure apparatus called an autoclave or a bomb. A great deal of experimental work has been carried out using Morey bombs [25,36,37a] and Tuttle-Roy test tube bombs [20,25,37a] (made by Tem-Pres Division, LECO Corp. [47]) shown in Figure 1 [37a] and Figure 2 [37a]. There are many articles related to autoclaves and bombs, and these vessels are extremely important in hydrothermal experiments. Thus, we explain these vessels in Section 4.1.

There are two types of systems that use the autoclave or the bomb: the batch system and the continuous system. In general, people in the industry prefer the continuous system while those in academic research prefer the batch system. These systems are explained in Section 4.2. As for heating, there are also two types: an external system and an internal system. These systems are explained in Section 4.3. Relating to pressuring, namely, the increase of the water vapor pressure in the autoclave or the bomb, there are also two methods: one involves heating and the other uses a pump or an intensifier.

Water becomes vapor at 100 °C under atmospheric pressure, but it still remains liquid at a high temperature under high pressure unless the conditions exceed the critical point. Therefore, solubility in water at high temperatures under high pressures is different from that in ordinary water. The solubility of almost all materials increases as temperature increases. This is one of the reasons for using the hydrothermal synthesis method.

In addition, the rate of reaction becomes high at high temperature.

The action of water depends on its surrounding conditions as described below:

1. Water functions as a solvent in a reaction.
2. Water functions as a starting material, namely, a substance that reacts with other starting materials.
3. Water accelerates the rate of a reaction as a catalyst or a mineralizer (a reaction promoter) to rearrange or recrystallize the atoms and ions of products.
4. Water functions as the medium and regulates temperature and pressure.

In the case of water vapor or steam, steam acts a medium to regulate temperature and pressure. In addition, there are cases in which steam functions as a starting material, or a catalyst, or a mineralizer. In other words, in some cases, hydrothermal synthesis involves H2O as a component of reacted products (i.e. solid phases) in the synthesis or a catalyst at elevated temperatures (>100 °C (or 60 °C)) and pressure (more than a few atmospheres) [44c].

The hydrothermal synthesis method is used for the synthesis of materials and the crystal growth of materials under high pressure and temperature of water. Section 4.4 deals with starting materials of hydrothermal reactions. As
shown in Table 3 [20,21], there are many types of reactions, and these reactions are explained in Section 5.7 of the first edition of this handbook. Moreover, example reactions are briefly described in Section 4.5. Characteristics of hydrothermal processing are shown in Table 4 [5,19,20]. In Section 4.6, zirconia-based powders are explained as example products of the hydrothermal reaction, hydrothermal homogenous precipitation. In addition, Section 4.7 describes the characteristics of hydrothermal powders and comparisons between hydrothermal and ideal powders.

4.1. Autoclaves or Bombs

According to Laudise and Nielsen [48], the most important apparatus for hydrothermal crystal growth experiments are autoclaves and closures. The ideal hydrothermal autoclave should have the following characteristics: (a) inertness to acids, bases, and oxidizing agents; (b) ease of assembly and disassembly; (c) ruggedness in order to require no treatments between runs; (d) enough length to produce large internal temperature differences within the autoclave and a diameter that is large enough to grow crystals of a sufficient size; (e) no leaks; and (f) unlimited pressure and temperature in the autoclave.

Obviously, no such vessel exists even though vessels to fulfill one or some of the above criteria are available. Table 5 [37a,44b] lists the limits of pressure and temperature of autoclaves, and Table 6 [25,36,37a,44b,46,49] lists sealing methods for autoclaves.

One needs an autoclave or a bomb in order to do typical hydrothermal research [4,25,36-46]. There are Morey bombs and Tuttle bombs or Tuttle-Roy bombs, and their names have originated from the names of researchers of hydrothermal work and inventors of autoclaves, Morey, Tuttle, and Roy. The Tuttle-Roy bomb is also called the Tuttle-Roy test tube reactor or the Tuttle-Roy test tube bomb.

Early studies related to hydrothermal systems at high temperature and pressure were carried out by Morey and coresearchers at the Geophysical Laboratory of the Carnegie Institution of Washington, Washington, DC, since 1913 [50].
The Morey type autoclave is quite easily assembled, and the gasket plate requires minimum mechanical precision. The first version of the Morey bomb was sealed off with a fixed volume of water, and hence, the pressure was fixed at a specific temperature. Therefore, the most important modification was made to introduce a drilled plunger in order to pressure the vessel outside it. Moreover, there were two additional problems inherent in design [51]. First, the thrust of the closure nut (“plunger” in Figure 1) on the soft metal sealing gasket produced leaks at a high pressure or with rapid temperature change. In addition, the first problem made the operation of the vessel time consuming because of not only leaks but also difficulties in removing the closure at the end of a run. The second problem was pressure measurement because the entire Morey bomb and closure lay within the furnace. At that time, the pressure was calculated from the pressure—volume—temperature (P—V—T) relationship assuming that

<table>
<thead>
<tr>
<th>TABLE 3 Hydrothermal Reactions [20,21]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrothermal synthesis</td>
</tr>
<tr>
<td>1. Hydrothermal crystal growth</td>
</tr>
<tr>
<td>2. Hydrothermal treatment</td>
</tr>
<tr>
<td>3. Hydrothermal alteration</td>
</tr>
<tr>
<td>4. Hydrothermal dehydration</td>
</tr>
<tr>
<td>5. Hydrothermal extraction</td>
</tr>
<tr>
<td>6. Hydrothermal sintering</td>
</tr>
<tr>
<td>7. Hydrothermal reaction sintering</td>
</tr>
<tr>
<td>8. Corrosion reaction</td>
</tr>
<tr>
<td>9. Hydrothermal oxidation</td>
</tr>
<tr>
<td>10. Hydrothermal precipitation</td>
</tr>
<tr>
<td>11. Hydrothermal decomposition</td>
</tr>
<tr>
<td>12. Hydrothermal electrochemical reaction</td>
</tr>
<tr>
<td>13. Hydrothermal mechanochemical reaction</td>
</tr>
<tr>
<td>14. Hydrothermal reaction under microwave conditions</td>
</tr>
<tr>
<td>15. Hydrothermal reaction under ultrasonic conditions</td>
</tr>
<tr>
<td>16. RESA; spark discharge in solution</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 5 Limits of Pressure and Temperature of Autoclaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure (atm)</td>
</tr>
<tr>
<td>Pyrex (5 mm as inner diameter, 9 mm as outer diameter)</td>
</tr>
<tr>
<td>Pyrex (5 mm as inner diameter, 9 mm as outer diameter)</td>
</tr>
<tr>
<td>Morey type (flat plate seal)</td>
</tr>
<tr>
<td>Welded Walker–Buehler</td>
</tr>
<tr>
<td>Delta ring</td>
</tr>
<tr>
<td>Bridgeman seal</td>
</tr>
<tr>
<td>Modified Bridgeman seal</td>
</tr>
<tr>
<td>Cold test tube</td>
</tr>
<tr>
<td>Stellite 25</td>
</tr>
<tr>
<td>Rene 41</td>
</tr>
</tbody>
</table>

(After Laudise and Nielsen [37a] and Somiya [44b])

<table>
<thead>
<tr>
<th>TABLE 4 Characteristics of Hydrothermal Processing Based on Characteristics of Powders [5,19,20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Powders formed directly from solution</td>
</tr>
<tr>
<td>2. Crystalline and amorphous powders formed depending on hydrothermal temperature</td>
</tr>
<tr>
<td>3. Particle size controlled by hydrothermal temperature</td>
</tr>
<tr>
<td>4. Particle shape controlled by starting materials</td>
</tr>
<tr>
<td>5. Easy to control chemical composition</td>
</tr>
<tr>
<td>6. In many cases, powders formed without water (i.e. processes to remove water are not necessary)</td>
</tr>
<tr>
<td>7. In many cases, powders formed without agglomeration because these powders are fine and anhydrous (i.e. the milling of powders is not necessary)</td>
</tr>
<tr>
<td>8. Highly reactive powders formed (i.e. calcination or preheating is not necessary before sintering)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 6 Sealing Methods [44b]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sealing methods</td>
</tr>
<tr>
<td>Flat plate closure (Morey type) [25,36,37a]</td>
</tr>
<tr>
<td>Cold-cone seat closure (Tuttle type) [46,37a]</td>
</tr>
<tr>
<td>Delta and “D” ring closure [37a]</td>
</tr>
<tr>
<td>Modified Bridgeman closure [37a]</td>
</tr>
<tr>
<td>Gray Loc closure [49]</td>
</tr>
</tbody>
</table>
nothing was lost during an experimental run after sealing a known volume of water in the vessel at the beginning of the run. In other words, the pressure was calculated based on a known ratio of the water to the vessel.

Later, these problems were solved in a modified design, and the above types of bombs have been used up to 3 kb and up to about 600 °C. Vessels made out of Inconel may be used under a pressure as high as 4 kb and at 600 °C. Morey vessels for academic experiments are used in order to obtain a lot of samples because of a large volume is a requirement for the vessel.

Later, a vessel different from the Morey bomb was designed by Tuttle, Roy, and other researchers. This type is called a Tuttle bomb or a Tuttle–Roy bomb. The Tuttle bomb is also called a cold seal pressure vessel. This type of vessel has the basic advantages of having no threaded parts that are heated and having sealing made by pressure on the cone-in-cone seal. The principal advantages of this type of apparatus are its simple constitution and a small size of the pressure vessel, which permits more rapid heating and quenching than externally heating vessels do (see Heating Methods in Section 4.3). A Morey vessel and the closure are within the furnace, while the closure of a Tuttle bomb vessel or a cold seal pressure vessel is placed outside the furnace. This type of vessel is simply a metal cylinder with an axial hole drilled from one end to within about 1/2 of the other end. The closed end is placed upward in the furnace with the pressure connection below and outside the furnace.

Roy et al. modified the Tuttle vessel to suspend a vessel in the furnace. They then named it a “test tube” pressure vessel for this unit. The first type required that a capsule with the starting powders be supported by a rod within the vessel, but an improved type does not require this capsule. The Tuttle–Roy bomb may be made cheaply out of ordinary stainless steel and bombs made out of Stellite 25 are suitable for experiments at 1 kb and 950 °C.

When the closed end of a Tuttle–Roy bomb is lowered, quenching becomes very efficient when one immerses the closed end in a bath of water. The use of a close-fitting rod in order to decrease the volume of the vessel enhances quick safety and greatly diminishes the effect of convection. These types of vessels have been used at a pressure as high as 6 kb to 7 kb and at 700 °C.

At present, one is able to get many kinds of autoclaves to cover different pressure—temperature ranges and volumes. In the USA, there are three companies: Tem-Pres, a division of LECO Corp. [47], Autoclave Engineers, a division of Snap-tite Inc. [52], and Parr Instrument Co. [53]. Each autoclave has limitations of pressure and temperature. These limitations depend on autoclave material, sealing methods, operating temperature, and time duration as shown in Figure 3 [37a] and Figure 4 [49], and Tables 5 and 6. In Japan, there are regulations for high pressure gas operations. In addition, safety is one of the important points for the operation of autoclaves.

4.2. A Batch System and a Continuous System [41]

Processes entail two techniques: One is a batch system, and the other is a continuous system. In the ordinary case, batch systems are common in experimental rooms. Some researchers, however, prefer continuous systems. Sakai Chemical Industry Co. Ltd. has produced barium titanate (BaTiO3) using a hydrothermal continuous system [54].
4.3. Heating Methods

There are two kinds of heating methods, the internal and the external heating method. The internal heating system contains heaters in a reactor, while the external heating system has an external furnace outside a reactor. The external heating system is simpler than the internal heating system, and hence, the external heating system is very commonly used around the world.

Heating methods are selected depending on operating pressure and temperature, the length of operating time, and the size of the reactor. The internal heating system is preferably used under conditions such as high pressure and temperature, long operating time, and large-volume autoclaves. A requirement of reactors for industrial products is large volume because of large-sized products, many products, or both. Thus the internal heating system is mainly used. Concerning academic research, the internal heating method may be used in order to use autoclaves under conditions up to 10 kb and 1500 °C.

4.4. Starting Materials

For hydrothermal experiments, the starting materials are very important. Requirements for the starting materials are as follows:

1. The use of high-purity starting materials.
2. The use of fine starting materials, namely, fine powders.
3. Accurate ratio of the starting materials for the chemical composition of a product.

There are six types of starting materials that are commonly used for research and industrial production: (a) glasses; (b) gels; (c) dry mixtures of oxides; (d) chemical salts, such as carbonates, hydroxides, nitrates, and sulfates; (e) natural minerals and rocks; and (f) metal (with regard to metal, there are cases in which small pieces and wires may be used instead of powders). Some starting materials, such as carbonates, easily absorb water, and hence, it would be better to use starting materials without water in order to correctly weigh the required substances. For example, one may heat starting materials in order to remove water before weighing them.

4.5. Hydrothermal Reactions

Table 3 lists the various hydrothermal reactions and some example reactions are introduced here. Among them, Reactions 1, 4, 6, 7 as listed in Table 3 are used to produce nonpowders (i.e., bulks) as final products. We also mention these reactions because some are industrially and scientifically important, and it would be valuable to introduce them. Important reactions under hydrothermal conditions are Reactions 1, 7, 10. From the industrial viewpoint, the first two reactions are important.

An example of Reaction 1 is the crystal growth of artificial quartz (SiO₂) [49,55]. Hydrothermal crystal growth is not a process for preparing powders. Reaction 2 is used to remove the impurities from objective products under hydrothermal conditions. An example is the removal of glassy materials from fluorphlogopite rock [55].

An example for Reaction 3 is the exchange reaction in rock, fluorine ion ↔ hydroxide ion (F⁻ ↔ OH⁻). Kaolinite is produced from feldspar [55].

An example for Reaction 4 is to prepare potassium hexatitanate (K₂Ti₆O₁₃; fiber) from hydrous titanium dioxide (TiO₂·nH₂O), potassium hydroxide (KOH), and water (H₂O) along with either magnesium (Mg metal) or zinc (Zn metal) as a dehydration reagent [56]. Example products were not powders but fibers as described above.

Examples of Reaction 5 are to obtain aluminum oxide or alumina (Al₂O₃) from bauxite (ore containing Al₂O₃) in...
alkali solution [57], and magnetite (Fe₃₋ₓO₄) from ilmenite (FeTiO₃) in KOH solution [58,59].

Reaction 6 [60] may be used to remove water from hydrated materials during hot pressing by using hydrothermal apparatus and retaining the water in the apparatus. In other words, hydrothermal sintering may be used for hot pressing of hydrates in order to obtain anhydrous sintered products. Examples of such products are porous silicon dioxide or silica (SiO₂) and dental materials, such as quartz (SiO₂). Hydrothermal sintering is not a process for preparing powders.

Final products of Reaction 7 are not powders, either. Reaction 7 is used to change metal powders or thin pieces of metal into oxide powders and sinter the oxide powders [61]. In other words, powders are intermediate products and sintered products are not powders. This sintering under certain conditions allows one to produce smaller grain sizes of sintered bodies than that of the starting metal [61]. Grain size increases as the sintering temperature increases.

From the scientific viewpoint, this is one of the most important hydrothermal reactions. R. L. Coble who was a professor of the MIT, an authority on the sintering of ceramics, and an inventor of transparent alumina “Luca-lox,” had mentioned that hydrothermal reaction sintering became a scientific research issue because particle size reduction from starting powders (chromium (Cr metal)) to intermediate powders (chromium oxide (Cr₂O₃)) was observed [62].

Examples for Reaction 11 are shown below:

\[ 3\text{FeTiO}_3 + \text{H}_2\text{O} \rightarrow \text{Fe}_3\text{O}_4 + 3\text{TiO}_2 + \text{H}_2 \] [59]

\[ \text{ZrSiO}_4 + 2\text{NaOH} \rightarrow \text{Na}_2\text{ZrSiO}_5 + \text{H}_2\text{O} \] [68]

(Sodium hydroxide (NaOH) is used as a catalyst in order to decompose zirconium silicate (zircon); (ZrSiO₄))

\[ \text{ZrSiO}_4 + \text{Ca(OH)}_2 \rightarrow \text{ZrO}_2 + \text{CaO} \cdot \text{SiO}_2 + \text{H}_2\text{O} \] [68]

An example of Reaction 12 is to produce BaTiO₃ by reaction between Ti (metal) and barium nitrate (Ba(NO₃)₂) [69].

An example of Reaction 13 is to produce barium hexafluoride (BaO·6Fe₂O₃) using chemicals, such as barium hydroxide (Ba(OH)₂) and iron (III) chloride (ferric chloride; FeCl₃) [70].

Special characteristics of Reaction 14 [71] are to save time, energy, and cost because of (a) rapid heating to reaction temperature, (b) low reaction temperature and pressure compared with conventional hydrothermal synthesis conditions, (c) the increase of reaction speed by one or two orders of magnitude compared with conventional hydrothermal synthesis, namely, short duration time, (d) formation of a novel phase of a material, and eliminating the metastable phase processes, and (e) environmentally safe and benign processes.

Three examples are explained below [71]. Products under microwave hydrothermal conditions with titanium tetrachloride (TiCl₄) in hydrochloric acid (HCl) at 164 °C for both 0.5 and 2 h were rutile (TiO₂). Meanwhile, products under conventional hydrothermal conditions at 164 °C for both 2 and 24 h were rutile and a small amount of anatase (TiO₂).

Monoclinic ZrO₂ was produced under microwave hydrothermal conditions with ZrOCl₂ solution at 194 °C for 2 h, but ZrO₂ was not produced under conventional hydrothermal conditions at 195 °C for 2 h.

Titanium oxychloride (TiOCl₂) reacts with Ba(OH)₂ to produce BaTiO₃ and a small amount of impurity under microwave hydrothermal conditions at 194 °C for 2 h, but BaTiO₃ is not produced under conventional hydrothermal conditions for 24 h even though the same sources with Ti and Ba were used.

Examples for Reaction 15 are shown below [72]:

\[ \text{Al}_2\text{O}_3 \text{ (manufactured by Rhone-Pouleene)} + \text{H}_2\text{O} \rightarrow \text{Al(OH)}_3 \text{ (bayerite)} \]

In addition, Al₂O₃ reacts with phosphoric acid (H₃PO₄) in water to produce hydrated aluminum phosphate (AlPO₄·xH₂O) (e.g. x = 1.67).

Example products of Reaction 16 [28] are oxides, carbides, and nitrides. For example, part of a Zr metal rod
in water becomes fine ZrO₂ powder in water by the RESA method.

Reaction 16 (the Iwatani–Ishibashi method) is similar to the RESA method as described in Section 3.2. A difference between the two methods is the type of solution used and material produced [29,30].

An example product of Reaction 16 [29,30] is aluminum hydroxide (Al(OH)₃) produced from aluminum (Al metal) and H₂O. Al(OH)₃ may be calcined to produce Al₂O₃. When Al which does not contain sodium (Na) is used as a starting metal to produce Al(OH)₃, Na-free Al(OH)₃ and then Na-free Al₂O₃ are produced.

### 4.6. Zirconia and Yttrium-Doped Zirconia Powders Manufactured by Hydrothermal Homogenous Precipitation

We explain zirconia (ZrO₂) and yttrium-doped zirconia (Y-ZrO₂) as sample commercial powders manufactured by one of the hydrothermal precipitation methods, namely, homogenous precipitation.

There are many articles related to hydrothermal zirconia and zirconia-related materials [73–78]. Chichibu Cement Co. Ltd. (presently Taiheiyo Cement Corp. through Chichibu-Onoda Corp.) and one of the authors (S. S.) produced zirconia (ZP 20 as in Table 7 and Figure 6), and partially stabilized zirconia (ZY 30) and stabilized zirconia (ZY 80) using yttrium by adopting the hydrothermal homogenous precipitation method with financial support from the Research Development Corp. of Japan (presently Japan Science and Technology Agency through the Japan Science and Technology Corp.) [79,80].

Figure 5 [79,80] shows the processes for preparing hydrothermal homogenous precipitation powders of ZrO₂ and Y-ZrO₂. Properties of ZrO₂ and Y-ZrO₂ prepared by using this method are shown in Table 7 [79,80]. Transmission electron microscopy (TEM) pictures of ZrO₂ (ZP 20) and Y-ZrO₂ (ZY 30, ZY 80) are shown in Figure 6 [79,80]. A picture of a sintered body of Y-ZrO₂ (ZY 30) at 1400 °C for 2 h is shown in Figure 7 [79,80]. These have very fine powders and grains.

### 4.7. Characteristics of Hydrothermal Powders [19,20]

Hydrothermal powders are directly formed from solution. These powders may be single crystals, polycrystals, amorphous forms, and metastable crystals depending on hydrothermal preparation temperature as the main condition, and on hydrothermal preparation time. In addition, anhydrous powders may be prepared. Particle size and shape are controlled by hydrothermal temperature and starting materials, respectively. Hydrothermal powders generally do not need milling because these powders are fine and anhydrous, and they are formed without agglomeration. Moreover, these powders are highly reactive during sintering and hence are generally sintered without calcination.

Some characteristics of hydrothermal powders are described below:

1. Fine powders.
2. Narrow particle size distribution.
3. No or weak agglomeration.
4. Good flow: the forming of powders is good.
5. No or few small-size pores in grains.
6. Good structural homogeneity.
7. Good sinterability.
8. Ability to become not only ordinary crystalline forms but also low temperature and metastable ones.
9. Ability to make composites, such as a mixture of organic and inorganic materials.
10. Ability to sinter any kinds of materials, such as high vapor pressure materials.

The ideal powder is assumed to be spherical, but hydrothermal powders are not generally spherical. In other words, the hydrothermal powder is a small fragment, and the shape of the hydrothermal particle varies as shown in Figure 6.

Among various characteristics given above, “1. Fine powders (fine particles <1 μm are easily formed)” (1. in
characteristics of ideal powders or quasi-ideal powders at the end of Section 2.2), “2. Narrow particle size distribution” (no. 3), “3. No or weak agglomeration” (no. 4), “4. Good flow” (no. 6), “6. Good structural homogeneity” (no. 10), structural uniformity among various types of uniformity), “7. Good sinterability” (no. 13) for hydrothermal powders are similar to those characterized as ideal or quasi-ideal powders. In addition, “5. No or few small-size pores in grains” is also a characteristic of ideal or quasi-ideal powders (no. 5 in the characteristics). Thus, some characteristics and properties of hydrothermal powders are similar to those of ideal ceramic powders.

5. SUMMARY

We described opinions of requirements for ceramic powders. Each scholar had his or her own opinion. We described synthetic methods, and characteristics and properties of powders as well as ideal powders. We also explained hydrothermal preparation methods and powders.

ACKNOWLEDGMENTS

One of the authors (S. S.) would like to express his appreciation to the persons and organizations that have helped him in his work.

REFERENCES


Chapter 11.1.2

Sol–Gel Process and Applications

Sumio Sakka
Professor Emeritus of Kyoto University, Japan

Chapter Outline

1. Introduction 883
2. Chemical Reactions in Sol–Gel Solutions 885
3. Formation of Shapes and Microstructures 890
4. Applications 892
5. Recent Topics on Sol–Gel Method 901
6. Concluding Remarks 904
References 904

1. INTRODUCTION

1.1. A Brief History of Sol–Gel Technology

In the early 1930s, Kistler made aerogels [1]. This may be regarded as the first practical application of the sol–gel method. The silica aerogel was prepared by methanol supercritical drying after solvent exchange of water glass-derived gel. In the 1950s and 1960s, Roy [2] used sol–gel-derived homogeneous powder for constructing phase diagrams relevant to ceramics. In the mid-1960s, Schroeder [3] developed sol–gel coating of soda-lime glass in order to modify the optical properties of glass by SiO2 and SiO2–TiO2 films. In the late 1960s, the sol–gel-derived powders were used as source for sintering ceramics. Mazdiyasni [4] prepared dense ferroelectric ceramics at low temperatures, using sol–gel-derived powder. Dislich [5] prepared a pyrex-type borosilicate glass lens by hot-pressing a mixture of sol–gel-derived powder at 630 °C. This fact had a big impact on material researchers working in the field of glass, and people started to do research on the sol–gel preparation of glasses and ceramics for the production of materials. We published a sol–gel paper [6] at 10th International Congress on Glass held in Kyoto in 1974. We found that this is the only paper on the sol–gel technology that was published at this Congress. In order to see the emerging state of this technology, we had to wait until the first Sol–Gel Workshop held in 1981, when the idea on a typical sol–gel method was formed.

In the first half of 1980s, organic–inorganic hybrids or microcomposites were added to the sol–gel products. Schmidt [7] reported on organic–inorganic hybrids in which organic species are bonded with inorganic species. Avnir [8] proposed sol–gel preparation of organic–inorganic microcomposites in which functional organic molecules or particles are dispersed. These additions very much promoted the importance of sol–gel science and technology, attracting chemists and chemical industries to the sol–gel technology. The next progress was brought about by invention of characteristic meso- and macroporous silica and other oxides based on the self-assembly of organic molecules [9] and concurrent gelation-phase separation [10] in the 1980s. Later development of organic–inorganic functional materials has been focused on fabrication of a large variety of organic–inorganic optical, chemical, and biological materials. Also, precise control of meso- and macropores and fabrication of various porous materials besides silica have been extensively developed.

1.2. Outline and Characteristics of Sol–Gel Process

The steps in the sol–gel processing of materials [11–13] and examples of the microstructures of final possible products are shown in Figure 1. The final products are dried gel, heated gel, and glass and ceramics, which are shown by the bold-lined rectangle in the figure.

A typical sol–gel method for fabricating materials starts with a solution consisting of metal compounds, such as metal alkoxides and acetylacetonates as source of
oxides, water as hydrolysis agent, alcohol as solvent, and acid or base as catalyst. Metal compounds undergo hydrolysis and polycondensation at near room temperature, giving rise to sol, in which polymers or fine particles are dispersed. Further reaction connects the particles, solidifying the sol into a wet gel, which still contains water and solvents. Usually, various shapes are formed during the sol to gel transformation. Vaporization of water and solvents produces a dried gel, one of final products. Heating of gels to several hundred degrees and higher temperatures produces dense oxide materials as final products.

These explanations, together with Figure 1, indicate that the sol–gel method is characterized by the low temperature for processing. It is noticed that the sol–gel science and technology is an important basis for nanotechnology, for most of gel products may consist of nanosized particles and pores, or have nanostructures. Since the sol–gel processing starts from the solution, the sol–gel processing is quite versatile; materials of various shapes and microstructures can be prepared. Table 1 shows the shapes of materials made by the sol–gel method.

Bulk bodies can be made by casting the gelling sol into a mold. Fibers can be drawn from the viscous sol, if the sol of appropriate composition is used. Coating films can be made by dip coating or spin coating of the sol. Unsupported films can be made by synthesizing the film at the interface between alkoxide solution and water. Membranes are prepared by pouring the sol onto the porous oxide with coarse pores. Particles with sharp size distribution can be precipitated and grown in the sol.

Examples of microstructures of materials produced by the sol–gel method are shown in Table 2. Three types of microstructures are classified: porous, dense, and organic–inorganic hybrid. Of these three, porous and organic–inorganic hybrid microstructures are obtained at low temperatures around room temperature. As shown in

---

**TABLE 1** Shapes of Materials Produced by Sol–Gel Method

<table>
<thead>
<tr>
<th>Shape</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk bodies</td>
<td>Rod(cylinder), plate, sheet</td>
</tr>
<tr>
<td>Fibers</td>
<td>Continuous fibers</td>
</tr>
<tr>
<td>Coating films</td>
<td>Thin films, thick films</td>
</tr>
<tr>
<td>Unsupported sheets</td>
<td>Sheets</td>
</tr>
<tr>
<td>Membranes</td>
<td>Microfilters</td>
</tr>
<tr>
<td>Particles</td>
<td>Spherical particles</td>
</tr>
</tbody>
</table>

---

**TABLE 2** Types of Microstructures of Materials Produced by Sol–Gel Method

<table>
<thead>
<tr>
<th>Microstructure</th>
<th>Examples of materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porous</td>
<td>Silica aerogels</td>
</tr>
<tr>
<td></td>
<td>Silica columns for liquid chromatograph</td>
</tr>
<tr>
<td>Dense</td>
<td></td>
</tr>
<tr>
<td>Glasses</td>
<td>Silica glass preforms for optical fiber</td>
</tr>
<tr>
<td>Polycrystalline materials</td>
<td>Ferroelectric coating films</td>
</tr>
<tr>
<td></td>
<td>Titanium oxide coating</td>
</tr>
<tr>
<td>Inorganic–organic hybrid</td>
<td>Silica glass with dispersed metal and semiconductor colloids</td>
</tr>
<tr>
<td>Organic–inorganic hybrids</td>
<td></td>
</tr>
<tr>
<td>Distributed organic and inorganic species bonded with each other</td>
<td>Polymethylmethacrylate-silica-titania</td>
</tr>
<tr>
<td>Inorganic or hybrid matrix with dispersed organic molecules and pigments</td>
<td>Silica film with dispersed organic pigments, Silica film with dispersed rhodamin B molecules</td>
</tr>
</tbody>
</table>
Figure 1, dried gels are regarded as one of the final products of sol–gel method. Generally, oxide gels are porous and the porosity creates the useful applications, such as aerogels of silica and other oxides and silica columns for high-performance liquid chromatograph.

When organic compounds, such as alkylalkoxysilanes, methylmethacrylates, polymethylsiloxanes, or organic dyes, are added to the starting solutions, organic–inorganic hybrids are formed. Such hybrids are possible to be fabricated because of the low-temperature nature of the sol–gel processing. Heat treatment of the organic–inorganic hybrids is usually limited to temperatures lower than about 200 °C. It is noted that organic–inorganic hybrids constitute one of the most important material areas in the sol–gel method. There are two kinds of organic–inorganic hybrids. One is a material in which organic and inorganic species bonded with each other are distributed and the other is an inorganic or hybrid matrix with dispersed functional organic molecules and pigments.

Heating of gels to several hundred degrees and higher temperatures causes sintering or collapse of a porous microstructure into a dense microstructure, producing glasses, ceramics, and inorganic–inorganic microcomposites, depending on the composition of materials.

With the advantages and characteristics mentioned above, the sol–gel method has been applied to processing of a great many different materials.

2. CHEMICAL REACTIONS IN SOL–GEL SOLUTIONS

Gelation of the solution is the basis of the sol–gel method. It is known that the microstructure, forming, and other properties of the gel are affected by the gelation condition, even if the composition of the starting solution is the same. Here the reactions occurring in the silicon alkoxide solution will be discussed.

2.1. Hydrolysis and Polycondensation

Gelation of silicon alkoxide solutions takes place as a result of the hydrolysis of the silicon alkoxides Si(OR)₄ and subsequent dehydration–polycondensation, leading to the formation of polymers and particles consisting of siloxane bonds. The reactions can be expressed by the following formula, if it is assumed that the reaction occurs separately:

\[ \text{Si(OR)}_4 + 4\text{H}_2\text{O} \rightarrow \text{Si(OH)}_4 + 4\text{ROH} \quad (1) \]
\[ \text{Si(OH)}_4 \rightarrow \text{SiO}_2 + 2\text{H}_2\text{O} \quad (2) \]

Actually, however, the condensation reaction might start at different stages of the hydrolysis of the silicon alkoxide, depending on the reaction condition. This means that the reaction condition affects the nature of the resultant siloxane polymers or particles as well as their aggregation state and, accordingly, governs the characteristics of the gel.

The hydrolysis in the conventional solution of the Si(OR)₄–ROH–H₂O–catalyst (acid or base) system is generally expressed as follows:

\[ \text{Hydrolysis} : \quad \equiv \text{Si–OR} + \text{H}_2\text{O} \rightarrow \equiv \text{SiOH} + \text{ROH} \quad (3) \]

An alkoxysilane molecule such as Si(OCH₃)₄ and Si(OCH₂CH₃)₄ has four alkoxyl groups, and so the hydrolysis occurs stepwise as seen below. Here, \( k_1, k_2, k_3 \), and \( k_4 \) are the kinetic constants for each step of hydrolysis:

First step

\[ \text{Si(OR)}_4 + \text{H}_2\text{O} \rightarrow \text{Si(OR)}_3(\text{OH}) + \text{ROH} \quad k_1 \quad (4) \]

Second step

\[ \text{Si(OR)}_3(\text{OH}) + \text{H}_2\text{O} \rightarrow \text{Si(OR)}_2(\text{OH})_2 + \text{ROH} \quad k_2 \quad (5) \]

Third step

\[ \text{Si(OR)}_2(\text{OH})_2 + \text{H}_2\text{O} \rightarrow \text{Si(OR)}_2(\text{OH})_2 + \text{ROH} \quad k_3 \quad (6) \]

Fourth step

\[ \text{Si(OR)}(\text{OH})_3 + \text{H}_2\text{O} \rightarrow \text{Si(OH)}_4 + \text{ROH} \quad k_4 \quad (7) \]

When the hydrolysis (4)–(7) proceeds, condensation of the hydrolysis product with an alkoxysilane molecule or another hydrolysis product starts to take place, resulting in the formation of siloxane bonding:

\[ \text{Condensation} : \quad \equiv \text{Si(OH)} + \text{ROSi} \equiv \rightarrow \equiv \text{Si–O–Si} \equiv + \text{ROH} \quad (8) \]
\[ \text{Condensation} : \quad \equiv \text{Si(OH)} + (\text{HO})\text{Si} \equiv \rightarrow \equiv \text{Si–O–Si} \equiv + \text{H}_2\text{O} \quad (9) \]

As reactions (8) and (9) progress, siloxane bondings increase and the hydrolysis products grow to oligomers, polymers, and particles, resulting in gelation. The above reactions indicate that various products are involved. This means that in order to control the microstructure and properties of the gel, information on many factors has to be obtained [14,15]. Effects of some of the factors will be discussed below.

2.1.1. Comparison of Acid Catalyst with Base Catalyst

Acids and bases are employed as catalyst for the reaction of the silicon alkoxides based on the idea that H⁺ and OH⁻...
ions catalyze the hydrolysis [14]. It is known that the reaction pattern of hydrolysis and subsequent polycondensation is different, according to whether the catalyst is acidic or basic [16—19]. It is assumed that in the presence of acid catalyst, such as HCl or HNO3, the hydrolysis of the silicon alkoxide is caused by the electrophilic attack of \( \text{H}_2\text{O}^+ \) ion toward OR, and so the reactivity decreases as the number of OR radicals on the Si decreases with the progress of hydrolysis. Using kinetic constants \( k \)'s corresponding to Formulas (4)–(7), this is expressed as

\[
 k_1 > k_2 > k_3 > k_4 \quad \text{(acid catalyst)} \tag{10}
\]

This shows that at the start of the reaction, that is, in preparation of the alkoxide solution, most of Si(OR)\(_4\) molecules rapidly change to Si(OR)\(_3\)(OH) molecules, which, then, more slowly change to molecules with more OH radicals. This situation is confirmed by [17–19]. It is important to note that the condensation starts among molecules with a smaller number of OH radicals.

On the other hand, the hydrolysis in the presence of base catalyst of NH\(_3\) takes place via the nucleophilic substitution by OH\(^-\) ions. Accordingly, the rate of reaction of an alkoxide molecule increases, due to a decrease in steric hindrance, as the number of OR radicals decreases with the progress of the hydrolysis. This is expressed by Formula (11).

\[
 k_1 < k_2 < k_3 < k_4 \quad \text{(base catalyst)} \tag{11}
\]

Consequently, completely hydrolyzed molecules such as Si(OH)\(_4\) are easily formed by the preferential hydrolysis of the rest of the OR groups belonging to a partially hydrolyzed Si(OR)\(_4–n\)(OH)\(_n\) (\( n = 1–3 \)) molecule, and some silicon alkoxide molecules remain nonhydrolyzed.

The difference in the reaction pattern would be revealed in the structure of reaction products. In the acid-catalyzed solutions, silicon alkoxide molecules with nonhydrolyzed alkoxyl groups polymerize with each other, producing polymers in which the degree of crosslinking is low; however, in the base-catalyzed solutions, Si(OH)\(_4\) molecules participate in the polycondensation reaction, producing highly crosslinked polymers. Keefer [20] and Sakka [21] showed, based on viscosity measurements, that one-dimensionally developed particles which give spinnability to the sol are formed in the acid-catalyzed solutions.

2.1.2. Effect of Type of Acid

The hydrolysis and condensation of silicon alkoxides using acid catalysts other than HCl, HNO3, or H\(_2\)SO\(_4\) may follow the different mechanism. It was shown by Pope [22] that in the HF-containing solution, the hydrolysis takes places based on the nucleophilic attack of F\(^-\) ions, in spite of the fact that the solution is acidic.

2.1.3. Effect of the Kind of Alkoxy Group in Silicon Alkoxide

For the rate of hydrolysis of alkoxysilane Si(OR)\(_4\), the following experimental results are obtained [16].

The rate of hydrolysis decreases as the alkoxy group becomes complicated:

\[
 \text{OC}_3\text{H}_7 \text{Si(OR)\(_4\)} \text{OC}_2\text{H}_5 \text{Si(OR)\(_4\)} \text{OC}_4\text{H}_9 \text{Si(OR)\(_4\)} \text{OC}_6\text{H}_{13} \tag{12}
\]

The rate is smaller for branched alkyl groups than for nonbranching ones:

\[
 n - \text{C}_4\text{H}_{9} > \text{sec} - \text{C}_4\text{H}_9 \tag{13}
\]

This can be explained based on the steric hindrance effect.

2.1.4. Effect of the Type of Alkylalkoxysilane

For alkylalkoxysilanes \((\text{CH}_3)\text{Si(OC}_2\text{H}_5)\)\(_4–\alpha\), the following results are obtained [7].

In HCl-catalyzed solution, the hydrolysis rate increases with increasing number of CH\(_3\) groups \( \alpha \) bonded to silicon directly.

\[
 \text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5) \quad \text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5)\text{Si(OC}_2\text{H}_5) \tag{14}
\]

On the other hand, in NH\(_3\)-catalyzed solution, the rate of hydrolysis decreases as the number of CH\(_3\) bonded to Si directly increases. These effects can be explained by the inductive effect.

2.2. Gelation in the Phase-Separable Solution

It was found that macroporous silica containing macropores with controlled size and size distribution are prepared through the sol–gel reaction accompanied by phase separation of the starting solution [23–25]. For this purpose, starting solutions containing water-soluble polymers or organic substances, such as polyacrylic acid, together with alkoxysilane as silica source are used. When polycondensation of silica components progresses in these solutions, phase separation due to spinodal decomposition into silica-rich phase and organic molecule-rich phase occurs. Each phase shows connected irregular shape. When the organic molecule-containing phase is dissolved away, silica-gel body with micron-size continuous pores (0.5–5 \( \mu \)m) results. Silica frameworks contain nanometer-size (5–50 nm) open pores. Accordingly, the silica bodies thus prepared are so called a double pore system, that is, bigger pores are used for transport of liquid substrate and fine pores are used for absorption—desorption of molecules, for instance, when used as column for chromatography. It is suggested that the volume and size of both macropores and micropores can be adjusted by controlling gelation and growth of phase separation.
2.3. Reaction for Nonsilica Oxides

Sol–gel method is useful for fabrication of functional materials, such as photocatalyst, nonlinear optical materials, ferroelectrics, and superconductors. For this purpose, sol–gel preparation of simple and complex nonsilica oxides, including TiO₂, ZrO₂, Al₂O₃, ZnO, WO₃, Nb₂O₅, rare earth oxides, and so on, have to be studied. Metal alkoxides are often employed for those oxides. Most of them, however, are unstable, that is, they are rapidly hydrolyzed and are easily precipitated, which makes it difficult to form homogeneous multicomponent oxide products. Methods of controlling reactivity of the transition metal alkoxides are introduced below.

2.4. Hydrolysis and Condensation of Transition Metal Alkoxides

According to Livage [26] and Sanchez [27], the rate of both hydrolysis and condensation of transition metal alkoxides is very large compared to that of silicon alkoxides. As an example, the hydrolysis rate of Ti(OEt)₄ is about five orders of magnitude greater than that of Si(OEt)₄. The high reactivity of transition metal alkoxide may lead to the oligomer formation such as formation of [Ti(OEt)₄ₙ]ₙ (n = 2 or 3) and [Ti(OiMe)₄ₙ]ₙ.

In order to avoid the precipitation by suppressing the rate of hydrolysis and polycondensation, chemical modification is applied to the transition metal alkoxides [28,29]. An example of the chemical modification is that 2-methoxyethanol HOC₂H₄OCH₃, diethanolamine (HOCH₂CH₂)₂NH, or triethanolamine is used as solvent in place of simple alcohols. For example, zirconium isopropoxide [Zr(OiPr)₄(iPrOH)]ₙ reacts with 2-methoxyethanol solvent and forms Zr(OC₂H₄OCH₃)ₙ(O'C₃H₇), which suppresses the hydrolysis. Another example is the modification of the coordination state of the metal [26]. It is known that carboxylic acid reacts with the titanium isopropoxide, modifying Ti(O-iC₃H₇)₃ to Ti(O-iC₃H₇)(O'COCH₃). Another method is the use of β-diketones, such as acetylacetone and ethylacetate [30] and β-diols [31] as a part of the solvent. These solvents are often used to prevent the precipitation by complexing the ligand of metal alkoxides with high reactivity.

Besides control of reactivity, the modification makes the otherwise insoluble metal alkoxide soluble in alcohol. Copper alkoxides such as copper methoxide and propoxide are solids which are insoluble in alcohol. Modification of these alkoxides by β-diketone makes them soluble in alcohol, which makes possible the use of copper alkoxides as starting material for sol–gel processing. Kakihana [31] introduce a great number of water-soluble titanium complexes which can be used as precursors for synthesizing titania and titanium-containing oxides through sol–gel method. Synthesis, coordination structure, solubility, stability, and application of the soluble titanium complexes are also discussed.

Many of the functional oxides such as ferroelectric materials contain two or more kinds of metals as ion. Alkoxide of each metal may not be so stable, undergoing a rapid hydrolysis and polycondensation. Therefore, it is not easy to obtain homogeneous products from the mixture of different metal alkoxides. It is considered that the use of heterometallic alkoxides might dissolve this problem. Heterometallic alkoxide molecule holds two or more different metal atoms in it and is quite useful for preparation of homogeneous gel which contains different metallic species. Then many heterometallic alkoxides have been synthesized. Kessler [29] lists up a large number of heterometallic alkoxides, a part of which is shown in Table 3.

### Table 3 Heterometallic Alkoxides for the Synthesis of Complex Oxides [Taken from Kessler’s Article [29]]

<table>
<thead>
<tr>
<th>Complex oxides to be synthesized</th>
<th>Heterometallic alkoxides as starting material</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaTiO₃</td>
<td>BaTi(O'Pr)₄(ROH)₂</td>
</tr>
<tr>
<td>BaZrO₃</td>
<td>BaZr(OH)₄(O'Pr)₃ ROH</td>
</tr>
<tr>
<td>MgAl₂O₄</td>
<td>MgAl(acac)(O'Pr)₄</td>
</tr>
<tr>
<td>Na₄WO₃</td>
<td>NaW(OR)₆ R = Me, Et, 'Pr, 'Bu</td>
</tr>
<tr>
<td>PbMg₁/₃Nb₂/₃O₃</td>
<td>MgNb₂(OR)₁₂ 2EtOH</td>
</tr>
<tr>
<td>PbMg₁/₃Ta₂/₃O₃</td>
<td>MgTa₂(OR)₁₂ 2EtOH</td>
</tr>
<tr>
<td>(Pb, La)(Ti, Zr)O₃ (PLZT)</td>
<td>Pb₂Ti₂O(OR)₁₂(Ο'Ac)₂</td>
</tr>
<tr>
<td>SrBi₂Nb₂O₆</td>
<td>SrNb₂(OR)₁₂ 2ROH</td>
</tr>
<tr>
<td>SrBi₂Ta₂O₆</td>
<td>SrTa₂(OR)₁₂ 2ROH</td>
</tr>
</tbody>
</table>

2.5. Formation of Organic–Inorganic Hybrid

Organic–inorganic hybrids have microstructures in which inorganic species like —Si—O—Ti—O— are chemically connected with radicals like CH₃ and C₆H₅ or organic polymers like poly(methylmethacrylate). Philipp [33] prepared a transparent, homogeneous organic–inorganic hybrid material, which may be used as hard contact lens, by hydrolysis and condensation of a mixture of epoxysilane, methacryloxy silane, silicon alkoxide, titanium alkoxide, and monomeric methacrylate with a small amount of peroxide polymerization catalyst. Since then, the organic–inorganic hybrids have been attracting much attention, becoming a very important area of sol–gel technology.
Sol—gel processing of organic—inorganic hybrids can be carried out by the following ways.

2.5.1. Use of Alkylalkoxysilanes Such as CH₃Si(OC₂H₅)₃

This is the simplest method of preparing hybrids [34]. An alkyl group or groups bonding directly to a silicon atom of the silicon alkoxide molecule remain bonded in the process of hydrolysis and polycondensation of alkoxide groups in the alkoxide-alcohol-water-catalyst solution, until they are decomposed around 200 °C or higher temperatures. When dimethyldiethoxysilane [34] with two methyl groups per Si atom is employed for sol—gel synthesis, tetraethoxysilane has to be introduced into the starting solution to form three-dimensional network structure.

2.5.2. Use of Polydimethylsilane

Polydimethylsilane—tetraethoxysilane solutions subjected to hydrolysis—condensation reaction give rise to soft and hard organic—inorganic solid materials [35], depending on the content of polydimethylsilane.

2.5.3. Radical Polymerization of Alkyltrimethoxysilane Followed by Hydrolysis

It is shown [36] that transparent flexible or hard organic—inorganic hybrid gels are prepared by radical polymerization of vinyltrimethoxysilane followed by hydrolytic polycondensation of polyvinyltrimethoxysilane. The gels contain both siloxane and carbon—carbon bonds and show mechanical properties depending on the degree of condensation of the siloxane network.

2.5.4. Use of Polysilazane

Perhydropolysilazane (PHPS) composed of Si—N bonds with Si—H terminal groups has been employed as source material for SiO₂ since coating film on silicon wafers made from PHPS solution was found to have soft and refractive index close to those of silica glass. According to Kozuka [37], coating films prepared from xylene solution of PHPS are converted to dense, hard silica films at room temperature by exposing to the vapor from aqueous ammonia.

It is demonstrated [37] that crack-free, transparent PMMA (polymethylmethacrylate)—silica hybrid coatings are prepared from the PHPS—PMMA—xylene solution at room temperature through the similar procedure. These hybrid coatings exhibit higher hardness and chemical durability than alkoxide-derived ones, which indicates the great advantage of PHPS over alkoxide. It is also suggested that owing to the hydrophobic nature, PHPS makes it possible to prepare silica hybrids with a large amount of hydrophobic organic molecules.

2.5.5. Synthesis of Hybrids Containing Characteristic Polymers

Organic—inorganic hybrids with interesting properties are synthesized by sol—gel processing. For example, polyvinylpyrrolidone (PVP)—SiO₂ hybrid [38], polyoxazoline (POZO)—SiO₂ hybrid, and diphenyldimethoxysilane—SiO₂ hybrid [39] are synthesized.

2.5.6. Synthesis of Hybrids Containing Characteristic Silicate Anion

This is a method of preparing organic—inorganic hybrids by bonding characteristic silicate anions, such as cage-like structured cubic octamer Si₈O₂₀₈⁵⁻, with functional organic molecules [40]. This procedure may be called a building-block approach. An example is shown [41]. A methanolic solution, in which cubic octamers are present as tetramethylammonium silicate, is added drop by drop to a 2,2-dimethoxypropane [(CH₃O)₂C(CH₃)₂] solution of dimethyl dichlorosilane [(CH₃)₂SiCl₂], being kept under stirring at room temperature. Then, organic—inorganic hybrid, in which octamers are bonded with Si(CH₃)₃(OSi) units, results.

2.5.7. Ordered—Mesostructured Organic—Inorganic Hybrid

In the 1990s, ordered mesoporous thin silica films were prepared by calcining ordered—mesostructured thin organic—silica hybrid fabricated from the silicate solution containing surfactants or liquid crystals as template [42]. Recently, ordered—mesostructured organic—inorganic hybrid materials themselves also attract a considerable attention. Many works on the formation of organic—silica hybrids with ordered mesostructure, using surfactants as template, are reported.

Shimojima [43,44] prepared ordered, nanоструктурированные silica-based hybrid materials by sol—gel synthesis from designed organosiloxane oligomer having a terminal phenyl group without the use of any special structure-inducing agent. The formation of ordered mesostructure was attributed to the formation of amphiphilic species during hydrolysis of the alkoxide. With similar idea, Shimojima [45] synthesized siloxane—organic hybrids with well-ordered mesostructures through the self-assembly of amphiphilic molecules that consist of cubic siloxane heads and hydrophobic alkyl tails. Besides silica-based hybrids, for instance, ZnO-based hybrids with aligned ZnO nanocrystals are formed with the assistance of a surfactant.
2.6. Nonhydrolytic Sol-Gel Reaction

Most of the sol-gel preparation of metal oxides is based on the hydrolysis and subsequent polycondensation of metal alkoxides. In the early 1990s, the group of Vioux [46,47] developed the nonhydrolytic sol-gel process, in which M-O-M bridges are formed as a result of the condensation between metal halide, usually metal chloride, and metal alkoxide with elimination of alkyl halide:

\[ MX + MOR \rightarrow M-O-M + RX \] (15)

Here M is a metal, X is a halogen, and R is an alkyl group. As an example, monolithic gel is produced through non-hydrolytic reaction between AlCl₃ and Al(OiPr)₃ in CCl₄-(C₂H₅)₂O solvent by heating at 110 °C for 24 h. The resultant gel remains amorphous after calcination at 750 °C for 5 h. Stability of the gel and the presence of pentacoordinated Al ions in resulting monolithic gel show the difference between hydrolysis-condensation process and nonhydrolytic process.

In his article, Vioux [47] indicate that a great majority of nonhydrolytic reactions for oxide formation so far studied are based on the reaction between chloride precursors and alkoxide, ether or alcohol, as expressed by the following equations, respectively:

\[ MCl_n + M(OR)_n \rightarrow 2MO_{n/2} + nRCl \] (16)

\[ MCl_n + (n/2)ROR \rightarrow MO_{n/2} + nRCl \] (17)

\[ MCl_n + (n/2)ROH \rightarrow MO_{n/2} + nRCl + nHCl \] (18)

Many kinds of metal oxides, such as SiO₂, Al₂O₃, TiO₂, ZrO₂, SnO₂, and Nb₂O₅, and mixed oxides, such as SiO₂-TiO₂, SiO₂-Al₂O₃, TiO₂-Al₂O₃, TiO₂-ZrO₂, and SnO₂-Al₂O₃, have been prepared by the nonhydrolytic process.

It is reported [48] that this method is suitable for the preparation of organic-inorganic hybrids. As an example, fabrication of low-melting hybrid glass will be cited. Niida [49] synthesized such glass by reacting the mixture of dimethylchlorosilane Me₂SiCl and anhydrous orthophosphoric acid H₃PO₄ in flask at temperatures higher than 40 °C under flow of nitrogen. The reaction follows Eqn (19):

\[ H₃PO₄ + R₂SiCl₃ \rightarrow (OH)₂OP-O-SiClR₂ + HCl↑ \] (19)

\[ R = CH₃ \]

Network structure consisting of P-O-Si bondings is developed as the reaction proceeds with escape of HCl, resulting in transparent solid glassy materials with a low glass transition temperature.

2.7. Formation of Nanopores in the Gel

2.7.1. Noncontrolled Pore Formation

Inorganic dried gel products are usually porous materials. Examples of the porosity are given in Table 4 for three kinds of silica gels, which have different starting compositions. It should be noted that all the three gels consist of granular particles and so, pores are connective. It is observed that the gel contains several tens volume% of porosity when silica gels are formed without any special procedure. Gel no.1 is transparent due to the small pores [50]. The gel doped with fluorescent and laser molecules is applied as photo-functional gel. Gel no.2 is translucent [51], due to larger pores of 16 nm and can be applied to the fabrication of transparent silica bodies. The particles of gel no.3 are very big with 5 μm diameter [52]. It is observed that these big particles are secondary particles consisting of very small primary particles.

2.7.2. Controlled Formation of Porous Materials with Ordered Mesopores

Porous materials with controlled mesopores have been attracting much attention. In these materials, uniform mesopores are orderly arranged like an array of atoms or ions in crystals [53]. The outline of this technology is introduced below, according to the early work on this matter [42]. The amphiphilic surfactant, such as C₁₆H₃₃(CH₃)₃NOH/Cl, is employed to make ordered structure. The process starts with tetramethylammonium

<table>
<thead>
<tr>
<th>TABLE 4 Three Kinds of Silica Gels with Different Pore Diameters and Porosities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composition of starting solution (molar ratio)</strong></td>
</tr>
<tr>
<td>Si(OCH₃)₄</td>
</tr>
<tr>
<td>H₃O</td>
</tr>
<tr>
<td>CH₃OH</td>
</tr>
<tr>
<td>(CH₃)₃NCHO</td>
</tr>
<tr>
<td>HCl</td>
</tr>
<tr>
<td>NH₄OH</td>
</tr>
<tr>
<td><strong>Property</strong></td>
</tr>
<tr>
<td>Average pore diameter</td>
</tr>
<tr>
<td>Porosity (Vol.%)</td>
</tr>
</tbody>
</table>
silicate solution containing surfactants. It is assumed that circular micelles with the center part consist of hydrophobic ends of surfactants and hydrophilic silicon containing periphery. These circles are assembled face to face, producing micellar rods or cylinders. Since the surface of the cylinder attracts tetramethylammonium ions and becomes rich in silica, the rods gather together, forming hexagonal arrays. When these hexagonal arrays are calcined to decompose internal organic parts, 2-dimensional hexagonal arrays of mesopores arranged in silica networks are obtained in coating film.

After this work an extensive research has been directed to fabrication of ordered mesoporous materials. At present, this process is classified into two types [44]: One is the use of surfactants as structure creating agents that enables the formation of lamellar, hexagonal, or cubic mesostructures. Another type is the formation of self-directed assembly of certain organoalkoxysilanes during hydrolysis and polycondensation, providing a one-step surfactant-free route to certain organoalkoxysilanes during hydrolysis and polycondensation reactons proceed, leading to an increase in viscosity. When the composition of the starting Si(OC₂H₅)₄—H₂O—CH₃OH—HCl solution is appropriate, fibers can be drawn from the viscous sol having a viscosity higher than 10 poises. Gel fibers drawn from the solution can be converted to silica glass fibers by heating at 800–850 °C. It is noted that the maximum heating temperature throughout the processes, that is, 800–850 °C, is much lower than the temperature at which silica fibers are drawn from molten quartz or silica glass rod.

As to the spinnability of the solution, Sakka [59] found that the starting solution of low water content that is catalyzed with acid becomes spinnable. This condition corresponds to those for the starting solutions in which long-shaped linear particles are formed [60]. Figure 2 compares the rheological properties of spinnable and nonspinnable sols [61]. It is observed that the spinnable sol shows Newtonian behavior at high viscosities, while nonspinnable sols show structural viscosity or thixotropy characteristic of granular particles at such high viscosities where fibers can be drawn if the composition is appropriate.

These conditions for the occurrence of spinnability and rheological behavior of spinnable sols can be applied to many sols of different compositions from silica, that is, alumina, titania, superconducting oxide sols. Fibers of various compositions were prepared by the sol—gel method [62].

3.3. Sol—Gel-Coating Films
Coating of glass, ceramic, metal, and plastic substrates by the sol—gel method is very useful for modifying properties of substrates with a large or small surface area or providing substrates with new active properties, which are needed for developing optical, electronic, chemical, and biological devices. The sol—gel method allows application of the film to nonheat-resistant substrates.

3.3.1. Formation of Coating Films
Dip coating [63,64], spin coating, laminar flow coating [65], printing, and spray coating are employed in sol—gel coating.
Dip and spray coatings are two representative coating methods that are used in sol–gel coating. Here, discussion will be made mainly with dip coating; however, most of the discussion on the relation of thickness with crack formation will be valid also for other coating methods.

### 3.3.2. Adherence and Thickness of the Film

In dip coating, a coating procedure consists of dipping of the substrate in the solution, drawing up of the substrate from the solution and heating at 200–500 °C for the adherence of the film to the substrate. The adherence is attributed to the formation of bonding $\text{M}–\text{O}–\text{M}'$ ($\text{M}$, $\text{M}'$ are metal atoms in the film and substrate, respectively), when both the film and substrate are inorganic oxides. Roughly speaking, the thickness of the film is proportional to the amount of the solution dragged by the drawn-up substrate, which increases with increasing solution viscosity and substrate drawing speed. Based on this idea and detailed discussion [63,64], the thickness is expressed by the following simplified form [50]:

$$t \propto \eta^m v^n$$

(21)

where $\eta$ is the viscosity of the sol and $v$ is the substrate drawing speed. It is shown that $m$ has a value between 0.5 and 0.6 [50,64], and $n$ is close to 0.5 in most cases [66].

### 3.3.3. Cracking of Film and Critical Thickness

It is known [67] that for one coating run, the maximum thickness of the coating films of SiO$_2$ and other oxides achieved without peeling off and crack formation, that is, the critical thickness is 0.2–0.5 μm. Occurrence of cracks is assumed to be caused by the stress formation upon heating of the film. According to Kozuka [68], in situ observation of gel films under heating indicates that macroscopic cracking occurs in the heating-up stage and that the cracking onset temperature is affected by heating rate, film thickness, water-to-alkoxide ratio, and humidity. There are two methods for increasing film thickness without cracking. One is the repetition of the coating procedure. Examples can be found for BaTiO$_3$ [50] and TiO$_2$–GeO$_2$ films [69]. It is assumed that for every subsequent coating procedure, one uses a new substrate consisting of combination of the original substrate and previously applied films. Another method of avoiding crack formation is to add organic polymers to starting components. The critical thickness is larger than 0.7 μm when silica films are prepared from a solution containing CH$_3$Si(OC$_2$H$_5$)$_3$ [70]. BaTiO$_3$ of 2.1 μm thickness is obtained without cracking by using the starting solution containing poly-pyrrolidone by the amount corresponding to PVP/Ti(OC$_3$H$_7$)$_4$ ratio = 0.5 [68,71].

### 3.3.4. Striations in the Coating Films

Radiative striations, which consist of radially extending neighboring ridges and valleys, are often observed in sol–gel spin-coating films. In general, striations are regarded as defects causing light scattering. Kozuka [68] examined the effect of coating condition on the formation and properties of striations, using the sol prepared from a solution of molar composition, Si(OC$_2$H$_5$)$_4$:H$_2$O:HNO$_3$:C$_2$H$_5$OH = 1:4:0.01:2. It was shown that striations are formed by local Marangoni convection, irrespective of the presence or absence of rotation of substrate, and that use of less volatile alcohols is effective in avoiding striation, because striations once formed diminish in the course of gelation.

Uchiyama [72] demonstrated that striations are formed also in dip-coating films when starting solutions of molar compositions, Ti(OC$_3$H$_7$)$_4$:H$_2$O:HNO$_3$:i-C$_3$H$_7$OH: PVP = 1:2:0.2:60:0.10.7, are used. The addition of PVP
(polyvinylpyrrolidone) is necessary to cause striations. In the case of dip coating, linear arrangements of striations are realized, because the sol on the glass substrate unidirectionally flows downward. The formation of the arrangement of linear striations on titania films might be applied as a method of making surface patternings.

3.3.5. Microstructure Formation of Coating Films

Various microstructures as shown in Figure 1 are realized in sol-gel-coating films. For inorganic oxide-coating films, the presence or absence of crystallization and the type of crystallization are important microstructure characterizing coating films. It was found with Li2B4O7-coating film on Si wafer prepared from metal alkoxides LiOCH3 and B(OC4H9)3 that three kinds of microstructures, that is, amorphous, polycrystalline, and crystal-oriented microstructures, are made by changing the acid content of the starting solution [73].

Electric properties of ferroelectric thin films largely depend on the microstructure, especially orientation of crystals, even if the same crystalline species are precipitated. In sol-gel fabrication of ferroelectric coating films, many factors, such as kind of source materials, composition of starting solution, chemical reaction in the solution, kind of substrate, temperature and time of heat treatment, atmosphere in heating, and so on, have to be considered. As an example, Schneller [74] demonstrated that crystal orientation of Pb(Zr,Ti)O3 films on platinum depends upon the atmosphere at heat treatment.

For instance, higher oxygen pressure hinders formation of Pt_xPb phase at Pt-film boundary, leading to decrease in (111) orientation and, consequently, polarization of the film is very much reduced.

4. APPLICATIONS

Sol-gel method is characterized by the low-temperature processing, homogeneous products, fabrication of nanostructured materials, and organic—inorganic hybrid formation. Accordingly, this method is applied to fabrication of a great many kinds of functional materials [75,76]. Tables 5–7 show some of the important sol-gel-derived materials with photonic and electronic functions, thermal and mechanical functions, and chemical, biochemical, and biomedical functions, respectively. It is observed from those tables that various functional materials are fabricated by sol-gel method.

In this section, description will be made on selected applications: phosphors, lasers, antireflecting coating films, coating films for automobile windows, photocatalyst, transparent conductors, ferroelectric materials, aerogels, and biochemical and biomedical functions.

4.1. Phosphors

Intensive emission and stability of emission are two important luminescent properties for good phosphors. To achieve improvement, many kinds of phosphors have been prepared by sol-gel method.

As example, white LED (light-emitting diode) CaAl2Si2O8: Eu2+, Mn2+ phosphors [78] were prepared by sol-gel method. Stoichiometric amounts of Ca(NO3)2·4H2O, Al(NO3)3·9H2O, Eu(NO3)2·6H2O and Mn(NO3)2·xH2O were dissolved in ethanol and Si(OC2H5)4 was added. The resultant mixtures were stirred for 2 h and then heated at 70 °C until homogeneous gels were formed. The gels crystallized at 1000 °C showed emission bands excited by ultraviolet light. In a similar manner, a red-emitting long afterglow

<table>
<thead>
<tr>
<th>TABLE 5</th>
<th>Sol–Gel-Derived Materials with Photonic and Electronic Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>Specific function or microstructure</td>
</tr>
<tr>
<td>Photonic</td>
<td>Generation of light</td>
</tr>
<tr>
<td></td>
<td>Transmission of light</td>
</tr>
<tr>
<td></td>
<td>Optical absorption</td>
</tr>
<tr>
<td></td>
<td>Electrochromism</td>
</tr>
<tr>
<td></td>
<td>Sensing</td>
</tr>
<tr>
<td></td>
<td>Antireflection</td>
</tr>
<tr>
<td></td>
<td>Photovoltaic effect</td>
</tr>
<tr>
<td></td>
<td>Photocatalysis</td>
</tr>
<tr>
<td>Electronic</td>
<td>Transparent conductor</td>
</tr>
<tr>
<td></td>
<td>Ferroelectric</td>
</tr>
</tbody>
</table>
Sr$_3$Al$_2$O$_6$: Eu$^{2+}$, Pr$^{3+}$ phosphor with single phase was synthesized by sol–gel methods. In this case, heating at 1200 °C for 2 h in the reducing atmosphere improved the light intensity and the light-lasting time (15 min) of these materials [109].

There are many kinds of organic fluorescent molecules. In order to improve the stability of such molecules, the use of inorganic materials such as silica as matrix was proposed. Reisfeld [110] prepared a sheet of silica doped with rhodamine 6G, applying to fluorescent solar collector. Afterwards, organic–inorganic hybrid materials are used as matrix in place of silica, since the organic molecules are well dispersed in the hybrids as seen in the following Section 4.2.

### 4.2. Lasers

Nd glass laser is the first laser to which sol–gel processing was applied [111]. It is known that in preparing Nd-doped silica glass by sputtering or melting, homogeneous doping is difficult, because rare-earth elements tend to form clusters in glass. This results in low threshold concentration of laser element for lasing. One may assume that the formation of the domains can be avoided by the use of sol–gel processing. However, this is not the case. Even with the sol–gel processing, addition of Al$_2$O$_3$ to glass is necessary to reduce clustering of Nd ions. The advantageous use of sol–gel method for rare earth doped silica glass is found in fabrication of planar waveguide for amplification [79] in WDM device for optical communication systems. When erbium-doped planar optical wave amplifier prepared by sol–gel method is used in place of the fiber amplifier, the amplification can be carried out within a smaller volume. Laser properties of planar waveguides are affected by the host material, and erbium-doped planar waveguides are fabricated on the host of GeO$_2$-SiO$_2$ [112] and aluminosilicate [113].

It is well known that incorporation of functional organic molecules in silica matrix [8] opened new fields for sol–gel method around 1984. This idea was applied to the development of lasers containing organic laser dyes [77]. First, solid-state tunable lasers were prepared instead of liquid
tunable lasers. The next development of dye lasers was to use organic—inorganic materials, especially organically modified silicates (ormosils). Those materials can dissolve higher amount of organic laser dyes. For this purpose, various organic laser dyes, such as perylimide, pyromethene, rhodamine, and cyanine dyes, were synthesized and used for making red, blue, and UV lasers.

A problem with dye lasers is their weak photostability. To improve the photostability of dye lasers, photostable matrix as well as photostable laser dyes have to be explored. Reisfeld [114] prepared photostable berylimide dyes. Yagi [39] indicated that incorporation of phenyl groups in organic—inorganic hybrid matrix improves photostability of dye lasers.

Transparent spherical particles containing organic laser dyes become micro-lasers [115]. The diameter of the micro-sphere lasers is several to several tens of microns. The spheres are made from organic—inorganic hybrids. Introduced lights go around by reflecting at the sphere surface within the sphere crust, emitting laser light.

4.3. Antireflective Coating

The reflection of light takes place at the surface of glass. The reflectivity is about 4% for one surface of soda-lime-silica glass having refractive index of about 1.52. In other words, about 4% of the incident light is lost at every surface of glass. The reflection of light causes decrease of measuring accuracy in optical systems, lowering of the oscillation threshold and increased damage of glass laser lenses in the laser fusion system, and difficulty in clear sight for glass-protected paintings in museum and showcases.

When the refractive index \( n_f \) and the thickness \( d_f \) of a film satisfy the following equations, antireflection is achieved for the light of wavelength \( \lambda \):

\[
\begin{align*}
    n_f &= \sqrt{n_g} \quad (22) \\
    n_f \cdot d_f &= 1/4 + m/2 \cdot \lambda \quad (23)
\end{align*}
\]

where \( n_g \) is the refractive index of the glass substrate. Since the refractive index of conventional soda-lime glass is around 1.52, the refractive index \( n_f \) desirable for the antireflective coating should be 1.23. It is difficult to achieve such a low value for the practically useful film, and so what to do is to make the refractive index of the coating film as low as possible.

When a two-layered film is designed, the first higher refractive index layer is deposited on the glass, and on this layer the second lower refractive index layer is deposited to achieve antireflection, as shown below:

\[
\text{Substrate/first higher index layer/second lower index layer} \quad (24)
\]

For a three-layered antireflective film, the film is constructed in the following way:

\[
\text{Substrate/first intermediate index layer/second higher index layer/third lower index layer} \quad (25)
\]

So far, various antireflective films have been prepared by sol–gel method. Mukherjee [116] prepared single-layered antireflective film of silica-rich porous glass on glass lenses for laser fusion system by the use of phase separation of deposited borosilicate glass film. A small reflectivity of 0.15—0.7% at the wavelength of 1.06 \( \mu \text{m} \) and an increase in the damage threshold of the fusion laser from 4 \text{J/cm}^2 to 22 \text{J/cm}^2 were achieved. Yoldas [117] deposited a porous antireflective film of pure silica on fused silica surface by sol–gel reaction involving \( \text{Si(OC}_2\text{H}_3)_4 \). The pore size and morphology were tailored by controlled heat treatment. Antireflection was achieved over the entire spectral range of silica with better than 99% transmission. Floch [118] developed a scratch-resistant one-layer antireflective coating film with a thickness corresponding to \( m = 0 \) in Eqn (22). The resulting coated glass can be applied to general uses, such as household articles, architectures and ophthalmic lenses, as well as fusion laser glasses. Yamazaki [84] developed single-layered antireflective coating film from the starting solution containing methyliethoxysilane (MTES) as well as tetraethoxysilane (TEOS) as the source of silica. The low refractive index is achieved by heating at 500 °C, which creates pores or voids as a result of decomposition of clusters of methyl groups remaining in the gel. This simple method can be applied to automobile windows. Three-layered antireflective coating films were developed in Schott of German [119]. The film has a constitution given by Formula (24), corresponding to glass substrate/\( \text{SiO}_2—\text{TiO}_2 \) layer/\( \text{TiO}_2 \) layer/\( \text{SiO}_2 \) layer. Based on this constitution, Schott fabricated commercial products “Amiran” for show windows and “Conturan” for a filter to increase CRT contrast [120].

4.4. Coating Films for Automobile Windows

Representative sol–gel-coating films which are applied to automotive windows are listed in Table 8.

4.4.1. Reflective Film

The reflective coating film as combiner of reflective screen in the head-up-display (HUD) system on the inner surface of windshield is used to display automobile information [121,122]. The coating was made by Central Glass Company in Japan and became a trigger for application of sol–gel coating for preparing various functional coatings on automobile windows. In the HUD system, the display light is reflected by the combiner film, reaching the
driver’s eye. It is designed so that the displayed information, such as speed of the car as image, appears in front of the driver. Complying with regulations concerning automobiles, TiO$_2$-SiO$_2$-coating films with refractive index of 2.05, thickness of 220 nm, and size of 4.5$''$ × 3$''$ were prepared.

### 4.4.2. Antireflective Film

For the antireflective film, refer to previous Section 4.3.

### 4.4.3. Colored Film

Neutral-colored glasses are important for car windows. Some passengers request having dark tinted windows for the privacy, comfort, and distinguished design. In order to accomplish this color by coating window glasses with gray color film, films consisting of two layers are applied to the glass; the inner layer consists of Cu–Mn–Co–Cr–O spinel-type oxide pigment and the outer layer consists of SiO$_2$ matrix containing oxides of copper, manganese, cobalt, and chromium [123]. After coating with 200 nm thickness inner layer and 30 nm thickness outer layer, the coating film is fired at 700 °C for a few minutes. After the treatment glass shows lower transmittance in the whole wavelength region, effectively shielding ultraviolet and infrared rays and darkening the interior of automobiles for privacy.

### 4.4.4. UV Shielding Film

UV shielding glass is needed for automobile windows, in order to protect passengers and internal commodities from the detrimental effect of UV rays. For this purpose, conventionally, UV absorbing agents were introduced into glass. However, application of UV shielding film on window glass is more convenient. Introduction of TiO$_2$ and CeO$_2$ is known to provide highly UV-absorbent coatings, which is much more stable than the UV shielding film based on organic UV absorbing dyes. Tomonaga [124] noticed that the film containing Ce and Ti has a high refractive index reaching 2.1, showing marked interference colors of reflected light. This intense coloring was suppressed by inserting an intermediate layer consisting of TiO$_2$-SiO$_2$.

### 4.4.5. IR-Shielding Film

IR (infrared ray) shielding film for automobile windows are needed to suppress the temperature rise within cars and keep comfort of passengers. For windshield, IR shielding is carried out through the intermediate film in glass, while IR shielding of front-door windows are carried out by applying a film to the surface of glass. At present, IR shielding for front doors [125] is based on ITO (In$_2$O$_3$-doped with SnO$_2$) fine particles in SiO$_2$ film. ITO particles absorb the IR of 0.3–2 μm wavelengths, while they do not absorb the visible lights so much. Therefore, the transmittance of the film over 70% required by the regulation on automobiles is obtained. To coat soda-lime glass with SiO$_2$–ITO infrared-shielding film, a solution containing 3-glycidoxypropyltrimethoxysilane and tetraethoxysilane as source of SiO$_2$ is used. The coating film heat-treated at 160 °C shows high strength and high pencil hardness of 9H, besides high IR-shielding property, that is, low transmittance in the infrared region of wavelengths longer than 1.2 μm.
4.4.6. Water Repellent Film

In order to drive safely and comfortably on rainy days without wipers, water-repellent surface of windshield of cars is required. Originally, water-repellent reagent was applied to the surface of glass, and this worked for a while. In the early 1990s, sustainable water-repellent coating films on glass substrate was developed, based on the sol–gel method [127]. The coating film is prepared by a two-step process. In the first step, porous SiO2 film is formed by heating alkoxy-derived SiO2 gel layer at 290 °C and in the second step, partially hydrolyzed fluoroalkyltrimethoxysilane is applied on the precoated porous SiO2 layer. The coating film showed excellent water repellency expressed by the contact angle of water as high as 100°.

Around the year 2000, it was discussed that for the clear sight the glass surface should have high water drop slidability as well as high water repellency. Asahi Glass Company [100] developed abrasion-resistant and chemically durable new water repellent film, which shows high water drop slidability as well and applied the film to automobile windows. Such films are also developed in other glass companies [126].

4.4.7. Hydrophilic Coating

Hydrophilic coatings are applied to the outside mirror of cars to enhance rearward visibility on rainy days [128]. When the mirror is covered with rain droplets, the mirror does not work, due to a cloudy surface. The hydrophilic coating spreads out rain drops on the mirror surface to thin water film. Presently, manufactured hydrophilic film consists of SiO2 matrix containing TiO2 fine particles, which show photocatalytic hydrophilicity. There is another type of film, in which the first layer on glass is TiO2 photocatalyst and the second layer on the first layer is very porous SiO2 hydrophilic layer.

4.4.8. Antifogging Film

An antifogging film [84] was developed based on polyurethane, which absorbs much water. The film is applied to the inner surface of windshield. The antifogging film of 20 μm thickness shows an outstanding antifogging performance even when the outside temperature is as low as the freezing temperatures.

4.5. Photocatalyst Film

4.5.1. Performance, Preparation and Application of Photocatalyst

Photocatalysts are defined as materials which decompose detrimental substances under the sun lights containing UV rays. Mainly, TiO2 is used as photocatalyst at present. Among polymorphs of TiO2, anatase phase shows the most effective photocatalytic effect. The effect of photocatalysts is explained by the following example of photocatalytic decomposition of surface stains, which occurs in photocatalytic self-cleaning coatings [129]. Pairs of electron and hole are excited by UV lights. The photo-generated electrons produce anion radicals like O2− by reacting with O2. The holes produce active OH radicals by reacting with water. Those active species decompose organic compounds, such as stains of oil and fat. Thus, due to its highly oxidizing and reducing power, photocatalyst is employed for decomposition of organic stains, elimination of bad smells, decomposition of nitrogen oxide (NOx), antibacteria, and so on.

4.5.2. Fabrication of Photocatalyst Film: Self-Cleaning of Windows

Fabrication of photocatalyst film will be discussed with self-cleaning window glass as example [87,130]. Starting solution containing Si(OC2H5)4, TiO2 fine particles, and SiO2 fine particles is applied onto plate glass by spray coating. TiO2 fine particles used are anatase-phase particles of 20 nm in diameter. The thickness of the film is 200–300 nm. Since a large size plate glass cannot be heated to 400–500°C without breaking, the anatase particles prepared separately by sol–gel method are used. The addition of silica to TiO2 is made, in order to keep hydrophilicity of the coating film. The prepared film becomes hydrophilic on exposure to UV rays in solar radiation, working as self-cleaning film [131]. Photocatalyst decomposes organic contaminants on the coating under sunshine, and rain enters between the non-decomposed contaminants and coating, eliminating the contaminant.

4.5.3. Improvement of Photocatalytic Activity

A large number of papers have been published on improvement of the activity of TiO2 photocatalysts. One of the methods is an increase in surface area of TiO2 photocatalyst. Luo [132] accomplished some improvement of catalytic activity by preparing TiO2 anatase nanocrystals with a diameter of 8–10 nm using a mixed template of polyethylene glycol and cetyltrimethylammonium bromide. The resultant TiO2 crystals showed high photocatalytic activity determined by reaction 3I− → I3− + 2e−.

Another method of improving photocatalytic activity is to dope TiO2 photocatalyst. As an example, Raileanu [133] prepared coatings of TiO2 doped with S and Ag for water purification, showing that the photocatalytic activity of the doped samples is higher than that of undoped sample in the test using degradation of organic chloride compounds from aqueous solution.
4.5.4. Visible-Light Photocatalyst

In TiO₂ anatase photocatalyst, which is currently available for practical use, only UV lights contained in the sun lights work for photocatalysis. On the other hand, visible-light photocatalysts [134] make use of visible lights, working inside of houses and buildings as well as outside, and so, their development is highly desired. Visible lights occupy most of the sun lights. At present, however, the efficiency of visible-light photocatalyst is relatively low, and exploration of highly efficient materials is needed. In addition, the visible-light photocatalyst films are necessarily colored and cannot be applied to windows which are basically colorless and transparent.

There are visible-light photocatalysts doped with transition element, such as Cr and V [135], and those doped with anion, such as N [136]. It appears that TiO₂₋ₓNx, that is, TiO₂ doped with N, is more efficient. Also, TiO₂ doped with two kinds of anions Br and N is proposed [86]. Besides these TiO₂-based materials, microcrystalline CuAl₂O₄ spinel [137] and In₂O₃₋ₓCaIn₂O₄ [138] are reported to show visible-light photocatalytic activity.

4.6. Transparent Conductive Film

Transparent conductive coating films are employed as transparent electrodes of solar cell, liquid crystal display panel, electrochromic devices [82], and so on. Most important ones are In₂O₃:Sn film (ITO film) and SnO₂:Sb (ATO film). Low resistivity is prerequisite for practical application of transparent conductive films. At present many of the transparent conductive films are fabricated by PVD methods such as sputtering and vacuum vapor deposition. It is known that the resistivity of ITO film prepared by the optimum condition for low resistivity is 1 × 10⁻⁴ ohm cm. It is generally believed that ITO shows lower resistivity than ATO.

Sol–gel preparation of ATO film was carried out in the mid-1980s [139]. The ATO film was prepared from a starting solution containing Sn(OC₄H₉)₄ and Sb(OC₃H₃)₃ as source of SnO₂ and Sb₂O₃, respectively. Resistivity of the film was 2.5 × 10⁻³ ohm cm, lower than that (7.5 × 10⁻³ ohm cm) of coating prepared on the high temperature substrate by spraying. Kodaira [140] used dip-coating method.

Sol–gel preparation of ITO film was published in 1982 [89]. Soda-lime glass substrate was coated with the solution of acetylacetonates of indium and tin by dip coating. Heating of the gel film at 500 °C produced ITO film thinner than about 10 nm. The lowest resistivity of 1.3 × 10⁻³ ohm cm was attained when 3–7% SnO₂ was added to In₂O₃. Furusaki [141] prepared ITO films from solution of In₂(SO₄)₃·nH₂O and SnSO₄ and from colloidal particles synthesized from In(NO₃)₃·3H₂O and SnCl₄. It is found that the resistivity of the film decreases upon heating in vacuum. The lowest resistivity obtained by [142] is 6–8 × 10⁻⁴ ohm cm. Lower resistivity of 2.7 × 10⁻⁴ ohm cm by dip coating [143] and 9.5 × 10⁻⁵ ohm cm [144] are reported. Examples of reported resistivities of ITO and ATO are shown in Table 9. It is desirable to continue research on this subject so that low resistivity can be obtained constantly for glass substrates of large size.

<table>
<thead>
<tr>
<th>No.</th>
<th>Kind of film</th>
<th>Thickness (nm)</th>
<th>Resistivity (ohm cm)</th>
<th>Remark</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>ATO</td>
<td>247</td>
<td>2.5 × 10⁻³</td>
<td></td>
<td>[139]</td>
</tr>
<tr>
<td>(2)</td>
<td>ATO</td>
<td>140</td>
<td>3 × 10⁻³</td>
<td>Lowest resistivity</td>
<td>[140]</td>
</tr>
<tr>
<td>(3)</td>
<td>ITO</td>
<td>&lt;70</td>
<td>1.3 × 10⁻³</td>
<td>Heated at 550 °C</td>
<td>[89]</td>
</tr>
<tr>
<td></td>
<td>ITO</td>
<td>120–14</td>
<td>2–4 × 10⁻³</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6–8 × 10⁻⁴</td>
<td>Heated in vacuum</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>ITO</td>
<td>2000</td>
<td>2.2 × 10⁻³</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 × 10⁻⁴</td>
<td>Heated in vacuum</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>ITO</td>
<td>286</td>
<td>1.4 × 10⁻³</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.7 × 10⁻⁴</td>
<td>Heated in N₂–1%H₂ gas at 600 °C (dip coating)</td>
<td>[143]</td>
</tr>
<tr>
<td>(6)</td>
<td>ITO</td>
<td>1.9 × 10⁻⁴</td>
<td>Lowest, Spray</td>
<td></td>
<td>[144]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.5 × 10⁻⁵</td>
<td>Heated in N₂–1%H₂ gas at 600 °C (Spray coating)</td>
<td></td>
</tr>
</tbody>
</table>
4.7. Ferroelectric and Related Dielectric Coating Films

4.7.1. Application of Ferroelectric Films

Besides applications as electro-optic, nonlinear optical and other optical films in photonics [80,145], ferroelectric and related dielectric films (later, simply called ferroelectric films) play important roles in electronics as device elements, such as capacitor, nonvolatile FeRAM (random access memory), piezoelectric transducer, pyroelectric temperature sensor, and element of surface elastic wave device.

4.7.2. Characteristics of Sol–Gel Method in Fabrication of Ferroelectric Films

Sol–gel processing makes it possible to easily realize: (1) precipitation of single phase perovskite by adjusting the composition of starting solution, reactions and condition of thermal treatment of the film; (2) preferred orientation of crystals in the film [146–148]; (3) formation of multilayered film consisting of layers of different kinds of ferroelectrics [149] and layers consisting of ferroelectric and dielectric films [150], and (4) amorphous film of the composition corresponding to a ferroelectric material [148,151]. Ohya [149] pointed out that in sol–gel prepared multilayered film consisting of PbTiO3 and PbZrO3 layers, the boundary between the two layers is distinct, while the boundary between PbTiO3 and SrTiO3 layers is not clear. It is discussed that in the latter multilayer film, Pb component diffuses into SrTiO3 layer.

4.7.3. Kind of Ferroelectric Films

Some of the examples of ferroelectric films prepared by sol–gel method are listed in Table 10. Originally, most of those ferroelectric materials were synthesized by conventional ceramic processing and later by physical and chemical vapor deposition. It is known that among ferroelectric materials shown in Table 10, barium titanate, lead titanate zirconate, lithium niobate, and bismuth titanate served as the basis of enormous kinds of ferroelectric materials. One aspect to be considered is inhibition of the use of lead-containing materials for environmental reason. Along with this, extensive efforts were made to find novel ferroelectric materials replacing lead titanate zirconate, for instance, and, as a result, new ferroelectric materials derived from bismuth titanate were discovered. An example is the sol–gel prepared CaBi4Ti4O12- coating film [161], which shows a high remnant polarization required for high-performance FeRAM (ferroelectric random access memory).

<table>
<thead>
<tr>
<th>No.</th>
<th>Composition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>BaTiO3</td>
<td>[50]</td>
</tr>
<tr>
<td>(2)</td>
<td>Ba0.5Sr0.5TiO3</td>
<td>[152]</td>
</tr>
<tr>
<td>(3)</td>
<td>PbTiO3</td>
<td>[153]</td>
</tr>
<tr>
<td>(4)</td>
<td>PbZr0.5Ti0.5O3 (PZT)</td>
<td>[153]</td>
</tr>
<tr>
<td>(5)</td>
<td>Pb(Mg1/3Nb2/3)O3 (PMN)</td>
<td>[154]</td>
</tr>
<tr>
<td>(6)</td>
<td>PbFe0.5Nb0.5O3 (PFN)</td>
<td>[155]</td>
</tr>
<tr>
<td>(7)</td>
<td>LiNbO3</td>
<td>[146]</td>
</tr>
<tr>
<td>(8)</td>
<td>KNbO3</td>
<td>[156]</td>
</tr>
<tr>
<td>(9)</td>
<td>Sr0.5Ba0.5Nb2O6 (SBN)</td>
<td>[157]</td>
</tr>
<tr>
<td>(10)</td>
<td>SrBi2Ta2O9</td>
<td>[158]</td>
</tr>
<tr>
<td>(11)</td>
<td>SrBi4Ti4O15</td>
<td>[159]</td>
</tr>
<tr>
<td>(12)</td>
<td>BaZr0.5Ti0.5O3</td>
<td>[160]</td>
</tr>
<tr>
<td>(13)</td>
<td>BiFeO3</td>
<td>[161]</td>
</tr>
<tr>
<td>(14)</td>
<td>BiFeO3–Bi4Ti3O12</td>
<td>[162]</td>
</tr>
</tbody>
</table>

4.8. Aerogels

Silica aerogels were made by Kistler [1,164] around 1930, using sol–gel method combined with supercritical drying. Aerogels are the first significant sol–gel product. No other method can produce aerogel.

4.8.1. Preparation of Aerogels

Kistler [2] used water glass as starting material to produce silica aerogels. At present, however, Techener’s method [165,166], which uses alkoxysilanes as starting material, is usually employed. First, a starting solution of the tetraalkoxysilane–alcohol–water system is prepared and is subject to hydrolysis–polycondensation. The resultant gel is then dried in an autoclave under the supercritical condition concerning methanol, that is, at higher temperature than the critical temperature $T_c = 239.4 ^\circ C$ and under higher pressure than the critical pressure $P_c = 80.9$ bar ($80.9 \times 10^5$ Pa). Drying under supercritical condition is necessary to avoid occurrence of capillary pressure which causes shrinkage and possible fracture of the gel. By this way, transparent aerogels of very large size, for example, $50 \times 60 \times 80$ cm, can be obtained without fracture and shrinkage. Later, an alternative method using drying under the CO2 supercritical condition, which is milder than the methanol supercritical condition, was proposed [167].
The $T_c$ and $P_c$ of CO$_2$ are 31.1 °C and 73.9 bar, respectively. The use of CO$_2$ supercritical drying makes it possible to carry out supercritical drying at low temperatures close to ambient temperature. To do this, alcohol is exchanged by CO$_2$ before drying.

### 4.8.2. Hydrophobic Silica Aerogel

Aerogels as prepared are usually transparent, but become translucent or opaque in a while. Aerogels are hydrophilic and in ambient atmosphere they absorb water, which causes aggregation of silica microparticles in aerogel body. To protect aerogels from water and keep the transparency, Yokogawa [93,168] prepared hydrophobic aerogel by trimethylsilylation treatment before supercritical drying. For trimethylsilylation, rinsed silica gel is immersed in hexamethyldisilazane, which reacts with hydroxy groups on the surface of silica microparticles, capping the reactive OH groups and changing the hydrophilic surface to hydrophobic surface, as follows:

\[
(\text{CH}_3)_3\text{SiNHSi(CH}_3)_3 + 2\equiv\text{SiOH} \\
\rightarrow 2\equiv\text{SiO}−\text{Si(CH}_3)_3↑+\text{NH}_3
\]  

(26)

Resultant silica aerogels are hydrophobic and are kept transparent. Research of hydrophobic silica aerogels is continued [169].

### 4.8.3. Properties and Applications of Aerogels

Silica aerogels are highly transparent due to the small particle size and small particle–particle distance, in spite of their high porosity. As shown in Table 11, aerogels are characterized by very high porosity, very low density, very low refractive index, and extraordinary low thermal conductivity. These interesting properties are related to applications of aerogels [168,170,171].

Due to their small refractive index, silica aerogels are used in elementary particle physics. Cherenkov counters for detection of relativistic particles need silica aerogels, because their refractive index is in a range which is not covered by compressed gases and liquids. Another application is cosmic dust catcher. Due to their porous structure, aerogels catch hypervelocity cosmic dusts of micron size.

### 4.8.4. Fabrication of Aerogels by Ambient Pressure Drying

The above-mentioned aerogels were prepared through drying in supercritical condition, that is, at high temperatures under high pressures. Even CO$_2$ supercritical condition includes high pressure reaching 74 bar (7.4 MPa). One can understand that drying under ambient pressure [172] is very effective in facilitating production of aerogels and reducing fabrication cost. Ambient pressure drying for synthesizing low-density gels was first attempted by [173], who reported transparent monolithic gels with similar properties as silica aerogels. Since then, a large number of papers have been published on this subject. According to Aravind, ambient pressure drying is now a requisite for further commercialization of aerogels.

The first method adopted for the success in ambient pressure drying is to strengthen the gel structure before ambient pressure drying [173]. When the alcogel (gel prepared in alcoholic solution) is immersed in a solution of tetraethoxysilane, additional precipitation of silica occurs and the gel structure is strengthened. In this way, shrinkage of the gel is reduced and aerogel is obtained. Hydrophobic capping of the gel surface before ambient pressure drying is effective in preventing gels from shrinkage. Capping is done by silylation of Si−OH groups on the surface of gels; H from the Si−OH group is replaced by stable Si−R group as follows:

\[
\text{Si−OH} \rightarrow \text{Si−O−Si−R}
\]

(27)

Among many silylating reagents, hexamethyldisilazane (HMDZ) and trimethylchlorosilane (TMDZ) are very powerful, effectively capping the OH group by trimethylsilylation. It is believed that during ambient pressure drying, repulsion works between trimethylsilylated hydrophobic groups, retaining the pore structure of gel. The gel obtained by [174] is 0.075 g/cm$^3$ in density, 97% in porosity and 0.07 W/mK in thermal conductivity. It should be noted that the increase of silylating reagent increases pore size, making the gel opaque.

Ambient pressure drying makes aerogel production cost-effective. In addition, the use of sodium silicate as precursor would make commercialization of aerogel easier. Also in the case of sodium silicate, silylation of the surface serves in reducing the shrinkage of gel [175−177]. By this way, aerogels with low density and high porosity similar to those of supercritically dried aerogels are obtained.

---

**TABLE 11** Some Properties of Silica Aerogel

<table>
<thead>
<tr>
<th>Property</th>
<th>Values or description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porosity</td>
<td>89.1−98.4 Vol.%</td>
</tr>
<tr>
<td>Density</td>
<td>0.035−0.24 g/cm$^3$</td>
</tr>
<tr>
<td>Refractive index</td>
<td>1.008−1.055</td>
</tr>
<tr>
<td>Thermal conductivity</td>
<td>0.015 W/mK*</td>
</tr>
<tr>
<td>Appearance</td>
<td>Transparent, translucent or turbid</td>
</tr>
</tbody>
</table>

*Lower than that of polyurethane foam.*
4.8.5. Aerogels Other than Silica

Aerogels other than silica draw much attention. Some examples are shown. In the TiO$_2$—SiO$_2$ system, TiO$_2$ microcrystals of anatase type are precipitated at the surface in the ethanol supercritical drying process [178]. Accordingly, the obtained aerogel with large surface area shows excellent photocatalytic performance. One can find an example [175], in which ambient pressure drying is done to produce TiO$_2$—SiO$_2$ aerogels. Vanadium oxide aerogel with high strength due to cross-linking structure is prepared [179]. Vanadia (VO$_3$) wet gels are prepared by copolymerization of a vanadium tripropoxide solution in acetone/water. Wet gels are subjected to cross-linking by diisocyanate and then changed to aerogels by supercritical drying using supercritical CO$_2$ fluid. High-strength vanadium aerogels are formed. Aerogels of alumina Al$_2$O$_3$ [180] and zinc ferrite Fe$_2$ZnO$_4$, a complex oxide, [181] were prepared.

Besides oxides, aerogels of sulfides are also prepared. For example, CdS aerogels are produced [182] by CO$_2$ supercritical drying of wet CdS gels.

Sol—gel method is also applied to produce aerogels of organic compounds. Light-weight organic aerogels are prepared by CO$_2$ supercritical drying of wet polydicyclopentadiene gels [183]. It is interesting to note that carbon aerogels are obtained by firing organic aerogels prepared from polybenzoxazine [184].

It is known that pure silica aerogels are brittle and easily fractured. In order to improve mechanical property of silica aerogels, organic—silica hybrid aerogels are prepared. Organic—silica hybrid aerogels were obtained [185] by alkylalkoxyxysilane precursors, such as methyltrimethoxysilane (MTMS) and methyltriethoxysilane (MTES), which have one nonhydrolyzable organic radical. These hybrid aerogels exhibited high porosity and low thermal conductivity. The incorporation of methyl radical into silica network improved mechanical property very much. Kanamori [186] prepared highly elastic organic—silica hybrid aerogels of composition CH$_3$SiO$_{1.5}$ from methyltrimethoxysilane by subcritical drying.

4.9. Gels with Biochemical and Biomedical Function

Sol—gel prepared biomaterials, that is, materials with biochemical and biomedical functions, will be described. These materials are fabricated by entrapping biological molecules or living tissues in organic—inorganic hybrid [187] or oxide gels. Entrapped matters are classified as follows: proteins [188,189], such as enzymes and antibodies, microorganisms, such as bacteria [190] and yeast [191], living tissues and organs, and antibacterial substances. Sol—gel-derived gels are suitable as host of biologically active entities, because gels are porous and the size and size distribution of pores can be controlled. The entrapped materials display the activity only when they contact the substrate solution passing through the pores.

4.9.1. Biosynthesis

Biosynthesis takes place in gels in which enzymes or enzyme-producing microorganism and cells are immobilized. The substrate solution continuously undergoes enzyme reactions, such as organic synthesis and degradation, during passing through the gel [192]. For this purpose, various enzymes, such as hemoglobin and mioglobin [193] and lipase [194], are entrapped in SiO$_2$ gel. Entrapping of microorganism by sol—gel method was originated by [195]. One of the many examples is entrapping of *Escherichia coli* bacteria in silica gel with gelatin [196]. Also, entrapping of plant cells [197] and animal cells [198] is studied.

4.9.2. Biosensing

Biosensor consists of a gel entrapping biomolecules, such as proteins, which react with molecules to be analyzed, resulting in a change in spectra of optical absorption or fluorescence [199]. Phycobiliprotein-doped gel [200], for instance, presents a sensor based on the change in fluorescence spectrum. Availability of carbon—enzyme hybrid electrodes is discussed with glucose biosensor [104].

4.9.3. Cleaning of Environments

For cleaning of environments, the toxic organic compounds remaining in soils are decomposed by enzymes or enzyme-producing microorganisms entrapped in gels. Also, sol—gel-derived material containing bacteria-killing compound is used to suppress the contamination by bacteria in environments. The activity of parathion hydrolase entrapped in SiO$_2$ gel is examined [201]. Atrazine is degraded by Pseudomonas strain ADP entrapped in gel [202]. Copper-containing silica-coating films, which show high antibacterial activity, are prepared on glass substrates by sol—gel method. Also, nanosized silver particles obtained as Ag/SiO$_2$ composite sol exhibit marked antimicrobial properties [203]. Ovesi [106] designed mesoporous titania photocatalyst film, which shows excellent antibacterial property to *Escherichia coli*.

4.9.4. Artificial Tissues and Organs

The application of sol—gel method consists in incorporating living cells, tissues, and organs into gels and examining their activity. Pope [195] incorporated silica gels entrapping bovine hepatocyte cells into animals. Sakai...
transplanted aminosilica gel encapsulated pancreatic islets into otherwise diabetic mice, demonstrating that no indication of diabetes appeared up to three months. Kokubo [204] fabricated CaO—SiO₂—P₂O₅ glass ceramics for bone implant by sintering and crystallizing melt-quenching glass and examined the formation of hydroxyapatite on the surface. Recently, sol—gel preparation of glasses for bone implant is attempted by many researchers [205,206]. Sol—gel-prepared bone implant is characteristic of porous nature.

4.9.5. Medical Diagnosis

Specific proteins entrapped in gels are used for diagnosis by immunoassay, which utilizes antibody—antigen reactions in porous gel matrix. In an example [207], echonococcosis antigens entrapped in silica gels are put in a container and diluted human sera from infected patient is poured into the container. A peroxidase conjugate is used for the optical detection of antibody—antigen associations. Other examples of immunoassays using the reaction of sol—gel-entrapped antigens with antibodies are studied by Livage [107] and Turnianski [208].

4.9.6. Drug Delivery

Sol—gel-derived silica is attracting attention as carrier for targeted and controlled drug delivery. As an example, Fidalgo [209] examine the wet sol—gel silica as delivery device for phenytin. Finnie [210] discuss biodegradability of sol—gel silica microparticles for drug delivery. The biodegradability is important, because nanoparticles taken up into cells must degrade to avoid the toxicity risk associated with accumulation of nondegradable fine particles.

4.9.7. DNA Array and Protein Array

An array of a large number of micro-DNA spots or wells prepared on a supporting plate made of silicon, slide glass, resin, and so on is called DNA array, DNA microarray, or DNA chip. Similar naming is given to protein and peptide. Microarray technology, based on DNA array [108], protein array [211], peptide array [212], and so on, allows fast, easy, and parallel detection of biological information in a single experiment under the same condition. It is noted that in microarray technology, the experiment is carried away with probes placed on several hundreds to several thousands of micro-wells which are fabricated on a support of several centimeter size [108]. DNA arrays are used for sequestering of DNA and interpretation of gene function. Protein arrays are used for characterization of proteins and interaction with other biomolecules. Peptide arrays are used for detection of proteins.

The role of sol—gel processing is to fix DNA or protein probes to solid supports. It is important in fixing biomolecules to preserve their biochemical functions, especially biological activity. For this purpose, the low-temperature processing characteristic of sol—gel method is quite suitable. However, the use of acid or base catalysts for conventional sol—gel reaction is prohibited. Also, alcohol cannot be used as solvent in starting solution. Therefore, usually, DNAs and proteins are fixed to the substrate by sol—gel-processed aminosilane or aminosilane-silica hybrid so that original functions of biomolecules may be preserved.

5. RECENT TOPICS ON SOL—GEL METHOD

Topics concerning novel materials produced by sol—gel processing and application of sol—gel processing to important advanced materials explored by conventional processing techniques will be described.

5.1. Elastic Aerogel

Kanamori [186] prepared novel aerogels with unusual mechanical strength against uniaxial compression. The novel aerogels having methylsilsesquioxane networks were prepared from CH₃Si(OCH₃)₃ (MTMS) solution containing surfactant and urea as phase-separation-inducing agent and as accelerator for the condensation reaction, respectively. The optimized aerogels dried under a supercritical condition showed not only similar properties as conventional silica aerogels such as high transparency and high porosity but also outstanding mechanical strength against compression. The compression test showed that an MTMS aerogel sample prepared from starting solution consisting of water—acetic acid—MTMS—CTAB (cetyltrimethylammonium bromide) shrinks up to 84% in linear scale upon uniaxial loading and springs back to the original size upon unloading. This phenomenon appeared really amazing when the author showed the experiment before us. This unusual mechanical behavior is attributed to the trifunctional flexible networks of MTMS aerogel, low silanol concentration, and high concentration of hydrophobic methyl groups. Therefore, aerogels prepared by ambient pressure drying and even xerogels show similar mechanical behavior, if they have the structure comparative to those prepared from MTMS by supercritical drying.

5.2. Migration of Components in Thin Films During Processing

We will introduce two examples of coating films, in which migration of particular chemical species toward the interface or the opposite surface takes place. Takami [213] and
Watanabe [87] prepared a photocatalyst film consisting of intermediate layer of organic—inorganic hybrid and TiO$_2$—SiO$_2$ top layer with photocatalytic property on polymer substrates of polyethylene terephthalate. The intermediate layer was prepared by applying a mixture of acetone solution of polymethylmethacrylate and hydrolyzed Si(OCH$_3$)$_4$—C$_2$H$_5$OH—H$_2$O—HCl solution onto the polymer substrate and drying at 70 °C for 10 h. The chemical analysis by X-ray photoelectron spectroscopy showed that the intermediate layer has compositional gradient in the direction of depth, that is, inorganic SiO$_2$ component is concentrated in the vicinity of the interface with the inorganic layer and organic polymer components are concentrated in the vicinity of the interface with the polymer substrate. This change of distribution of chemical species is explained by the adsorption of polymers in the intermediate layer to the polymer substrate. This phenomenon is beneficial for the photocatalyst film, because organic species is apart from the top photocatalyst layer and, so, is not decomposed by TiO$_2$.

In the course of the study of organic—inorganic hybrid coating prepared from solutions containing phenyltriethoxysilane and tetraethoxysilane on polycarbonate substrate, Mizuta [214] obtained the result which suggests that concentration of phenyl groups might be higher in the vicinity of the interface between the film and the polycarbonate substrate rather than in the vicinity of top of the free surface, which promotes the adhesion of the film to the substrate. It is noted that in both cases the migration of the particular chemical species is beneficial to formation and properties of the coating film. This may happen to be the case. However, it should be remembered that such phenomena may not be limited to the above two examples.

5.3. Melting Gels

The so-called melting gels [215] are rigid at room temperature, and soften and resoften at temperatures around 110 °C. However, melting gels have a consolidation temperature higher than around 150 °C, above which they no longer resoften. Typically, nonporous transparent melting gels are prepared by a sol—gel method using hydrolysis—polycondensation of a combination of monosubstituted and di-substituted siloxanes. Accordingly, these gels are organic—inorganic hybrid.

Methyl-substituted melting gels [216] were prepared, using, for example, methyltriethoxysilane (MTES) and dimethyldiethoxysilane (DMDES). The resultant gels showed the consolidation temperature of 135—160 °C. Phenyl-substituted melting gels were prepared by Jitianu [215], using phenyltriethoxysilane (PhTES) and diphenyldiethoxysilane (DPPhDES), for instance. The resultant gels showed the consolidation temperature of 135—150 °C. Melting gels of the polyphenylsiloxanes (PhSiO$_3$)$_n$ composition are shown [217,218] to be prepared from the solution containing C$_2$H$_5$OH, H$_2$O, HCl or NH$_3$, and phenyltriethoxysilane (PhSi(OEt)$_3$). Poly(benzylsiloxanes) melting gels are prepared [219] from benzyltriethoxysilane.

Coating film is made by pouring the fluid gel onto the substrate of silicon wafer and polycarbonate and heat-treating it at consolidation temperature. Thick films, thicker than 1 mm, can be prepared without crack formation. This is because the presence of substituted methyl groups lessens the strains that often produce cracks in sol—gel materials [214,216]. The films are hydrophobic and water repellent, because methyl groups are concentrated at the surface of the films. Based on the above properties, melting gels can be applied for hermetic barrier coatings, sealing of microelectronics [220], anti-adhesion coatings, and fine patterning by stamping [221].

5.4. Sol—Gel Synthesis in Ionic Liquid

Ionic liquid is defined as salt which is in the liquid state at ambient temperatures or below 100 °C. The melting temperature of a salt becomes low and the salt becomes ionic liquid when the constituent ions are replaced by big organic ions. Since stable liquids were discovered in the early 1990s, extensive research works have been developed as novel electrolyte materials. The cations of ionic liquids are classified into imidazolium and pyridinium ions and anions are classified into tetrafluoroborate and halide ions. Applications of ionic liquids are found in the fields of extraction and separation processes, synthetic chemistry, catalysis, and electrochemistry.

According to [194,222], ionic liquids have been applied to sol—gel synthesis of SiO$_2$ materials with novel texture and structure since early 2004. Ionic liquids, such as 1-butyl-3-methylimidazolium tetrafluoroborate (BuMeImBF$_4$) and 1-hexa-decyl-3-methylimidazolium chloride (HedeMeImCl), enable room-temperature synthesis of monolithic mesoporous silica (pore size >2 nm) and superporous silica with a lamellar texture (pore size 1—2 nm). Ionic liquids are efficient solvents for the synthesis of ambient pressuredried aerogels, which have a specific pore volume similar to supercritically dried aerogels. It is interesting to note that ionic liquids can act as a catalyst for both hydrolysis and polycondensation, because their anions are good Lewis base and their cations work as Brønsted acid. This means that hydrolysis and polycondensation of silica gel precursors such as tetramethoxy silane (TMOS) and methyltrimethoxysilane (MTMS) can be made. Karout [223] prepared aerogels from ionic liquid—SiO$_2$ precursor (tetrathoxysilane, methyltrimethoxysilane)—water solution to compare the effect of the kind of ionic liquid. CO$_2$ supercritical drying of gels was applied. The ionic liquids used were 1-butyl-3-methylimidazolium
tetrafluoroborate \( [(BuMeIm^+)BF_4^-] \), N-butyl-3-methylimidazolium tetrafluoroborate \( [(BuMePy)^+BF_4^-] \), 1-hexyl-3-methylimidazolium chloride \( [(HeMeIm^+)Cl^-] \), and 1-butyl-3-methylimidazolium chloride \( [(BuMeIm^+)Cl^-] \). The resulting materials presented a new type of mesoporous aerogel textures, indicating that ionic liquids can be used as template to produce a regular array of pores in the mesoporous range.

Donati [224] prepared silica xerogels by the sol–gel technique using tetraethoxysilane as precursor and hydrofluoric acid as catalyst, together with an imidazolium as ionic liquid. Three different anions of the ionic liquids led to different silica products of a compact lamellar monolith, a free-flowing powder of aggregated spheric particles, and solid aggregates with honeycomb structure.

In view of the micromorphology and forms of silica products described above, processing with ionic liquids may be applied to formation of catalyst, catalyst support, chromatographic material, immobilizer, optics and electronics, and template for the formation of advanced materials.

### 5.5. Sol–Gel Nanotubes

Carbon nanotubes have attracted much attention owing to their versatility in the field of advanced technology. In the area of inorganic materials, imogolites that are naturally occurring aluminosilicate hydrate nanotubes are familiar and their artificial synthesis is developed for practical application. Imogolites are nanotubes which have the diameter of 1.8–2.2 nm and the length of several to several tens of nm. Triggered by these nanotubes, many other nanotubes have been prepared and some of them have been prepared by sol–gel technique. Table 12 shows examples of inorganic nanotubes prepared by sol–gel method.

Silica nanotubes with coiled channels in the walls were prepared [225], using self-assemblies of a gelator L-PHePyBr as template from the solution tetraethoxysilane, ammonia, and hydrochloric acid. It is noted that the structure of silica nanotubes obtained is not simple but the wall has a helical structure. TiO\(_2\) nanotubes were prepared by Kasuga [226]. TiO\(_2\) nanotubes are made by treating sol–gel-derived fine TiO\(_2\)–SiO\(_2\) powders of the size less than 45 μm prepared with 5–10 M NaOH aqueous solution for 20 h at 110 °C. It was shown that TiO\(_2\) nanotubes have high photocatalytic activity due to their large surface area.

B-doped TiO\(_2\) nanotubes [227] prepared by the combination of sol–gel process with hydrothermal treatment showed significantly enhanced photocatalytic activity compared with pure TiO\(_2\) nanotubes.

ZnO hexagonal microtubes were fabricated [228] by microemulsion-based hydrothermal method using aqueous reverse micelles as template. Vanadium oxide nanotubes were prepared [229] by mixing hexadecylamine with V\(_2\)O\(_5\)·nH\(_2\)O gels followed by hydrothermal treatment.

Lead zirconate nanotubes were prepared [230] using anodic porous alumina template. The nanotubes obtained have heights of 8–30 μm and average pore diameters in the range 110–200 nm. DC electrophoresis served in growing PbZrO\(_3\) nanotubes. SrBi\(_2\)Ta\(_2\)O\(_9\) (SBT) nanotubes were fabricated [158] by dipping the anodic porous alumina template into an SBT sol and calcining at 700 °C. Homogeneous SBT sol was synthesized using ethoxytantalum, strontium acetate, and bismuth subnitrate as starting materials, ethylene glycol as solvent, and glacial acetic acid as catalyst. Resultant nanotubes had smooth morphologies and well-defined diameters corresponding to the diameter of the applied template.

Nanotubes of WS\(_2\) were prepared by Li [231]. They were obtained by heating (800 °C in Ar) tungsten sulfide composites prepared based on template self-assembly of WS\(_2\)\(^{2-}\) and cationic surfactants in solution.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Diameter</th>
<th>Length</th>
<th>Remark</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO(_2)</td>
<td>25 nm</td>
<td>0.2 μm</td>
<td>Gelator template</td>
<td>[225]</td>
</tr>
<tr>
<td>TiO(_2)</td>
<td>8 nm</td>
<td>0.1 μm</td>
<td>Anatase phase</td>
<td>[226]</td>
</tr>
<tr>
<td>TiO(_2) : B</td>
<td>30 nm</td>
<td>&gt;0.5 μm</td>
<td>Anatase phase</td>
<td>[227]</td>
</tr>
<tr>
<td>ZnO</td>
<td>500 nm</td>
<td>1 μm</td>
<td>Hexagonal microtube</td>
<td>[228]</td>
</tr>
<tr>
<td>V(_2)O(_5)</td>
<td>50–130 nm</td>
<td>1–10 μm</td>
<td>Hydrothermal treatment</td>
<td>[229]</td>
</tr>
<tr>
<td>PbZrO(_3)</td>
<td>150 nm</td>
<td>&gt;5 μm</td>
<td>Anodic alumina template</td>
<td>[230]</td>
</tr>
<tr>
<td>SrBi(_2)Ta(_2)O(_9)</td>
<td>250 nm</td>
<td>&gt;1 μm</td>
<td>Anodic alumina template</td>
<td>[158]</td>
</tr>
<tr>
<td>WS(_2)</td>
<td>5–37.5 nm</td>
<td>0.2–0.5 μm</td>
<td>Template self-assembly</td>
<td>[231]</td>
</tr>
</tbody>
</table>
5.6. Fuel Cells

Fuel cells attract much attention as environmentally-friendly power source replacing fossil fuels such as petroleum and natural gas. Oxide fuel cells (OFCs) using oxide ion-conducting zirconia as electrolyte work at high temperatures of 800–1000 °C and are employed for large-scale power generation, while proton-exchange membrane fuel cells (PEMFCs) work at low temperatures below or just above 100 °C, being aimed at the use in automobiles and houses. Sol–gel technology finds the application to both types of fuel cells [232]. At present, however, great efforts are directed to improvement of the electrolyte of PEMFC in the field of sol–gel processing.

A PEMFC consists of an electrolyte membrane and carbon paper electrodes contacting its both sides. At the anode, hydrogen molecules are oxidized into protons and the electrolyte membrane: 

\[ \text{H}_2 \rightarrow 2\text{H}^+ + 2e^- \]  

at anode (28)

Electrons are taken out of the anode as electric current. Protons diffuse through the membrane, reaching the cathode, and react with oxygen molecules and electrons supplied via the cathode on the catalyst, being reduced to water:

\[ 2\text{H}^+ + 1/2\text{O}_2 + 2e^- \rightarrow \text{H}_2\text{O} \]  

at cathode (29)

Commercially available PEMFCs use ion-exchange membrane of fluororesin represented by Naftion® which shows high proton conductivity at low temperatures near room temperature, but at temperatures higher than 80 °C the conductivity and chemical durability of the film decrease. The problem with conventional PEMFCs which are operated at temperatures lower than 80 °C is CO poisoning of Pt catalyst. In order to improve conversion efficiency and avoid CO poisoning of catalyst, PEFMC has to be operated at higher temperatures of 130–150 °C, retaining some water for high proton conductivity. Porous inorganic or organic–inorganic hybrid membranes prepared by sol–gel processing are assumed to satisfy these conditions.

Tadanaga [233] prepared proton conductive inorganic–organic hybrid films, which show the ionic conductivity of about 1 \times 10^{-2} \text{ S cm}^{-1} under 25% relative humidity at 130 °C, using epoxycyclohexylethyltrimethoxysilane and orthophosphoric acid. A PEMFC fuel cell using these films shows a power of about 40 mW cm⁻² with current density of 110 mA m⁻². Daiko [234] prepared highly conductive P₂O₅–SiO₂ gels by sol–gel method for PEMFC. Xiong [235] prepared highly conductive P₂O₅–SiO₂ gels with ordered mesopores of 5–7 nm using template. Very interesting electrolytes which show high ion conductivity without water were fabricated [236]. They are P₂O₅–SiO₂ gels prepared by the use of ionic liquid of the imidazolium system. Their conductivity at 250 °C is 1 mS/cm in the order of magnitude, which is shown without any humidity.

6. CONCLUDING REMARKS

The sol–gel method supports advanced technologies by supplying advanced nanomaterials. It was shown that sol–gel processing not only creates novel materials but also supplies better methods of fabricating novel materials explored by conventional sintering and sputtering.

REFERENCES

Chapter | 11.1.2 Sol–Gel Process and Applications


Chapter | 11.1.2 Sol–Gel Process and Applications


Chapter 11.1.3

Colloidal Processing Fundamentals

David McKinney and Wolfgang Sigmund

1 MM Virtuoso, Inc., Gainesville, Florida 32606, United States, 225 Rhines Hall, Department of Materials Science and Engineering, University of Florida, P.O. Box 116400, Gainesville, Florida 32611-6400, United States

Chapter Outline
1. Introduction of Powder Processing 911
2. Colloidal Processing Concepts and Theories 911
3. Rheological Properties of Ceramic Slurries 915
4. Processing Additives 918
5. Forming Techniques 921
6. Summary and Outlook 925
References 925

1. INTRODUCTION OF POWDER PROCESSING

Ceramics have prohibitively high melting temperatures, far exceeding most other materials. Hence, they typically cannot be processed via melts such as metals and polymers. Instead, 90% of ceramic products begin in powder form. To get to the finished product, practitioners process the powder through several steps. Put simply, ceramic powders are added to liquids, dispersed, formed or cast, dried, and sintered. Advanced ceramics require optimization throughout the entire process. As in baking bread, the final product’s integrity and quality depend on precise execution of process steps. This is vitally important, because with knowledge and understanding of ceramic processing, we can identify the exact step where something went wrong and fix it the next time.

Here, we present colloid processing fundamentals, rheological properties of ceramic slurries, information on processing additives, and forming techniques. As we lay out the background science of ceramic processing, we also present general applications for key formulas and techniques. Moreover, we define and establish exact meanings for words common to several disciplines, but which mean different things to scientists in different fields. We hope that interdisciplinary researchers adopt these and standardize their use.

2. COLLOIDAL PROCESSING CONCEPTS AND THEORIES

Nearly all ceramics begin in powder form, and today manufacturers can produce them nearly infinitely fine. Indeed, advanced ceramic powders contain submicron- and even nanometer-sized particles. Usually, the first processing step is to add the powder to a liquid to form slurry. Chemically, slurries are akin to mixtures in that the particles do not dissolve into the liquid. Physically, the ceramic particles are ideally dispersed evenly throughout the liquid. In such systems, particles and liquid remain separate entities. By definition, small particles dispersed in liquid form either suspensions or colloids. Suspensions contain particles larger than 1 μm and colloids contain particles 1 nm to 1 μm in size. Also, technically, particles in a suspension can settle and those in a colloid should not.1

Understanding colloid and surface science is important to ceramic processing because this provides the key to optimizing a ceramic product’s characteristics and controlling for flaws later on in the process. The term colloid stems from the Greek word kolla, meaning glue.

1. One often sees the terms ceramic suspension used synonymously with ceramic colloid. For precision’s sake, we will be consistent and use ceramic colloid throughout.
When particles are smaller than 100 μm in diameter, surface forces govern their behavior. Unlike gravitational forces, surface forces make particles stick together, hence the term colloid. In ceramic processing, these sticky surface forces must be controlled.²

**Stability** is the most desired property in a ceramic colloid, and interparticle forces play an important role in stabilization [1]. Tiny particles dispersed in a liquid medium spontaneously clump together due to attractive surface forces. Thus, particle surfaces must be modified to produce repulsive potentials that counteract attractive potentials. Below we introduce attractive and repulsive surface forces that ceramic processors use to optimize ceramic colloids.

### 2.1. Attractive Surface Forces

Dutch scientist and Nobel prize winner Johannes Diderik van der Waals (1837–1923) discovered attractive forces between atoms and molecules in the 1870s. He made changes to Boyle’s ideal gas law to make it work for real gases with polar behavior, e.g. H₂O, over a broader temperature range and first proposed attractive intermolecular forces. Since then, several physicists have explained their origins.

Typically, scientists discuss three types of **van der Waals forces**, all electromagnetic and dipole interactive: Keesom forces, Debye forces, and London or dispersion forces. The first two require at least one permanent dipole, while the third and strongest depends only on transient dipoles inducing dipoles in neighboring atoms or molecules. Another Dutch scientist, Hugo Christiaan Hamaker (1905–1993), described these energies more quantitatively in 1937. He showed that the energy of van der Waals attraction $E_{vdW}$ equals a materials constant $A$ times distance $d$ and body geometry-dependent function $G(d)$:

$$ E_{vdW} = A \times G(d) $$  (1)

We now know $A$ as **Hamaker’s constant** [2], and though this material-specific constant is fundamental, theoretical and experimental access to Hamaker constants still proves difficult even today. As such, we know Hamaker constants for a limited number of materials only. For more detailed information on Hamaker constants and van der Waals forces in ceramic processing, refer to two excellent review papers by Bergström and French [3,4].

### 2.2. Repulsive Surface Forces

As mentioned earlier, stability is the most desired property in a liquid ceramic colloid. To achieve stabilization, one needs to create mutually repulsive potentials on particle surfaces to counterbalance attractive van der Waals potentials. There are four ways to do this: charge ions on particle surfaces (electric double-layer stabilization), adsorb polymer chains onto particles (steric stabilization), combine the two mechanisms (electrosteric stabilization), or add non-adsorbing nanosize objects to the dispersing medium (depletion stabilization).

#### 2.2.1. Electric Double-Layer Forces

A colloidal particle’s surface charge forms the basis of electric double-layer (EDL) stabilization. Surface charge magnitude determines repulsion magnitude between particles, thereby influencing the stability of the colloid. A perfect stoichiometric crystal is neutral, but as soon as the particle surface oxidizes and hydrolyzes, hydroxyl and other groups form that develop charges when immersed in protic liquid. Surface group ionization or dissolution, specific ion adsorption, and ion exchange on the surface are major sources of these charges for any ceramic [5].

After oxidation, counterions (oppositely charged ions) in the electrolyte containing liquid attract to a particle’s surface charges. This forms a fixed, or Stern, layer on the particle surface and adds a diffuse second layer in solution. The Debye length Eqn (2) describes the ion double layer’s (EDL’s) dimension, thickness. For a particle surrounded by a diffuse double layer in a shear field, shear slippage occurs at some distance away from the surface, which is where electric potential, or zeta potential, can be measured. Zeta potential is believed to be close to the Stern layer’s potential, and when zeta potential is zero, the corresponding pH level is called the point of zero charge (PZC), or iso-electric point (IEP) (Figure 1).

**Point of zero charge** refers to having neither negative nor positive charges on the surface. IEP applies to complex ceramic surfaces, and the sum of negative and positive charges yields net zero charge. This means locally charged regions are still available at the IEP. Moreover, EDL-stabilized ceramic colloids become unstable at the PZC or IEP. Usually, this causes rapid coagulation; however, some systems, like silica in aqueous media $\phi_0 < 50mV$, may require days for coagulation to happen at the IEP. We discuss rates of coagulation later in this chapter. Table 1 gives PZCs and IEPs of various compounds.

Theoretically, the relationship between surface charge density and a dispersed particle’s surface potential is given by [6]

$$ \sigma_0 = (8\pi e_0 R T c)^{1/2} \sin \left( \frac{zF \phi_0}{2RT} \right) $$  (2)

where $\varepsilon$ is the medium’s dielectric constant, $\varepsilon_0$ the permittivity of free space, $R$ the gas constant, $F$ the Faraday constant, $c$ the electrolyte ions’ concentration, and $z$ the

---

² Force means different things to physicists, surface scientists, and ceramists. To be clear, this chapter deals with variations in electromagnetic force only, i.e. longer-range forces that are not gravitational.
Another important relationship to note is that ion surface concentration positively correlates to repulsive pressure $P$ (i.e. ion surface concentration is proportional to repulsive pressure $P$). Therefore, distances between particles become very small, and both ion surface concentration and repulsive pressure approach infinity with high solids loading [7,8]:

$$P = -\frac{p_skT}{D}$$  \hfill (5)  

However, this infinite pressure assumes that surface charge remains constant. This does not actually occur. When repulsive pressure $P$ forces two particles close to each other, counterions adsorb onto their original sites. Hence, as $D$ approaches 0, $\sigma$ also falls, and surface potential remains constant. This effect is known as charge regulation. Furthermore, if no binding occurs, surface charge density remains constant, and counterion concentration approaches a constant value $2\sigma/eD$ at the smallest possible separation distance.

In summary:

- constant charge EDL repulsion is always higher than constant potential EDL repulsion.
- real surfaces show intermediate behavior between constant charge and constant potential [8].

### 2.2.2. Repulsion between Double Layers

Two Russian scientists, Derjaguin and Landau, and two Dutch scientists, Verwey and Overbeek, simultaneously calculated the total forces acting between two particles in the 1940s; hence, the theory of interaction between electric double layers is referred to as DLVO theory. DLVO theory states that the sum of attractive potential energy $V_A$ and repulsive potential energy $V_R$ gives total interaction energy $V_T$ (Figure 2):

$$V_T = V_A + V_R$$  \hfill (7)  

Osmosis causes EDLs to repel particles and overcome attractive van der Waals forces. When two particles with like charges approach one another, their diffuse double layers overlap. This ionic cloud overlap causes local increase in ionic strength, which depletes liquid in the region, yielding an osmotic pressure. Finally, liquid diffuses into the region to equilibrate the concentration, pushing particles apart. If repulsive potential is large enough to overcome attractive van der Waals potential, the colloid has EDL stabilization (Figure 3).

### 2.2.3. Steric and Electrosteric Forces

Sometimes EDLs do not create a large-enough repulsive potential to overcome van der Waals potential. In such
cases, other techniques can be used to stabilize the ceramic colloid: steric stabilization and electrosteric stabilization. Steric forces depend on added polymers (organic macromolecules). They adsorb onto ceramic particles’ surfaces. When two particles with adsorbed polymers approach each other, their layers overlap and penetrate. As they cannot move about freely, conformational entropy decreases. The particles, essentially, cannot come closer than roughly twice the layer thickness. If this distance exceeds the range of van der Waals forces, particles will not stick.

Furthermore, a combination of electrostatic and steric forces can stabilize colloids — electrosteric stabilization. Some polymer molecules have groups that ionize in polar solvents to create polyelectrolytes (charged polymers). Solvent conditions decide the fraction of dissociated functional groups \( \propto \), and as \( \propto \) increases, the polymer charge can vary from zero to highly positive or negative. Figure 4 shows a schematic of the dissociation of poly acrylic acid (PAA) segments. In such cases, repulsion is greater due to the strong long-range EDL repulsion between charged molecules. If the particle itself is charged, it produces long-range repulsive potential when particles approach one another; if the polymer is charged, it produces an additional short-range repulsive potential (Figure 5).

2.2.4. Depletion Stabilization

Lastly, osmotic repulsion concepts described above for ions in solution or adsorbed polymers apply to dispersed nanoparticles or nonadsorbed, dissolved polymers. For depletion stabilization to work, nonadsorbed nano-objects have to fit in the space between two particles. For recent advances, refer to Rahaman’s Ceramic Processing [9].

2.3. Coagulation Kinetics

In the preceding section, we described forces that contribute to colloidal stability. This section covers the
kinetics of unstable colloids. We repeat, stability is the most desired property, yet most colloids are thermally unstable. Particles remain dispersed only when the potential barrier is $kT$. If the potential barrier is $\ll kT$, particles stick. Thus, it is important to understand a colloid’s rate of coagulation. Technically, a reversible adhesion is called coagulation; by definition, a coagulate forms when particles in an aqueous slurry adhere to each other due to van der Waals attraction. If this association is irreversible, the effect is called flocculation.

Colloid particles do not exhibit prescribed kinetic energy. At any given point in time, some particles may have higher energy than others. Napper and Hunter extensively studied the effect of DLVO theory on kinetic behavior of colloidal particles [10]. Stability ratio $W$ is given by

$$ W = \frac{\text{Number of collisions between particles}}{\text{Number of collisions leading to coagulation}} \quad (8) $$

The rate of slow coagulation is given by

$$ R_s = \frac{R_f}{W} \quad (9) $$

where $R_s$ is the rate of slow coagulation (when $VT \gg kT$) and $R_f$ the rate of fast coagulation (when $VT \ll kT$).

### 2.3.1. Rate of Fast Coagulation

Smoluchowski addressed the rate of coagulation without any repulsive force, i.e. fast coagulation [11]. He calculated a particle’s flux when it comes in contact with a fixed particle, and the rate of fast coagulation is given by

$$ R_f = 8\pi Da v_0^2 \quad (10) $$

where $D = kT/6\pi\eta a^2$ (m$^2$/s) (assuming uniform spheres), $a$ is the particle radius (m), $v_0$ is the initial particle concentration (number/m$^3$), and $\eta$ is the viscosity of dispersing medium (Pas).

Time for bulk concentration of particles to reduce by half is given by

$$ t^{1/2} = \frac{3\eta}{4kT v_0} \quad (11) $$

Thus, smaller particles have faster coagulation rates than larger particles.

### 2.3.2. Rate of Slow Coagulation

With a repulsive potential energy barrier $\gg kT$, the rate of coagulation becomes orders of magnitude smaller than in Eqn (10), i.e. slow coagulation. Consider the energy barrier as the activation energy needed for a coagulation event to take place. In this case, the equation for the rate of slow coagulation becomes

$$ R_s = \frac{4\pi Da v_0}{\int_{2a}^{\infty} \exp \left( \frac{VT}{r^2} \right) \, dr} \quad (12) $$

where $VT$ is the magnitude of the potential energy barrier and $r$ the distance from the particle center.

Substituting the values of $R_s$ and $R_f$ in Eqn (9), the value of stability ratio becomes

$$ W = \frac{R_f}{R_s} = 2a \int_{2a}^{\infty} \exp \left( \frac{VT}{r^2} \right) \, dr \quad (13) $$

This shows that the stability ratio is proportional to the particle radius, indicating that larger particles have a stronger tendency to coagulate.

### 3. RHEOLOGICAL PROPERTIES OF CERAMIC SLURRIES

Rheology is the science of deformation and flow of matter [8]. Practitioners study it to help characterize and optimize particle dispersion, especially in high solids loading and
low-viscosity colloids, which describes ideal ceramic colloids. **Solids loading** refers to volume percent of particles in the slurry. Viscosity describes flow behavior of liquids and slurries, suspensions, and colloids. As solids loading increases, the viscosity also increases.

When a fluid’s viscosity is independent of the shear rate, the fluid exhibits Newtonian behavior (e.g. water). However, colloids are more often non-Newtonian. This formula applies to viscous fluids in which viscosity changes with shear rate at a constant temperature. Viscosity can be defined as

\[ \eta = \frac{\tau}{\dot{\gamma}} \]  

where \( \tau \) is the shear stress, \( \dot{\gamma} \) the shear rate, and \( \eta \) the shear viscosity.

If viscosity and shear rate positively correlate, a fluid is shear thickening. Colloids exhibit this behavior at very high solids loadings. If viscosity and shear rate negatively correlate, a fluid is shear thinning, or pseudoplastic. Finally, other curves are formed when the particle network creates internal stress (Figure 6).

Two major approaches attempt to characterize behavior of colloid systems: hard sphere and soft sphere. The hard-sphere approach is simplistic and assumes the sole interaction among particles is repulsion. The soft-sphere approach assumes a potential with a barrier, a primary minimum like in DLVO theory, and applies better to real systems.

3.1. Hard-Sphere System and Solids Loading-Dependent Viscosity

For dispersions of noninteracting particles, viscosity depends on solids loading. Increased solids loading increases viscosity. Einstein proposed a method to describe this phenomenon. It is given by

\[ \eta = \eta_s(1 + k_1\phi) \]  

where \( \eta \) is the slurry viscosity, \( \eta_s \) the suspending medium viscosity, \( k_1 \) the Einstein coefficient (2.5 for spheres), and \( \phi \) the volume fraction of solids in slurry.

However, this equation does not account for particle size influence. Hence, it holds only when dispersed particles do not interact, e.g. dilute systems. When the system involves interaction between particles, the situation becomes more complex, and the contributions of particle interactions must also be included. If we include higher-order terms of \( \phi \), the extended Einstein equation becomes

\[ \eta = \eta_s(1 + k_1\phi + k_2\phi + \ldots) \]  

A convenient way of expressing viscosity is by dividing slurry viscosity \( \eta_s \) by suspending media viscosity \( \eta_1 \). This is a dimensionless quantity called relative viscosity \( \eta_r \).

Thus, the effect of volume fraction of solids on viscosity behavior is best studied in relation to the **maximum packing fraction**. With the addition of more and more particles, at some point they become jam-packed, resulting in a continuous, three-dimensional contact throughout the system. This immobilizes the fluid. At this point, when no flow takes place, the colloid has maximum packing fraction \( \phi_m \).

Krieger and Dougherty introduced a model for hard spheres:

\[ \eta_r = \left(1 - \frac{\phi}{\phi_m}\right)^{-[\eta]\phi_m} \]  

where \([\eta]\) is the intrinsic viscosity. However, this model applies only to monosize spheres and does not consider particle-size distribution of ceramic powders.

---

3. Authors often synonymously use concentration, but this term technically refers to solution molarity.

4. Sigmund et al. illustrate similar curves for viscoelastic fluids [4].

5. Green body refers to a ceramic product after forming but before drying.
3.2. Effects of Particle-Size Distributions

Most commercial ceramic powders have a continuous particle-size distribution. The effect of particle-size distribution on viscosity becomes pronounced at very high solids loading, close to the colloid’s maximum packing fraction. Typically, a polydisperse system’s viscosity is lower than a monodisperse system’s at identical volume fraction, as interstitials between larger particles can be occupied by smaller particles [15]. Also, as the Krieger–Dougherty Eqn (17) suggests, viscosity of monodisperse suspensions reaches a very high value at low solids loading.

For efficient processing, however, high solids loadings are preferable. Colloids consisting of mostly smaller particles exhibit a lower maximum solids loading, thixotropy, or plasticity. This may be caused by the increase in specific surface area binding the dispersing medium. Colloids with a higher percentage of large particles tend to be dilatant. Dilatancy limits processing and, thus, maximum solids loading. If medium-size particles dominate, distribution resembles more and more a monodisperse system. Thus, limited numbers of particles can be packed into the colloid with a limit of about 64 vol% for monosize randomly packed spheres [16].

3.3. pH Variations of Viscosity

A colloid’s pH significantly affects the system’s viscosity. In the case of oxides, the viscosity also negatively correlates with absolute value of the zeta potential [12]. Viscosity of alumina colloids increases with increases in pH, up to a certain value. Beyond a certain pH, viscosity becomes inversely proportional to pH. The pH at which viscosity reaches maximum value corresponds with the colloid’s IEP (Figure 7).

3.4. Steric System Rheology

In cases of dispersions stabilized by the addition of polymers, their rheology depends on three factors:

- Flory–Huggins parameter $\chi$. If this parameter is $<0.5$, then repulsion between particles will be strong enough for them to remain dispersed.
- Thickness of the polymer’s layer adsorbed on the particles. The adsorbed layer’s thickness should be large enough to overcome attractive van der Waals forces and prevent particles from sticking. Overlapping of polymer layers is crucial, which is the prime cause of repulsion.
- Ratio of particle diameter to adsorbed layer thickness [16]. The thickness of the adsorbed layer decreases as particle size increases. Hence, dispersions of small particles, due to large thickness of adsorbed dispersant layers, tend to act like soft spheres.

3.5. Electrosteric System Rheology

Polyelectrolytes are widely used as dispersants for high solids loaded colloids (>50 vol%). They combine principles of EDL and steric stabilization, or electrosteric stabilization, and they depend on pH and ionic strength, as Naito et al. discuss [17]. At low solids loading (~20 vol%), viscosity is relatively low, and it is affected very little by pH changes. As solids loading increases, however, pH affects viscosity significantly. The amount of added polyelectrolyte also has a profound effect on colloidal rheology. It should be optimized to just saturate the surface. Additional polyelectrolytes result in excess amounts of polymer in the system, and excess polymer can cause depletion flocculation in high solids loaded systems.

Conformation of adsorbed polyelectrolyte also plays an important role in rheological behavior of electrosterically stabilized colloids, and, in turn, polyelectrolyte conformation depends on the system’s pH. A detailed study of adsorption behavior on Al$_2$O$_3$ shows that polyelectrolyte adsorption on particles increases as pH decreases. Typically, a 10-fold increase of adsorbed amount is observed from the uncharged to the charged state [18,19]. When pH increases or decreases beyond zero charge, the fraction of the polyelectrolyte dissociated moves toward 1. Hence, charges in the polyelectrolyte repel each other and the molecule stretches. At this moment, two models exist: the charged polyelectrolyte adsorbs flat on the surface or the polyelectrolyte adsorbs in a tail-like brush structure.
Conformation shape of the adsorbed polyelectrolyte highly influences dispersion quality. Which types of structures — flat, pancake-like, or brush-like — are achieved depends on adsorption conditions and the materials involved. For pancake-like adsorption, the polymer only contributes short-range repulsive force, and EDL forces of the charged polyelectrolyte mainly contribute to stabilization via long-range interactions. For brush-style structures, the repulsion is much stronger, and true electrosteric contributions are present. Polyelectrolytes can also be used as dispersants when they are uncharged, i.e. at their PZC. However, they will favor coil-like conformations. Hence, much higher molecular weights will be needed to achieve thicker layers of adsorbed polymer coils, and steric forces predominantly contribute to stabilization [1].

4. PROCESSING ADDITIVES

Almost all powders start agglomerated in the process. To overcome attractive van der Waals forces, additives must be used to disperse them. Additives come in a wide variety and impact the particle-dispersing medium or particle–particle interfaces. Both organic and inorganic additives can be used to improve the properties of final ceramic green bodies and sintered products [1]. For example, ceramists use polymeric additives as they provide superior mechanical properties for handling. However, at the same time, such compounds introduce specific problems related to their size. In one such case, the dry powder direct compaction process, ceramists add organic binders to control packing uniformity as fine ceramic raw powders are processed into granules. Each granule should consist of hundreds of homogeneous distributions of particles and binder. However, in reality, granules might be hollow or show strong gradients of binder and fine particles on the surface. Additionally, when dispersing ceramic powders in solvent, dispersants make a high solids loaded colloid stable by adjusting interparticle forces. For certain casting processes, e.g. tape casting or injection molding, selection of suitable additives is vital. The following section describes commonly used additives.

4.1. Dispersing Media/Solvents

Dispersing medium is a liquid that determines the slurry’s flow characteristics. It also acts as solvent for organic additives. However, it sometimes behaves favorably by dissolving the ceramic powders also. A good dispersing medium must:

- dissolve organic additives,
- have low viscosity,
- not foam,
- have a low boiling point,
- not react with the powder, and
- be nonvolatile and nontoxic.

The right dispersing medium for slurry has to meet all these characteristics for the given powder [20,21]. For example, either water or ethanol is adequate for an alumina system, whereas a BaTiO3 system may need a nonaqueous mixture of trichloroethylene ethanol to avoid dissolution of Ba2+ ions in water. Generally, blends of alcohol and organic solvent, based on powder and dispersant colloidal chemistry, make up the system. This allows other additives to dissolve, and it keeps viscosity low during milling.

It is important to note that dispersing medium choice defines both attractive and repulsive forces. High-dielectric versus low-dielectric solvents considerably affect van der Waals interactions [1,22]. Some media allow steric stabilization only, while others allow any choice, including combinations. Moreover, if capillary forces are elected for compaction of the particle system, then the dispersing medium’s properties for capillarity are utmost important [1,23–25].

4.2. Dispersants

Dispersants, molecules or polymer chains of typically low-molecular-weight organic chemicals, absorb onto ceramic particles. Dispersants carry functional groups for anchoring on surfaces and have buoy moieties. The buoy performs best if additional highly polar groups are available, especially carboxylic acid groups, as these develop negative surface charges in protic dispersing media above pH 3. Typical examples include citric or maleic acids. Especially useful in tape-casting processes (covered later in this chapter), dispersants keep a slurry uniform and homogeneous. Equal dispersion helps provide uniform packing in a green body, and it yields uniform shrinkage during drying, binder burnout, and sintering for equiaxed particles. Complete surface coverage of powder with dispersant optimizes dispersion. This avoids bridging and achieves uniform surface properties defined by the dispersant. Complete surface coverage does not require large amounts; for submicron powders, typically less than 1 wt% suffices.

As with other additives, dispersants are system specific; however, Menhaden fish oil can be used for a variety of ceramic materials in nonaqueous dispersing media. It is a highly purified, nondrying oil derived from Brevoortia tyrannus, a herring-like fish found off United States’ Eastern coast. When auto-oxidized it transforms into a useful dispersant for ceramic processing. Auto-oxidation causes a poorly controlled polymerization of highly unsaturated triglycerides. The process is stopped at a certain viscosity. Ceramists, technically, describe Menhaden fish oil as a polydisperse polymeric mixture (dispersity above 40,
whereas standard synthetic polymers are around 2) having a molecular weight ($M_w$) above 100,000. Unfortunately, even after oxidizing, the oil still contains fatty acids, non-reacted triglycerides, and an intermediate molecular weight fraction. This causes multitudinous problems due to the nonreproducibility of the oil and its colloidal stabilizing ability. Also, though not endangered, Menhaden is over-fished and populations dwindle sizably each year. Synthetic substitutes are now available.

Other organic dispersants include polymers and polyelectrolytes with $M_w$’s from 2,000 for highly charged polyelectrolytes and 100,000 for polymers. Particle size and dispersing medium (main parameters controlling van der Waals forces) determine the size and required charge of a successful dispersant. In general, the repulsive barrier must be $\geq -1.5 \kappa T$ to have enough Brownian energy to move particles out of attraction and achieve free flow.

Typical polar or dissociable functional groups for dispersant include hydroxyl ($-\text{OH}$), carboxyl ($-\text{COOH}$), sulfonate ($-\text{SO}_3^-$), sulfate ($-\text{OSO}_3^-$), ammonium($-\text{NH}_4^+$), amino ($-\text{HN}_2$), imino ($-\text{NH}-$), and polyoxyethylene ($-\text{CH}_2\text{CH}_2\text{O}-$) groups. Such nonionic macromolecules in solid/solvent systems mainly contribute to steric stabilization [25]. Table 2 lists commonly used dispersants.

A ceramic material’s solubility also strongly influences a dispersant’s effect. Solubility is specific and covers many orders of magnitude. It depends on the material and the dispersing medium’s pH. Higher solubility increases counterions, which compress the EDL. Under these conditions, a purely EDL-stabilized system will collapse. Steric-stabilized systems will also change because solubility and polymer conformation depend on ionic strength. Polymers may salt out, i.e. coagulate. Such actions sometimes occur unwanted, or they may be especially designed, e.g. in direct casting or constant volume processes.

### Table 2: Common Dispersants Used in Ceramic Processing

<table>
<thead>
<tr>
<th>Low molecular weight</th>
<th>Large molecular weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium pyrophosphate</td>
<td>Poly(acrylic acid) (PAA)</td>
</tr>
<tr>
<td>Ammonium citrate</td>
<td>Poly(acrylic acid) (PMAA)</td>
</tr>
<tr>
<td>Sodium citrate</td>
<td>Ammonium polyacrylate</td>
</tr>
<tr>
<td>Sodium tartrate</td>
<td>Sodium polyacrylate</td>
</tr>
<tr>
<td>Sodium succinate</td>
<td>Sodium polysulfonate</td>
</tr>
<tr>
<td>Glyceryl triolate</td>
<td>Poly(ethylene imine)</td>
</tr>
<tr>
<td>Phosphate ester</td>
<td>Menhaden fish oil</td>
</tr>
<tr>
<td></td>
<td>Random copolymers</td>
</tr>
<tr>
<td></td>
<td>Comb polymers</td>
</tr>
</tbody>
</table>


### 4.3. Binders

Dry ceramic particle networks are brittle and exhibit low compression strengths, typically less than megapascals. To improve handling, ceramists add **binders** to toughen and strengthen the ceramic, but which burnout in the process. No other material group except organic polymers yields such properties. Polymer strengths derive from their size, i.e. molecular weight. Polymer solubility in the dispersing medium and slurry viscosity limit the polymer’s size; hence, binders are polymers with typically $M_w \sim 100,000$.

In tape-casting processes, a binder adds strength to the dried ceramic tape. This is important especially for storing and machining dry tape [20]. Since the binder is yet another polymer in the slip, surface chemistry must be tailored and processing hierarchy maintained to achieve the final product’s desired, well-defined nanostructure. Ceramists add dispersants first, then the binder. Any other sequence yields unfavorable nanostructures on particle surfaces, which lead to high viscosities and coagulation.

The binder’s polymer properties allow the ceramic to maintain its strength and shape until firing [27]. However, binders increase slurry viscosity and change flow characteristics. Thus, binder—dispersant and other additives in the ceramic mixture must be compatible to avoid phase separation as the dispersing medium evaporates. A high molecular weight and a low glass transition temperature are advantageous. Binders should:

- form a strong, cohesive tape when dry,
- be miscible or emulsible in the solvent,
- be inexpensive, and
- readily burnout without leaving carbon residue [28].

In fine powder granulation, e.g. spray drying, binders bond hundreds of fine particles to form larger (>50 $\mu$m) granules [23]. The binder temporarily bonds the powder particles and gives the green body good handling and storage characteristics. This reduces cracks, pinholes, and other defects in the unfired and fired parts. Table 3 gives typical recipes for spray granule additives, and Table 4 lists binders commonly used in ceramic processing.

### 4.4. Plasticizers

Solvents called **plasticizers** lower the glass transition temperature of polymers. Such plasticizers have low vapor pressure and typically exhibit a high boiling point (above 250 °C). Also, they are typically low-molecular-weight organic compounds. They function to increase chain mobility of binder molecules. With an optimal concentration...
of plasticizer, intermolecular forces between polymers weaken, and interchain distances between polymers lengthen, thereby increasing intermolecular mobility.

Compatibility, efficiency, and stability determine plasticizer choice. Most are either glycols or phthalates, e.g. polyethylene glycol or dioctylphthalate [20]. Often, binders require a plasticizer to improve flexibility and workability of the dried, green ceramic tape because glass’s brittle behavior at room temperature is not conducive to later processing steps. External plasticizing is most common. In this way, binder polymers mix with low-molecular-weight plasticizers (Mw ~ 300–400) or with oligomers (Mw ~ 3000–4000) [8]. Too much plasticizer degrades the binder’s strength, which will also cause problems later during machining. Therefore, the binder/plasticizer ratio must be carefully monitored to provide optimal properties. Furthermore, the plasticizer must be emulsible in the same solvent used to process the binder. Traditionally, two different plasticizers are added to slurry, although one suffices. This tradition might have a positive impact on drying and burnout.

<table>
<thead>
<tr>
<th>Additives</th>
<th>wt.% of additives</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyvinyl alcohol (PVA)</td>
<td>83.6</td>
<td>Binder</td>
</tr>
<tr>
<td>Glycerin</td>
<td>8</td>
<td>Plasticizer</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>4</td>
<td>Plasticizer/Lubricant</td>
</tr>
<tr>
<td>Tri ammonium citrate</td>
<td>4</td>
<td>Dispersant</td>
</tr>
<tr>
<td>Ammonium stearate</td>
<td>0.2</td>
<td>Wetting agent</td>
</tr>
<tr>
<td>Octanol</td>
<td>0.2</td>
<td>Deoamer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4 Common Binders Used in Ceramic Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binders</td>
</tr>
<tr>
<td>Poly(vinyl alcohol)</td>
</tr>
<tr>
<td>Poly(acrylic acid)</td>
</tr>
<tr>
<td>Polyacrylamide</td>
</tr>
<tr>
<td>Poly(ethylene glycol)</td>
</tr>
<tr>
<td>Poly(ethylene imine)</td>
</tr>
<tr>
<td>Poly(methylacrylic acid)</td>
</tr>
<tr>
<td>Polyvinylpyrrolidone</td>
</tr>
<tr>
<td>Methylcellulose</td>
</tr>
<tr>
<td>Hydroxypropylmethylcellulose</td>
</tr>
<tr>
<td>Hydroxyethylcellulose</td>
</tr>
<tr>
<td>Sodium carboxymethylcellulose</td>
</tr>
<tr>
<td>Ethyl cellulose</td>
</tr>
<tr>
<td>Starches</td>
</tr>
<tr>
<td>Dextrins</td>
</tr>
<tr>
<td>Sodium alginate</td>
</tr>
<tr>
<td>Ammonium alginate</td>
</tr>
<tr>
<td>Poly(vinyl butyral)</td>
</tr>
<tr>
<td>Poly(vinyl formol)</td>
</tr>
<tr>
<td>Poly(methyl methacrylate)</td>
</tr>
<tr>
<td>Gum Arabic</td>
</tr>
<tr>
<td>Guar gum</td>
</tr>
</tbody>
</table>
4.5. Lubricants

Lubricants reduce friction between particles or between particles and the die wall in extrusion and injection-molding processes. Stearic acid, stearates, and various waxy substances are commonly used [6,12,20,29–31].

4.6. Other Additives

Other additives can be used to enhance a final product. Wetting agents, or homogenizers, increase emulsibility of many liquids in slurry. They also contribute to high-quality tape surfaces. Cyclohexanone is a common wetting agent. Defoaming agents, such as dimethylsilicones, prevent foaming and allow trapped air bubbles to escape easier on degassing. They commonly have to be used in aqueous slurries or with submicron powders. Defoaming agents work to remove surplus surfactants from bubble surfaces, increase bubble interfacial energies, and make bubbles unstable. However, having several organic additives makes predicting a defoaming agent’s effectiveness complex. Table 5 shows examples of organic additive combinations used in industrial pressing powders.

5. FORMING TECHNIQUES

Wet powder processes disperse ceramic powders in solvent to form a stable, high solids loaded slurry for consolidation. Drained casting, tape casting, and constant volume techniques all consolidate colloids into designed, functional shapes. Each technique differs from the other primarily by the type of mold. In all cases, after a wet green body is formed, it is molded and dried to form a dry green body. Drained casting techniques use porous molds (e.g. slip casting) to consolidate the slip, tape-casting techniques do not use a mold, and constant volume techniques use nonporous molds.

5.1. Drained Casting Techniques

Drained casting techniques rely on phenomena that separate a ceramic colloid’s liquid and solid components.
at the macroscopic level. Various techniques for liquid-phase removal have been developed, and all seek to consolidate particles into a green body of desired form with sufficient mechanical strength for mold removal and handling.

5.1.1. Slip Casting

In slip casting, the solvent drains through the mold’s porous walls. Capillary forces depend on pore size, $\alpha$, and the surface tension at the liquid—vapor interface, $\gamma_{LV}$, which cause a pressure gradient in the liquid and drive its flow:

$$ P_c = \frac{2\gamma_{LV}}{\alpha} \quad (18) $$

Often, the growth rate of the slip-cast layer perpendicular to the mold surface exhibits a parabolic decrease over time. This makes slip casting a comparably slow technique and less suitable if thick, consolidated layers are required.

5.1.2. Osmotic Consolidation

Osmotic consolidation exploits the osmotic pressure of a colloid in contact with a polymer solution via a semi-permeable membrane. Here, differences in chemical potentials force the solvent to flow through the membrane into the polymer solution. The resulting osmotic pressure can be defined as

$$ \Pi = \frac{\mu_s - \mu_p}{V_m} \quad (19) $$

where $\mu_0$ is the chemical potential of the solvent in the ceramic suspension, $\mu_p$ the chemical potential in the polymer solution, and $V_m$ the molar volume of the solvent. Osmotic pressures are significantly higher than capillary pressures in slip casting.

5.1.3. Filter Pressing

Alternatively, an external, mechanical pressure can be applied to the colloid, which drives the solvent through a filter membrane—filter pressing. Consolidation rates are comparably high for filter pressing because employed pressures are usually about one or two orders of magnitude more than slip casting. Liquid flux through the porous particle layer is a function of pressure gradient $\nabla P$ permeability $D$, and liquid viscosity $\eta$, according to Darcy’s law:

$$ J = \frac{D \nabla P}{\eta} \quad (20) $$

5.1.4. Centrifugal Casting and Electrophoretic Deposition

Macroscopic separation of colloidal solid and liquid phases can also be achieved with the help of body forces acting on dispersed particles. Analogous to particle sedimentation caused by gravitational forces, centrifugal casting consolidates particles in a rotating mold. The settling velocity of a single spherical particle in the centrifugal force field can be approximated by a modified Stokes equation for particle movement in dilute colloids:

$$ v = \frac{2\Delta \rho R^2 \omega^2 z}{9\eta} \quad (21) $$

where $\Delta \rho$ is the density difference between solid and liquid, $R$ the particle radius, $\omega$ the angular velocity, and $z$ the distance from the rotation axis. Since liquid is not forced through a porous layer, this method offers the advantage of a linear relation between cast layer growth and casting time. As in slip casting, centrifugally cast bodies may suffer from inherent density gradients. An electric field can be used in a similar manner to accelerate particles suspended in a liquid if the particles carry a surface charge.

A corresponding technique used to consolidate particles into dense layers is called electrophoretic deposition. Similar to centrifugal casting, the accelerating force is balanced by viscous forces, resulting in a steady-state particle velocity:

$$ v = \frac{\varepsilon_r \varepsilon_0 \zeta E}{f_H \eta} \quad (22) $$

where $\varepsilon_r$ is the relative dielectrics constant, $\varepsilon_0$ the permittivity of free space, $\zeta$ the particle’s zeta-potential, $E$ the applied electric field strength, and $f_H$ the Henry constant (ranges from 1 to 1.5 depending on particle size and ionic strength).

5.1.5. Attraction and Repulsion Formulas

Applied consolidation pressure must exceed yield stress of the space-filling particle network in colloids with primarily attractive interparticle forces. The relation between yield stress, $\tau_Y$ and solids volume fraction can often be described by an empirical power law of the type

$$ \tau_Y \propto \phi^n \quad (23) $$

where $n$ is a constant fitting parameter.

Consolidation pressure must exceed osmotic pressure to achieve dense particle packing in colloids with primarily repulsive interparticle forces. Hard-sphere system osmotic pressure, $\Pi$, is a function of solids volume fraction, $\phi$, namely

$$ \Pi(\phi) = \frac{RT\phi(1 + \phi + \phi^2 - \phi^3)}{V(\phi_{\text{max}} - \phi)^3} \quad (24) $$
where \( R \) is the gas constant, \( T \) the temperature, \( V \) the molar volume of the colloid phase, and \( \Phi_{\text{max}} \) the maximum solids loading [32].

### 5.2. Tape-Casting Techniques

Liquid casting techniques allow easy manipulation of ceramic shapes before sintering. Tape casting uses wet casting principles originally used for drained casting to make flat, thin ceramic tapes as small as a few microns (1 \( \mu \)m to 1 mm). Such parts could not otherwise be fabricated until recently with new extrusion techniques. In tape casting, a ceramic slurry deposits on a flat carrier substrate, and the solvent simply evaporates. Solvent removal rate mainly depends on temperature, convective and diffusive solvent transport at the liquid–vapor interface, and internal capillary pressures in porous layers of cast tape. Tape cast products include dielectric capacitors, piezoelectric actuators, and thick and thin films. The chief market today remains electronics and radio applications.

#### 5.2.1. Slurry Composition

Although shaping, binder burnout, and sintering are important stages in determining the final properties of a tape cast item, original slurry composition influences the final product most. Slurry variables for a tape-casting run must be optimized for viscosity, particle size, and miscibility. In general, desirable viscosity is 3–5 Pas with powder particle size of 0.5–2 \( \mu \)m. However, these properties vary depending on the main powder, casting speed, thickness, and final application. Common additives to tape cast processes include dispersants, binders, plasticizers, etc.

#### 5.2.2. Common Problems and Defects

Variables involved in tape casting must be carefully monitored and controlled to prevent defects (cracks, nonuniformity, delamination, surface roughness, etc.) in the final product. Most problems originate with slurry composition. For instance, a common tape-casting problem is excessive sticking of cast tape to the belt. Adding more dispersant and plasticizer alleviates this problem. However, adding too much causes the cast tape to release from the belt early and curl up. Other signals include weak tape (caused by too little binder) and brittle tape (caused by too little plasticizer). Small deviations in additive composition greatly change the cast tape and final product.

#### 5.2.3. Waterfall Technique

Analogous to glazing ceramic tile, the waterfall casting technique utilizes a pump to pour a curtain of slurry onto substrate, then the slurry curtain dries. The slurry curtain, or waterfall, must remain thin and uniform to avoid splashing and other nonuniformities. This technique produces porous electrodes for fuel cells and multilayer ceramic capacitors [33].

#### 5.2.4. Paper Casting

In this process, low-ash paper is passed through ceramic slurry. Slurry sticks to the paper, passes through a dryer, and rolls onto a reel. The paper is removed during binder burnout. This process is used for honeycomb structures such as heat exchangers and catalytic converters [33].

### 5.3. Constant Volume Techniques

#### 5.3.1. Gel Casting

Gel casting developed in Canada in the 1960s. Since then, it evolved into an attractive forming process for manufacturing near-net-shape, very large, high-quality, complex ceramic parts with defined threshold strength [34]. The process does not remove any component or shrink the powder compact in mold. Rather, dispersing medium either is solidified via polymerization or via trapping molecules or voids within the particle system.

Numerous gel casting processes developed over the past 20 years to achieve solidification, starting with polymerization of monomers/oligomers or crosslinking polymers. Some use internal chemical reactions that form salt or change the system’s pH. Others use solubility, adsorption, or dispersant conformation changes of the ceramic material itself. Also, the dispersing medium can be frozen [1].

Differences in dispersing and gelation methods cause various challenges later in process. As examples: if more than 1 wt% organics are needed, the burnout may take days to weeks (depending on the size); if a chemical reaction facilitates consolidation, chemical gradients must be controlled to achieve uniform packing, bubble formation, or additive precipitation suppression. Gel casting’s foremost challenge is time needed to consolidate slurry in mold. Compared to pressing techniques, direct casting can take twice to thrice as long. However, the final products form at near-net-shape and with utmost quality. This saves substantial machining time and costs. Lastly, monomer toxicity remains problematic and hinders universal adoption. However, research into organic gelation compounds promises success [35].

#### 5.3.2. Extrusion

Extrusion as a manufacturing technique produces continuously long bodies of materials (metals, plastics, ceramics, etc.) with a common cross section by pressing material through shaped nozzles. Highly ductile metals (aluminum) are extruded cold (room temperature) to form hollow bars and tubes for low-strength applications such as window
frames. Other metals (copper and brass) are extruded hot to form water pipes and pipe fittings. Thermoplastics become viscous with heat and are extruded in many forms. Ceramic materials are extruded to form square-section bars, which can be cut into bricks; some (viscous alumina strips) are extruded into complex, thin-walled honeycombs and used as support media for catalysts [6,12].

In a conventional extrusion process for ceramic green bodies, ceramic paste is prepared by first adding polymers and other additives to ceramic powder. Then, the paste is heated, and a rotating screw (auger or piston) pushes it through a shaped die (nozzle) to form a green ceramic string, which takes the cross-sectional shape of the extrusion die. It can be no smaller than the die’s orifice. Next, air or water cools the profiled string, and it gets cut into desired lengths. All sorts of profiles can be made, and production volumes are normally high. Water (traditional ceramics) or organic materials (advanced ceramics) provide adequate plasticity for forming. The difficulty with extrusion is removing the organic material prior to firing without causing cracks or distortions.

During extrusion, the augers’ (or pistons’) driving force must exceed the paste’s resistive force and the die’s wall friction. Pressure is highest in the barrel and decreases along the extruder’s axis. Specific flow behavior depends on the die’s geometry and the material’s viscosity.

Co-extrusion in microfabrication (MFCX), a new method to produce axis-symmetric ceramic objects with micrometer-size features in two dimensions, can produce several shapes in alumina and PbO-containing ferric ceramics with carbon black as fugitive (material removed later in process). MFCX processes use simple round or square dies for size reduction. Shaping is done by using a shaped object of two or more materials as the extrudate. The primary material is the desired ceramic compound. The second material is a fugitive substance, which fills space and has similar viscosity as the primary ceramic. The two are co-extruded, cross-sectional dimensions are reduced, and the fugitive material is later removed. Composites with layers as fine as 8 μm have been made with this process [36–38].

5.3.3. Injection Molding

Injection-molding processes produce ceramic parts with complex shapes that other forming techniques cannot. Essentially, feed material consists of a ceramic colloid with binder, dispersant, and plasticizer additives. It is fed into a heated chamber to soften or liquefy the binder (necessary for optimum mold filling). A plunger or screw forces the feed material through a nozzle into the mold cavity and fills it. Next, the material cools until the binder solidifies. Finally, the mold opens, and the ceramic part is removed.

Suitable rheological properties of the feed material are required during mixing and injection. This means that injected material must have sufficiently low viscosity at relevant shear rates and temperatures. Accordingly, not only must the binder viscosity be low, but solids fraction of the feedstock also needs to be well dispersed in the binder matrix phase. Furthermore, viscoelastic properties of the feed material have to be considered; e.g., time and temperature-dependent relaxation of shear stresses after mold filling might lead to unwanted dimensional changes. Binder properties are of considerable importance in subsequent processing stages; e.g., burnout behavior of the binder can pose limitations on the geometry and the dimensions of the ceramic part [39].

Although it is comparably easy to switch production to a different product simply by exchanging the mold and feedstock, limiting factors of injection molding include cost competitiveness and difficulty producing large parts. An injection-molding machine’s design must account for high pressures. Moreover, precise control of injected material’s mass and temperature as well as position of the plunger/screw are essential for manufacturing high-quality parts. Finally, surfaces that contact the highly abrasive feed material wear, thus increasing costs [6].

5.3.4. Pressing

Dry powder processing involves direct compaction of either raw fine particles or granular powders in a rigid die. Common methods include die pressing, cold isostatic pressing (CIP), hot pressing (HP), and hot isostatic pressing (HIP). HP and HIP enable the green body to be formed and sintered simultaneously, but they require special apparatus.

Die pressing produces simple-shaped parts and can easily be automated. This process applies vertical stresses directly to ceramic powder granules in mold to generate loose bonding between particles. This gives the compact some mechanical strength for handling. During die pressing, the ceramic powder granules’ shapes change over three steps: (a) packing of granules, (b) deformation of granules, and (c) sliding between granules to decrease void size between deformed granules.

When stress is applied from one direction, e.g., the upper side of the mold, this is referred to as uniaxial pressing. The process automates fast and easy, quite an advantage, but sundry disadvantages also exist. Stress gradients result in density gradients in the compact green body. Sometimes, lamination may also occur because of inhomogeneous distribution of applied stress. Biaxial pressing, applied stress from above and below, minimizes the above problems. However, for very large parts, density gradients still exist, especially in areas along the mold’s walls due to friction.

CIP utilizes liquid media hydrodynamic forces to apply stress to compacted ceramic powder in all
directions. Therefore, it forms parts with an elongated dimension, relatively complex parts, or parts that may not be easily die-pressed. CIP processes are either wet-bag or dry-bag. In wet-bag processes, powder is filled into a mold, then stress is applied omnidirectionally. Wet-bag presses usually have operating pressures up to 200 MPa, and wet-bag presses operating up to 500 MPa are used for pressing large blocks. Such systems require that the ceramic-powder-filled rubber mold be immersed in a liquid medium. Then, force applied to the liquid medium passes on to the rubber mold and is equal in all directions. This method produces compact ceramic green bodies with fewer density gradients. However, it encounters problems when the part is too complicated and/or the surface finish is not good. In dry-bag processes, forces apply radially through a liquid medium between a flexible mold and a rigid shell.

6. SUMMARY AND OUTLOOK

The push toward nanotechnology in recent years changed the landscape of ceramic processing. Designed nanostructures’ fabrication methods were the focal point of research, and less attention was afforded to nanopowder processing. Still, tremendous challenges exist for processing nanoparticles that need to be addressed. With nanotechnology’s transition from research to industry and production, a new wave of research to enhance nanoparticle processing knowledge is expected.

Future efforts will require more computational approaches since slurries are so complex. Specifically, inadequacies of current measuring techniques limit our understanding of what happens when particles come within close proximity of a few nanometers. Furthermore, simplification of averages, which works for long distances, no longer applies. Thus, studying nanoparticles with proximities of a few nanometers will help achieve dense structures with acceptable shrinkage.

REFERENCES


Chapter 11.1.4

Solvothermal Synthesis of Metal Oxides

Masashi Inoue
Professor Emeritus of Kyoto University, Japan

1. INTRODUCTION

Syntheses of metal oxides are usually performed by calcination of suitable precursors, such as hydroxides, nitrates, carbonates, carboxylates, etc. Since this process requires large thermal energies, sintering of the product particles takes place and aggregated particles are usually obtained. Liquid-phase syntheses allow the formation of metal oxides at relatively lower temperatures, and products with well-defined morphologies can be obtained. The sol–gel method (or alkoxide method) is one of these methods [1,2], which enables the formation of metal oxides at low temperatures. However, this method usually gives amorphous products, and calcination of the products is required to obtain crystallized products.

This chapter covers solvothermal methods. Various compounds have been prepared by solvothermal reactions: metals [3,4], metal oxides [5–11], halides [7], chalcogenides [12,13], nitriles [12,14,15], phosphides [16], open-framework structures [17,18], oxometalate clusters [19,20], organic–inorganic hybrid materials [18,21,22], metal organic frameworks (MOF) [23], and carbon nanotubes [24,25]. Most solvothermally synthesized metal oxides are composed of nano- or microparticles with well-defined morphologies. The distribution of the particle size of the product is usually quite narrow, and the formation of monodispersed particles is frequently reported [26,27]. When the solvent molecules or additives are preferentially adsorbed on (or have a specific interaction with) a certain surface of the products, growth of the surface is prohibited and therefore products with unique morphologies may be formed by the solvothermal reaction [12,28,29]. Thus, nanorods [29], nanowires [30], tubes [31], and sheets [32] of various types of products have been obtained solvothermally.

Since 1984, we have been exploring the synthesis of inorganic materials in organic media at elevated temperatures (200–300 °C) under autogenous pressure of the organics using an apparatus shown in Figure 1 [33]. This technique is now generally called the “solvothermal” method [14]. The term “solvothermal” nominally means reactions in liquid or supercritical media at temperatures

![Figure 1](image_url)  
**Figure 1**: Geometry of our apparatus for solvothermal reaction. a) For the hydrolysis of alkoxide by water dissolved from the gas phase (Section 3.2.5), water was added in the gap between the autoclave wall and test tube.
higher than the boiling point of the medium. Some researchers define “solvothermal reaction” as a reaction that is carried out in a closed vessel. Reactions in a closed system using autoclaves or sealed ampoules and in an open system using a flask equipped with a reflux condenser sometimes give completely different results, especially when low-boiling-point products such as water are formed as by-products and affect the reaction. However, most researchers use the word “solvothermal” without showing evidence that the reaction behavior in the closed system is different from that in the open system.

The density of the liquid solvent is essentially unchanged at above or below the boiling point because the compressibility of the liquid is quite small. The effect of pressure on the reaction kinetics is determined by the activation volume, relative volume of the activated complex at the transition state to the volume of the starting molecule(s). However, in order to measure the effect of reaction pressure, the GPa-scale pressure is required. This means that the autogenous pressure created by the vapor pressure of the solvent has only a minor effect on the reaction rate. Therefore, there is no need to differentiate the reactions at the temperature above and below the boiling point. We previously proposed a broad definition of “solvothermal” reaction as a reaction in a liquid (or supercritical) medium at high temperatures [34]. However, this definition is so broad that flux methods are also included in this category, and a review paper discussed the flux method as a solvothermal method [35]. However, flux rather reflects the nature of melt, not solution, and, therefore, in this article, solvothermal reaction is defined as a solution-based reaction at high temperatures.

The prefix “solvo-” means any kind of solvent; various solvents such as water [36–40] ($T_c$, 374; $P_c$, 218 atm), ammonia [41–43] (bp, 78 °C; $T_c$, 132 °C; $P_c$, 113 atm), sulfur dioxide (bp, 10 °C, $T_c$, 157.5 °C, $P_c$, 78 atm), hydrofluoric acid [44] (bp, 19.5 °C; $T_c$, 188 °C, $P_c$, 188 atm), and many organic compounds have been used for the solvothermal reaction. Hydrothermal reactions are one type of solvothermal reaction; in this chapter, however, “solvothermal” is used to refer to reactions in organic solvents. When alcohols and glycols are used as the reaction media, the reactions are called “alcoholothermal” and “glycothermal” reactions, respectively.

In 2004, Cooper et al. [45] reported a new type of solvothermal synthesis in which an ionic liquid or eutectic mixture (they used the mixture of choline chloride and urea in a 1:2 ratio, which has a melting point of 12 °C and exhibits unusual solvent properties that are very similar to those exhibited by the ionic liquid [46]) was used as both the solvent and the structure directing agent in the synthesis of zeolites. This methodology was termed ionothermal synthesis and has received growing interest for synthesizing open-framework structures such as zeolites, aluminophosphates, and organic—inorganic hybrid materials, many of which have novel structures [47,48].

2. SAFETY CONSIDERATION

2.1. High-Pressure Experiments

To carry out reactions at temperatures higher than the boiling point of the reaction medium, pressure vessels (autoclaves) are usually required. Some researchers favor the use of sealed ampoules of glass or silica, but these experiments should be carried out with great care because the ampoules are easily broken by the internal pressure of the reaction medium. To avoid explosion of the ampoules, they may be placed in an autoclave together with a suitable medium to create a vapor pressure to balance the internal pressure of the ampoule.

Thermal expansion of liquid is an important factor that should be taken into account to carry out experiments safely. The volume of liquid increases with the increase in temperature, and thermal expansion is drastic near the critical point. When the volume of the liquid reaches the volume of the vessel, a further increase in the reaction temperature resulted in a drastic increase in the pressure of liquid, because the compressibility of liquid is quite small. Therefore, the percentage of fill of the reaction medium in the autoclave is an important factor. For the hydrothermal system, the Kennedy diagram [49] can be used for estimation of the reaction pressure. For organic solvents, however, such diagrams are not available. Therefore, one should estimate the density of the reaction medium at the desired reaction temperature by using a suitable method (for example, the Hanson diagram [50]). From the density of the liquid, one can calculate the maximum volume that can be charged to the reaction vessel. However, the volume of the medium should be smaller than the maximum volume, so that a reaction does not create unexpected high pressure.

2.2. Organic Medium

Various organic solvents have been applied for the syntheses of inorganic materials [34]. As ceramists may not be familiar with organic solvents, some important features are summarized here.

From a thermodynamic point of view, most organic compounds have an inherent tendency to decompose into carbon, hydrogen, nitrogen, and so forth, at high temperatures,
although the reaction is endothermic. Methane is the most stable aliphatic hydrocarbon; however, it can decompose into carbon and hydrogen at temperatures $>360^\circ C$ in the presence of a suitable catalyst such as nickel or iron [51]. Therefore, most of the organic compounds, even though they may be stable at room temperature, act as reducing agents at high temperatures. Therefore, the presence of any oxidants in the reaction system of the solvothermal reaction using organic solvents cannot be recommended. The gas phase in the reaction vessel must be completely purged with an inert gas such as nitrogen. Since organic compounds easily react with oxygen with highly exothermic nature (i.e., combustion of organics), the reaction continues until completion. Note that even though one could conduct an experiment safely, this fact does not guarantee the safety of another experiment.

When an organic molecule contains oxygen atoms in the structure, it may spontaneously decompose into carbon dioxide, water, nitrogen, and so forth, and the reaction becomes exothermic. To assess the intrinsic nature for the exothermic decomposition of organics, oxygen balance may be used. Oxygen balance ($OB$) is defined as the weight of oxygen that is excess or deficient for the complete decomposition of 100 g of a compound and is given by the following equation:

$$C_{x}H_{y}O_{n} + (x + y/4 - z/2)O_{2} \rightarrow xCO_{2} + y/2 \text{H}_{2}O + n/2N_{2}$$

$$OB = -1600(x + y/2 - z)/M_{w}$$

where $M_{w}$ is the molecular weight of the compound. A compound having an $OB$ value near zero has a high tendency to explode.

One should also consider the oxygen balance of the reaction system. For the reaction of a nitrate of a divalent metal ion with ethylene glycol, for example (Eqn 3), oxygen balance is zero, indicating that this reaction system can explode:

$$\text{M(NO}_{3})_{2} + \text{HOCH}_{2}\text{CH}_{2}\text{OH} \rightarrow \text{MO} + 2\text{CO}_{2} + 3\text{H}_{2}\text{O} + \text{N}_{2} \quad OB = 0\%$$

The reaction shown in Eqn (3) may take place via the following equation:

$$\text{M(NO}_{3})_{2} + \text{HOCH}_{2}\text{CH}_{2}\text{OH} \rightarrow \text{M(OH)}_{2} + \text{O}_{2}\text{NOCH}_{2}\text{CH}_{2}\text{ONO}_{2}$$

The last compound of the equation is a nitrate ester of ethylene glycol and is known as nitroglycerol, a highly powerful explosive that was used as an ingredient of dynamite. Even though excess ethylene glycol is used, where the $OB$ value of the whole reaction system is highly negative, the reaction system should be carefully treated as a mixture of nitroglycerol and ethylene glycol.

For the choice of the solvent, a less toxic solvent should be selected, for example, toluene instead of benzene. Benzene is highly toxic and it causes fatal damage to hematopoietic organs. $n$-Hexane is one of the most common aliphatic solvents; however, this compound has a specific toxicity that is not shown by branched $C_{6}$ hydrocarbons or by other straight-chain paraffins such as pentane and heptane. The origin of this toxicity is well established. Through the metabolic system of human beings, this compound is oxidized by oxygenase to 2,5-hexanedione (Eqn 5), which attacks the peripheral nervous system [52]. Therefore, heptane should be used in place of hexane:

$$\text{Cytochrome P-450} \quad \text{H}_{3}\text{C}\text{C} = \text{C} = \text{O} \quad \text{O} \quad \text{CH}_{3}$$

Thermal stabilities of the solvents in terms of decomposition and polymerization should also be taken into account. Spontaneous decomposition of organic compounds was discussed earlier. Cyclic ethers such as tetrahydrofuran (THF) may cause ring-opening polymerization which has a severe exothermic nature. Dimethyl sulfoxide (DMSO) is an aprotic polar solvent and is widely used in organic chemistry. However, the thermal stability of DMSO is slightly lower than that of other aprotic polar solvents such as $N,N$-dimethylformamide (DMF) and hexamethylphosphoramide (HPMA). Moreover, this compound may cause an explosive reaction resulting in the formation of dimethyl sulfide when it contacts compounds such as acid halides.

$$\text{O} \quad \text{CH}_{3}\text{SCH}_{3} \quad \text{H}_{3}\text{C} \quad \text{H}_{3}\text{C} \quad \text{N-CHO}$$

Dimethyl sulfoxide (DMSO); bp = 189 °C

Dimethylformamide (DMF); bp = 153 °C

$$\text{H}_{3}\text{C} \quad \text{N-P} \quad \text{N} \quad \text{CH}_{3} \quad \text{H}_{3}\text{C} \quad \text{CH}_{3} \quad \text{H}_{3}\text{C} \quad \text{CH}_{3}$$

Hexamethylphosphoramide (HPMA); bp = 233 °C

Before performing solvothermal reactions in any type of solvent, ceramists or inorganic chemists who are not familiar with the organic compounds are recommended to consult with organic chemists and refer to the material safety data sheets available from the supplier of the chemicals.
2.3. Surface Acidity of the Products

Some mixed oxides such as amorphous silica–alumina have strong surface acidity. Since the particles of solvothermal products are usually covered with organic moieties, the surface acidity does not work well. However, when the reaction system contains a small amount of water, it may hydrolyze the surface organic moieties, resulting in exposure of an acidic surface. An acidic surface possibly decomposes an organic solvent such as alcohol resulting in the formation of water. The thus-formed water further hydrolyzes surface organic moieties, resulting in facilitating the decomposition of the organic solvent. Therefore, the reaction pressure may suddenly begin to increase. This event has occurred, and, therefore, reaction pressure should be monitored during the course of the solvothermal experiments. The reaction pressure is usually determined by the vapor pressure of the solvent. However, under the solvothermal conditions, the reaction pressure may be much higher than expected from the vapor pressure of the solvent because of the decomposition of the solvent.

3. SOLVOTHERMAL SYNTHESIS OF METAL OXIDES – CASE STUDIES

3.1. Glycothermal Synthesis of α-Alumina

3.1.1. Reaction of Microcrystalline Gibbsite

In 1989, Inoue et al. [5] briefly reported that the reaction of microcrystalline gibbsite (a modification of Al(OH)₃) in 1,4-butanediol or 1,6-hexanediol at ~300 °C for 2 h (glycothermal reaction) yields α-alumina through the glycol derivative of boehmite as an intermediate (Figure 2) [21,53,54]. This intermediate is a class of intercalation compound [55], in which some of the hydroxyl groups in the layers of boehmite (a modification of aluminum hydroxide oxide, AOOH) are substituted with ω-hydroxylalkoxyl groups [21,33]. In other words, the alkyl groups are bound to the boehmite layers through covalent bonding. The particles of starting gibbsite are hexagonal palates, while the intermediate phase has a characteristic thin wrinkled sheet morphology. The produced α-alumina particles are hexagonal plates with a relatively large surface area (Figure 3). This indicates that the dissolution–recrystallization mechanism takes place.

3.1.2. Thermodynamical Consideration

Hydrothermal reactions of aluminum compounds have been well studied, and this reaction easily yields boehmite at relatively low temperatures (~200 °C) [56]. The equilibrium point between diaspore (another polymorph of AlOOH) and α-alumina + H₂O under the autogenous vapor pressure of water was reported to be 360 °C [57]. However, near the equilibrium point, the transformation rate would be very sluggish, and only a small conversion of diaspore would be attained. Therefore, a much higher temperature is required to achieve complete conversion of diaspore into α-alumina.

Since boehmite is slightly less stable than diaspore (i.e., boehmite is a metastable phase), the hypothetical equilibrium point between boehmite and α-alumina under the hydrothermal conditions would be slightly lower than that for the diaspore–α-alumina system. However, α-alumina...
Solvothermal Synthesis of Metal Oxides

Figure 4: Comparison between glycothermal and hydrothermal boehmite into more than 10 h are required for the complete conversion of a low temperature. Actually, Yanagida et al. reported that would not be formed by a hydrothermal reaction at such a low temperature. Actually, Yanagida et al. reported that more than 10 h are required for the complete conversion of boehmite into \( \alpha \)-alumina, even with a reaction at 445 °C in a 0.1 N NaOH solution and in the presence of seed crystals [58]. On the other hand, under glycothermal conditions, complete conversion of gibbsite into \( \alpha \)-alumina is attained by the reaction at 285 °C for 4 h.

To the best of our knowledge, this reaction is the first example of crystallization of a thermodynamically stable metal oxide in an organic solvent at a temperature lower than that required by the hydrothermal reaction. In other words, this example showed that dense metal oxides can crystallize in organic solvents at temperatures lower than those required for crystallization in water.

The difference between these two reactions may be attributed to the activities of water present in the reaction systems, since the overall reaction is dehydration of aluminum hydroxide. However, intentional addition of a small amount of water caused the enhancement of \( \alpha \)-alumina formation rather than the retardation expected from the equilibrium point of view [59]. Therefore, facile formation of \( \alpha \)-alumina under glycothermal conditions is attributed to the thermodynamical instability of the intermediate, the glycol derivative of boehmite, as compared with the well-crystallized boehmite formed under hydrothermal conditions (Figure 4). The glycol derivative of boehmite has Al–O–C bonds and therefore is more unstable with respect to boehmite itself. Thus, conversion of the former compound into \( \alpha \)-alumina has a much larger driving force. These arguments lead to an interesting prediction that if the formation of well-crystallized boehmite from gibbsite is somehow prevented under the hydrothermal conditions, the formation of \( \alpha \)-alumina would take place at much lower temperatures.

The glycothermal reaction of gibbsite is strongly affected by the particle size of the starting material [59]. When gibbsite with a particle size less than 0.2 μm is used, complete conversion into \( \alpha \)-alumina is attained. When coarse gibbsite is used for the reaction, \( \alpha \)-alumina was not formed at all but the product is \( \gamma \)-alumina together with well-crystallized boehmite [59]. The latter compound is partially formed by thermal dehydration of coarse gibbsite via intraparticle hydrothermal reaction [60, 61]. Since the reaction temperature is close to the temperature for thermal dehydration of gibbsite, the formation of the intermediate, the glycol derivative of boehmite via dissolution of the starting material competes with the thermal dehydration of gibbsite yielding \( \gamma \)-alumina and well-crystallized boehmite. Therefore, the small particle size of gibbsite, which facilitates the dissolution rate, is essential for the formation of \( \alpha \)-alumina. When thermal dehydration of gibbsite into \( \gamma \)-alumina takes place, the thus-formed \( \gamma \)-alumina facilitates epitaxial decomposition of the glycol derivative of boehmite into the transition alumina. Therefore, coarser gibbsite as the starting material of the glycothermal reaction does not give \( \alpha \)-alumina, even at higher reaction temperatures.

### 3.1.3. Occlusion of Foreign Ions

When precipitates are formed from alkaline solutions, a significant amount of alkali cations is incorporated into the precipitates. Since commercial gibbsite is produced by crystallization from the sodium aluminate solution, the gibbsite samples are always contaminated with sodium ions. This can be easily recognized by the fact that the mother liquor after the hydrothermal conversion of gibbsite yielding boehmite is highly alkaline. On the other hand, the mother glycol solution after the glycothermal conversion of gibbsite into \( \alpha \)-alumina is not alkaline. This result indicates that a significant amount of sodium ions is occluded in the glycothermal \( \alpha \)-alumina.

Some researchers claimed that one of the specificities of the solvothermal reaction is the formation of contamination-free products. However, this is not a general characteristic of the solvothermal reactions but rather specific to the hydrothermal reaction. Solvothermal products obtained in organic solvents easily occlude various foreign ions because the solubilities of these ions in organic solvents are rather low. This is a drawback of the solvothermal reactions but is a merit when preparing composite oxides. When one of the component ions has high solubility in water, these ions are easily expelled from the products under hydrothermal conditions. On the other hand, the solvothermal method gives a convenient route for preparing composite oxides that are difficult to prepare by the hydrothermal method.

### 3.1.4. Effects of Seeds

We reported that the reaction of aluminum alkoxide in glycols also yields the glycol derivatives of boehmite [54, 55]. The products have the same morphology as that of the products obtained from gibbsite. However, the reaction of aluminum alkoxides in glycols does not yield \( \alpha \)-alumina even at elevated reaction temperatures. To account for these results, two other factors must be considered [59]. First, water formed as a by-product in the formation of the
glycol derivative of boehmite facilitates crystallization of $\alpha$-alumina by increasing the dissolution rate. Intentional addition of a small amount of water actually lowers the $\alpha$-alumina crystallization temperature; however, excess water causes the formation of boehmite, which is fairly stable under the reaction conditions and contaminates the final product.

Another point is the presence of $\alpha$-alumina nuclei in the starting gibbsite sample. The starting gibbsite is commercially produced by milling coarser gibbsite. Prolonged grinding or milling of gibbsite particles causes mechano-chemical transformation into $\alpha$-alumina [62]. Therefore, $\alpha$-alumina-like domains are formed at this stage and act as the nuclei of $\alpha$-alumina in the glycothermal reaction. Kaiser et al. [63] and Bell et al. [64] also addressed the importance of seed particles on the formation of $\alpha$-alumina and on the size of glycothermally prepared hexagonal $\alpha$-alumina platelets.

Although the glycol derivative of boehmite prepared from aluminum alkoxides cannot transform into $\alpha$-alumina because of the absence of $\alpha$-alumina nuclei and water, the reaction of a mixture of gibbsite (microcrystalline) and aluminum alkoxide in the presence of a small amount of water gives $\alpha$-alumina with larger crystallite sizes than those obtained from gibbsite alone [59].

Kaiser et al. [63] reported that the reaction of a 0.6-μm-sized gibbsite in 1,5-pentanediol did not yield $\alpha$-alumina but that tohdite (5Al₂O₃·H₂O) was obtained. This is presumably because the crystal size of the starting material is insufficiently small for glycothermal synthesis of $\alpha$-alumina. They also noted that the presence of $\alpha$-alumina seed crystals in the reaction system yields $\alpha$-alumina.

Cho et al. [65] reported that glycothermal treatment of gibbsite prepared by precipitation of aluminum nitrate with potassium hydroxide yields $\alpha$-alumina, even though the particle size of the gibbsite is much larger than 0.2 μm. However, their starting material seems to be contaminated with potassium ions, which seem to play an important role in the nucleation of $\alpha$-alumina. They noted that the produced $\alpha$-alumina contained 0.09 wt% potassium.

3.1.5. Morphology of Glycothermal $\alpha$-Alumina

Bell and Adair [66] examined the effects of various additives on the morphology of the glycothermally obtained $\alpha$-alumina. They concluded that alcohols specifically adsorb on {01T12} habit planes (see Figure 3) to promote a platelet morphology, while the addition of glacial acetic acid causes specific adsorption-stabilizing {01T12} habit planes resulting in the formation of acicular morphology. They also reported that hydroxide ions and pyridine also affect the morphology of the $\alpha$-alumina particles because of their preferential adsorption of a specific surface.

Adair et al. [67] also reported the effects of process conditions such as the stirring rate and seed loading on the morphology of $\alpha$-alumina formed by the glycothermal reaction. Other researchers also discussed the morphology of glycothermally prepared $\alpha$-alumina [68–70]. Li and Choi [71] discussed the effect of adsorption of 1,4-butanediol on the morphology of the $\alpha$-alumina particles on the basis of molecular dynamics calculation. However, these discussions neglect the effect of alkali ions remaining in the $\alpha$-alumina and further study is required to completely understand the morphology of $\alpha$-alumina formed under glycothermal conditions.

3.1.6. Gas–Liquid Equilibrium

It was reported that the addition of THF (a by-product formed from 1,4-butanediol; vide infra) and nitric acid had essentially no effect on the morphology of the product [66]. In the solvothermal reaction, the gas–liquid equilibrium should be considered as well as the adsorption of solvent and additive molecules on the surface of the starting material or the product. Since THF has a much lower boiling point than 1,4-butanediol, most of the THF molecules are present in the gas phase, thus having negligible effect on the morphology of the product. As for the effect of nitric acid, the authors attributed the result to the nonadsorbing electrolyte property of nitrate ions. However, nitrate complex of various metal cations is known and, therefore, this explanation is questionable. Nitric acid is a strong oxidant, and, therefore, is easily reduced by 1,4-butanediol, resulting in the formation of NO and N₂O. These by-products easily escape to the gas phase. The nitric acid oxidation of organic compounds is a severely exothermic reaction and it is not recommended to carry out such an experiment in a closed vessel, although the authors carried out the reaction with only a small amount of nitric acid.

3.2. Titania

3.2.1. Thermal Decomposition of Titanium Alkoxide in an Inert Organic Solvent

We examined the solvothermal reaction of titanium isopropoxide (TIP) and titanium oxyacetylacetonate (TiO(acac)₂) in various organic solvents, including toluene, 2-butanol, and glycols, and found that some of the reactions yield anatase crystals [72].

Thermal decomposition of metal alkoxides is usually applied for synthesizing oxide films or aerosols by the chemical vapor deposition (CVD) method [73]. Whereas CVD reactions of metal alkoxides under reduced pressure usually give amorphous products, solvothermal reactions of alkoxides in inert organic solvents, such as toluene, in some cases, yield crystalline products. For example, thermal decomposition of aluminum isopropoxide, TiO(acac)₂, and zirconium isopropoxide in toluene at 300 °C yields $\chi$-alumina [74], anatase [72], and tetragonal zirconia [75],...
respectively. This means that solvent molecules present in the reaction system facilitate the crystallization of the product.

This reaction is strongly affected by the structure of the alkyl groups of the alkoxides. For aluminum and zirconium alkoxides, reactivity of the alkoxides increases according to the following order of alkyl structure: primary < secondary < tertiary. This result indicates that heterolytic cleavage of the C–O bond yielding carboxylation and metal oxo anion (≡M–O−; Eqn 6) is the key step, and stability of the carboxylation determines the reactivity of the metal alkoxides [74]: Since the solvent molecules cannot stabilize these charged species, however, the formation of an ion pair seems to be most probable. The anion (≡M–O−) attacks another molecule of the alkoxide yielding ≡M–O≡ bond and alkoxide anion (RO−). On the other hand, the carboxylation gives olefin and proton via E1 reaction. It is known that dissociative reaction (E1 reaction) is favored by the high temperature rather than the associative route (SN1 reaction) required by the formation of ether from two molecules of metal alkoxides. The RO− anion gives alcohol by reaction with the proton:

\[
\begin{align*}
\text{M(OR)ₙ} & \rightarrow \text{M–O–M} + RO^- \\
\text{M(OR)ₙ} & \rightarrow \text{M} + \text{H}^+ + \text{ROH} \\
\text{M} + \text{H}^+ + \text{ROH} & \rightarrow \text{Olefin} + \text{H}^+
\end{align*}
\]

Shulman et al. [76] examined the kinetics of the thermal decomposition of aluminum alkoxides and found that the activation entropy is positive, suggesting a transition state less organized than the initial state. Based on these findings, they ruled out the possibility that the reaction takes place via a cyclic mechanism (Eqn 6) and proposed a carboxylation mechanism, which is consistent with our conclusion.

However, this reaction is strongly affected by the metal cation of the alkoxide, and TIP was found not to decompose in toluene at 300 °C [72]. The factors of metal cations controlling the decomposition of alkoxides have not yet been fully elucidated; electronegativity and the oligomeric structure of the alkoxides seem to determine the reactivity of metal alkoxides.

Following the preliminary work [72], Kominami et al. [77] further examined the solvothermal decomposition of titanium tert-butoxide in toluene at 300 °C, and found that this reaction yields nanocrystalline anatase. Note that the lowest temperature required for the formation of crystalline titania by CVD synthesis was reported to be 400–450 °C [78–80].

Payakgul et al. [81] found that the reaction of titanium n-butoxide in toluene at 300 °C initially forms an amorphous product, which then crystallizes into anatase. Kim et al. [82] reported that the thermal decomposition of TIP in toluene is affected by the TIP/toluene ratio, and found that anatase crystals with 10–30 nm size can be obtained at 250 °C for 3 h, when TIP/toluene ratios are in the range of 10/100–30/100. We reported that TIP was not decomposed in toluene even at 300 °C, although the reaction was carried out with a low TIP/toluene ratio [72]. Since this procedure is severely affected by the presence of a trace of water, further work is desired to determine why high or low loading of TIP does not give anatase crystals. Investigation on the effect of the structure of alkyl groups on the reactivity of titanium alkoxides is also desired, because the reactivity tendency may be different from those found in aluminum and zirconium alkoxides.

Parala et al. [83] synthesized anatase nanoparticles with 3–5-nm particle size by thermal decomposition of titanium tri-isopropoxidedimethylaminoethoxide (Ti(O(Pr)₃)(OCH₂CH₃NMMe₂)) and diethylaminoethoxide (Ti(O(Pr)₂)(OCH₂CH₃NMe₂)) in trioctylphosphine oxide (TOPO) at 325 °C. They reported that the product showed a broad emission centered at 370 nm and that large spherical submicron particles are obtained by slow evaporation of a dispersion of titania nanocrystals in toluene [83].

3.2.2. Hydrolysis of Titanium Alkoxides by Water, Homogeneously Formed in Organic Media

Hydrolysis of metal alkoxide with water formed by the reaction of the solvent is a variation of the sol–gel synthesis of metal oxide. Water formed homogeneously in organic media under the solvothermal conditions hydrolyzes metal alkoxides and facilitates solvothermal crystallization of metal oxides. This method can be regarded as the sol–gel version of the homogeneous precipitation method, where precipitants are formed homogeneously in the reaction system avoiding the heterogeneity caused by the addition of precipitants to the reaction system. To the best of our knowledge, this strategy was first applied by Fanelli and Burlew [84], who examined the reaction of aluminum sec-butoxide in 2-butanol at 250 °C and reported the formation of alumina with quite a large surface area. In this
reaction, 2-butanol decomposes yielding butenes and water. Before the work of Fanelli and Burlew [84], Takahashi et al. [85] noted that the presence of 2-propanol in the CVD reaction facilitates the decomposition of TIP because of the formation of water from alcohol. Following the work of Fanelli and Burlew [84], Inoue et al. examined the reaction of TIP in 2-butanol at 250 °C and found that anatase with 22 nm crystallite size was formed [72].

Coutecuisse et al. [86] reported a flow reactor for the formation of titania by the decomposition of TIP in 2-propanol at 260–300 °C. They reported that the rate-determining step is the thermal dehydration of titanium hydroxide formed by the hydrolysis of the precursor alkoxide [87]. They also found that an increase in the supercritical fluid density decreased the overall reaction rate, but this was not adequately explained. Kominami et al. [88] examined the reaction of various titanium alkoxides in secondary alcohols and concluded that the reaction takes place by dehydration of the alcohols followed by hydrolysis of the alkoxides. They also found that the products had high activities for photocatalytic mineralization of acetic acid. They explained the results by the high crystallinity and high surface area of the products [88].

Besides dehydration of alcohols, esterification can be used for this strategy. Thus, Invanda et al. [89] synthesized titania (anatase + amorphous phase) by the reaction of TIP in mixtures of carboxylic acid and alcohol. Water formed by esterification of the solvents hydrolyzed the alkoxide yielding metal oxide. Cozzoli et al. obtained nearly spherical nanocrystals by the reaction of TIP in oleic acid and ethylene glycol [90].

\[
\begin{align*}
\text{CH}_3\text{(CH}_2\text{)}_7\text{C} & = \text{C}(\text{CH}_2\text{)}_7\text{COOH} \\
\text{Oleic acid; bp} & = 360 \degree \text{C}
\end{align*}
\]

Zhong et al. [91] synthesized anatase nanorods by calcination of the precursor obtained by solvothermal treatment of a mixture of TIP, acetic acid, 2-propanol, and aniline, in the presence of a drop of H$_2$SO$_4$ at 100 °C. H$_2$SO$_4$ was added to accelerate the esterification reaction. Because of the amorphous nature of the solvothermal product, mechanisms for the formation of nanostructure and role of amines were not elucidated.

Titanium alkoxides are highly active homogeneous catalysts for esterification and transesterification [92–95], and the catalyst activity of TIP was reported to be 100 times higher than that of H$_2$SO$_4$ when activity was compared based on mole of the catalyst [94].

Aldol condensation may also be used for the supply of water. However, actual chemistry is quite complex and titanium-oxo cluster \([\text{Ti}_3\text{O}(\text{OiPr})_7(\text{O}_3\text{C}_9\text{H}_{15})]\) was identified by the reaction of TIP with acetone at room temperature for 24 h [96]. The moiety, \(\text{O}_3\text{C}_9\text{H}_{15}\), in the cluster is formed by the aldol condensation at both methyl groups of acetone. Garnweitner et al. [97] reported the formation of anatase by the solvothermal reaction of TIP in acetone at 130 °C, and quite complex mixtures of the aldol condensation products were obtained.

Li et al. [98] reported a unique method: they found that anatase nanoparticles were formed by the reaction of titanium n-butoxide with linoleic acid and triethylamine in the presence of NH$_4$HCO$_3$, which spontaneously decomposes at 150 °C to produce water.

\[
\begin{align*}
\text{CH}_3\text{(CH}_2\text{)}_4\text{C} & = \text{C}(\text{CH}_2\text{)}_7\text{COOH} \\
\text{Linoleic acid; bp} & = 365 \degree \text{C}
\end{align*}
\]

3.2.3. Glycothermal Reaction of Titanium Alkoxides

The reaction of TIP and TiO(acac)$_2$ in 1,4-butanediol at 300 °C yields anatase with relatively small crystallite sizes [72]. This is a simple extension of glycothermal reaction of aluminum compounds, and the first step of this reaction is an alkoxyl exchange reaction that takes place at lower temperatures [99]:

\[
\text{M(OR)}_n + n\text{R'}\text{OH} \rightleftharpoons \text{M(OR')}_n + n\text{ROH} \quad (7)
\]

This reaction is in equilibrium, and therefore the composition of alkyl groups in the coordination sites of the metal is determined by the relative amount of the two alkyl groups in the reaction system. Since the alcohol derived from the alkoxides has a lower boiling point than glycol, the alcohol will escape from the reaction system to the gas phase (see Section 3.1.6), and therefore the corresponding alkyl groups are completely expelled from the product.
The crystallite size of the product obtained from aluminum alkoxide increased in the following order: \( \text{HO(CH}_2\text{)}_3\text{OH} < \text{HO(CH}_2\text{)}_2\text{OH} < \text{HO(CH}_2\text{)}_3\text{OH} < \text{HO(CH}_2\text{)}_4\text{OH} \). The physical properties of the products and the alumina derived by calcination thereof varied according to this order [53,100]. These results suggest that the development of the product structure is controlled by the heterolytic cleavage of the \( \text{O}--\text{C} \) bond of the glycoxide intermediate \( >\text{Al}--\text{O}--(\text{CH}_2)_n\text{OH} \) formed by alkoxyl exchange between aluminum alkoxide and the glycol used as the medium. In the solvolytic reaction of \( >\text{Al}--\text{O}--(\text{CH}_2)_n\text{OH} \), aluminate ion \( (>\text{Al}--\text{O}^-) \) is considered to be a leaving group, and because of the electron-withdrawing effect of the intramolecular hydroxyl group the cleavage of the \( \text{O}--\text{C} \) bond proceeds more readily with increasing the carbon number of the glycol. When the number is 4 the cleavage of the \( \text{O}--\text{C} \) bond is accelerated by participation of the intramolecular hydroxyl group [101], yielding protonated tetrahydrofuran, which was actually detected in the supernatant of the reaction mixture (see, Figure 5). Similarly, reaction of titanium alkoxides proceeds yielding anatase.

Kang et al. [102] reported that the anatase sample prepared from TIP using 1,4-butanediol has a high photochemical activity for the decomposition of \( \text{CHCl}_3 \). Klongdee et al. [103] also reported that the glycothermally catalyzed reaction \( \text{CHCl}_3 \) proceeds faster than under nonhydrolytic sol-gel conditions. They also noted that needle-like anatase is prepared by topochemical decomposition of the glycolate precursor [105]. By the addition of titanium alkoxide to ethylene glycol \( \text{EG}/\text{Ti} > 1000 \) preheated at 170 °C, Jiang et al. [106] prepared titanium glycolate nanowires with an average diameter of 50 nm, which can easily be transformed into polycrystalline anatase nanowires by calcination. They noted that when precursor concentration is higher, much thicker and shorter nanowires were formed [106].

Kominami et al. [107] reported that solvothermal treatment of \( \text{TiO(acac)}_2 \) in ethylene glycol at 300 °C in the presence of sodium laurate \( \text{[CH}_3\text{(CH}_2)_{10}\text{COONa]} \) and a small amount of water yielded microcrystalline brookite having an average size of 14 nm × 67 nm without contamination of other \( \text{TiO}_2 \) phases.

### 3.2.4. Reaction of Titanium Chloride

Since this method is closely related to the nonhydrolytic sol-gel reaction [11,108,109], the basic chemistry of the reaction is discussed first. This method was initially used for the synthesis amorphous gels, but nowadays many papers report on the formation of crystalline products by the solvothermal treatment of the nonhydrolytic sol-gel systems.

The initial step of the reaction of metal chloride with alcohol or ether is the formation of alkoxy groups by the alcoholysis or etherolysis of metal halides:

1. \( \equiv\text{M}--\text{X} + \text{ROH} \rightarrow \equiv\text{M}--\text{OR} + \text{HX} \) (alcoholysis) \( \tag{8} \)
2. \( \equiv\text{M}--\text{X} + \text{ROR} \rightarrow \equiv\text{M}--\text{OR} + \text{RX} \) (etherolysis) \( \tag{9} \)

The subsequent step is the formation of an oxo bridge between two metal atoms by eliminating a small organic molecule as shown in the following equations:

1. \( \equiv\text{M}--\text{OR} + \text{RO}--\text{M} \equiv \rightarrow \equiv\text{M}--\text{O}--\text{M} \equiv + \text{RO} \) (ether elimination) \( \tag{10} \)
2. \( \equiv\text{M}--\text{OCOR} + \text{R'O}--\text{M} \equiv \rightarrow \equiv\text{M}--\text{O}--\text{M} \equiv + \text{R'OCOR} \) (ester elimination) \( \tag{11} \)
3. \( \equiv\text{M}--\text{Cl} + \text{RO}--\text{M} \equiv \rightarrow \equiv\text{M}--\text{O}--\text{M} \equiv + \text{RCl} \) (alkyl halide elimination) \( \tag{12} \)

In the ordinary sol-gel reactions, this step is performed by the reaction of two of \( \equiv\text{M}--\text{OH} \) species yielding \( \equiv\text{M}--\text{O}--\text{M} \equiv \) and water and is known as the condensation reaction.

1. **Alkyl halide elimination**

Arnal et al. [110] examined the reactions of \( \text{TiCl}_4 \) with TIP (Eqn 12), ethers (Eqn 9 → Eqn 12), and alcohols (Eqn 8 → Eqn 12) in a sealed glass tube at 110 °C for 7 days.
Titania samples prepared by these methods strongly varied in structure depending on the reaction used. For example, the reaction of TiCl₄ with diethyl ether affords anatase, and the reaction of TiCl₄ with ethanol led to rutile, whereas the reaction with tert-butyl alcohol resulted in the formation of a mixture of rutile and brookite. They concluded that the effect of the alcoholic medium is predominant for low-temperature crystallization of titania. However, we believe that hydrochloric acid liberated from the reaction facilitates the crystallization of titania through the bond breaking and forming equilibrium (as shown in Eqn 13):

\[ \equiv \text{Ti} - \text{O} \equiv \text{Ti} + \text{HX} \rightleftharpoons \equiv \text{Ti} - \text{OH} + \text{X} \equiv \text{Ti} \quad (13) \]

Arnal et al. [111] subsequently examined the solution chemistry of the alkyl halide elimination (Eqn 12) and the etherolysis (Eqn 9) by means of NMR spectroscopy. They found that the etherolysis (Eqn 9) of TiCl₄ with disisopropyl ether (\(\text{Pr}_2\text{O}\)) yielding Ti(O\(\text{Pr}\))Cl₃ is quite fast even at room temperature, while the second alkoxylation step is much slower and is not completed when condensation (Eqn 12) starts. The main reaction that takes place at room temperature in a mixture of TiCl₄ and TIP is the redistribution of O\(\text{Pr}\) and Cl ligands. In both cases (etherolysis or direct condensation), the condensation reaction is slow at room temperature, and it takes place at 100 °C after an induction period. Thus, the true precursors in both nonhydrolytic routes (alkoxide or ether) are titanium chloropropoxide species TiCl₄\(\equiv\text{O}\)\(\equiv\text{Ti} - \text{Cl}_3\equiv\text{O}\)\(\equiv\text{Pr}_3\) in equilibrium. Another interesting point is the autocatalytic nature of the condensation reaction. Usually, such behavior is explained by the reaction catalyzed by the reaction products. However, the condensation rate was not modified when the reaction was performed in the presence of titania or isopropyl chloride, the final products of the condensation. Therefore, Arnal et al. concluded that the intermediate oxochloropropoxide species (Cl₃\(\equiv\text{O}\))\(\equiv\text{Ti} - \text{O} \equiv \text{Cl}_3\equiv\text{O}\)\(\equiv\text{Pr}_3\) are much more reactive toward condensation than the starting chloropropoxide, possibly because they are better Lewis acids [111]. These results clearly show that complex chemistry exists in these reactions, although the overall reactions are simple.

As for the ether elimination reaction (Eqn 10), Mutin and Vioux [11] wrote in their recent review that ether elimination has been postulated in the formation of metal oxoalkoxides from alkoxides. They also noted that methyl and benzyl groups favor the ether elimination route. Niederberger and Garnweitner [112] found that the reaction of TIP in benzyl alcohol [\(\text{C}_6\text{H}_5\text{CH}_2\text{OH}\); \(\text{bp} = 205 ^\circ\text{C}\); causes severe eye irritation and may cause skin irritation] at 200 °C proceeds by ether elimination mechanism, yielding dibenzyl ether together with crystalline anatase.

As mentioned in Section 3.2.1, we proposed a heterolytic-cleavage mechanism for thermal decomposition of metal alkoxide (Eqn 6). In alcoholic media, this reaction proceeds more easily because the intermediates, carbocation and metalxo anion, are stabilized by salvation. The overall reaction of TIP can be written as follows:

\[ 2\equiv\text{Ti} - \text{O} - \text{CH}(\text{CH}_3)_2 \rightarrow \equiv\text{Ti} - \text{O} \equiv + (\text{CH}_3)_2\text{CHOH} + \text{CH}_3\text{CH} = \text{CH}_2 \quad (14) \]

Since the benzyl group cannot give olefinic products, the cation formed from this group reacts with alkoxide anion (RO⁻) yielding ether molecules.

Following intensive studies on the formation of oxide gel by the direct reaction between metal halides and alcohols or alkoxides [110,111], Trentler et al. [113] in 1999 reported synthesis of “hydroxyl-free” anatase nanocrystals based on the alkyl halide elimination process (Eqn 12). Titanium halide was mixed with TOPO in heptadecane [\(\text{C}_6\text{H}_5\text{CH}_2\text{OH}\); \(\text{bp} = 301.8 ^\circ\text{C}\)] and heated to 300 °C under dry nitrogen and a titanium alkoxide was then rapidly injected into the hot solution. Anatase with a crystallite size of less than 10 nm in diameter was obtained, although there was considerable size distribution. However, the yields of the products were less than 50%. The reaction rate dramatically increased with greater branching of the alkyl group of the alkoxide, which is in good agreement with the reactivity of metal alkoxides for thermal decomposition as discussed in Section 2.1.

Moreover, the boiling points of titanium halides are relatively low (TiCl₄, 130 °C; TiBr₄, 231 °C), and the reaction rate was not affected by the nature of X. The thermal decomposition of titanium alkoxide probably took place under their reaction conditions, and hydrogen halides liberated from TiX₄ is facilitate the decomposition of titanium alkoxide, while Ti atoms from TiX₄ seem not to be incorporated in formation of anatase.

Trentler et al. made three important points: (1) Use of a high-boiling-point organic solvent such as heptadecane, which permits the high-temperature reaction under ambient pressure, (2) use of a capping (passivating) agent such as TOPO, which prevents the coagulation of nanoparticles during the synthesis, and (3) injection of the reactant to the preheated organic solvent, so that a huge number of nuclei of the product are formed. As mentioned in Section 1, the effect of pressure is negligible, and use of an open vessel (glass apparatus) or closed vessel (autoclave) is not particularly important unless the low-boiling-point by-products affect the reaction: In this case, the reaction in an open vessel seems to facilitate the elimination of surface hydroxyl groups. For the second point, the use of phosphine-based capping agents possibly limits the application of the product. For the third point, injection of a large quantity of the reactant is impossible; therefore, this method cannot be applied for large-scale synthesis. Nevertheless, this paper had an enormous impact upon researchers in material chemistry, and this methodology is
now called an “injection method” and has been applied for synthesizing various metal oxides.

Jun et al. [114] used a similar method for preparing anatase nanocrystals: Dioctyl ether \([\text{CH}_3(\text{CH}_2)_7\text{O}(\text{CH}_2)_{7}\text{CH}_3]; \text{bp} \approx 286–287 \degree\text{C}\) was used as a solvent and lauric acid \([\text{CH}_3(\text{CH}_2)_{10}\text{COOH}; \text{mp} = 44.8 \degree\text{C}; \text{bp} = 298.9 \degree\text{C}\) and TOPO as selective and nonselective surface capping agents, since the former capping agent selectively binds to \{001\} faces through a bridge-bonding mode. The effect of the relative ratio of two types of capping agents was examined and it was found that truncated octagonal pyramids were formed when lauric acid was added to the reaction system. Long rods and branched rods were also formed (Figure 6).

Koo et al. [115] found that when a solution of TiCl\(_4\) in 1-octadecene \([\text{CH}_3(\text{CH}_2)_{15}\text{CH}==\text{CH}_2]; \text{bp} \approx 315 \degree\text{C}\) and a solution of TIP in the same solvent were simultaneously injected into preheated oleylamine using two syringe pumps, the injection rate of the two solutions affected the morphology as well as the crystal structure: The decreased injection rate preferred the formation of rutile.

\[
\text{CH}_3(\text{CH}_2)_7\text{NH}_2
\]

Oleylamine; \text{bp} = 196 \degree\text{C}/15 \text{mmHg}

Causes skin burns and may be harmful

In contact with the skin

Following the procedure of Trentler et al. [113], Vukičević et al. [116] prepared an aqueous suspension of anatase nanorods (1-dodecylamine \([\text{CH}_3(\text{CH}_2)_{14}\text{NH}_2]; \text{bp} = 247–249 \degree\text{C}; \text{causes skin burns}\) was also used as the capping agent together with TOPO). After the reaction, the products were dispersed in CHCl\(_3\) and irradiated with UV light in the presence of water. Photocatalytic decomposition of the capping agents causes migration of anatase nanorods to the aqueous layer.

2. Reaction of titanium halide in alcohols

Wang et al. [117] reported the reaction of TiCl\(_4\) in various alcohols at 100 \degree\text{C} and 160 \degree\text{C}. Methanol, ethanol, 1-propanol, and 2-propanol gave anatase, and butanol and octanol yielded rutile, while ethylene glycol yielded a mixture of rutile and anatase. The products consisted of spherical particles, but agglomerated tenuous fibers were obtained in octanol (rutile) and 2-propanol (anatase). The quite low crystallization temperature of titania and the low decomposition temperature of the alkoxides derived from TiCl\(_4\) are mediated by HCl liberated from TiCl\(_4\). They also noted that the phase formed by the reaction is affected by the concentration of HCl, a lower concentration of HCl favoring the formation of anatase.

Niederberger et al. developed a nonhydrolytic route by the use of benzyl alcohol [118–121], and titania can be crystallized at temperatures as low as 40 \degree\text{C} [112,118]. Benzyl cation is stabilized by delocalization of the positive charge to the phenyl ring, and therefore the cleavage of C–O bond of this alcohol proceeds facilely.

They also functionalized the surface of the nanoparticles of titania by the addition of catechol ligands such as dopamine and 4-tert-butylcatechol to the reaction mixture. This was achieved because the TiCl\(_4\)/benzyl alcohol system proceeds at low reaction temperatures. This reaction provides a simple route to surface-functionalized nanoparticles that are dispersible either in water or in organic solvents [120]. They also found that when the same reaction is performed in the presence of 2-amino-2-(hydroxy-methyl)-1,3-propanediol [(HOCH\(_2\)\(_2\)CNH\(_2\)] or 2-amino-1,3-propanediol [(HOCH\(_2\)\(_2\)CHNH\(_2\)], the as-synthesized titania nanocrystals assemble into nanowires upon redispersion in water. These nanowires are composed of a continuous string of precisely ordered nanoparticles assembling along the [001] direction through oriented attachment [121]. They concluded that the anisotropic assembly is a consequence of the water-promoted desorption of the organic ligands selectively from the \{001\} faces of the crystalline building blocks together with the dissociative adsorption of water on these crystal faces. These processes induce the preferred attachment of the titania nanoparticles along the [001] direction (Figure 6) [122].

One shortcoming of this method is that the product was contaminated with chloride ions. Li et al. [123] obtained phase-pure anatase nanocrystals by mixing of TiCl\(_4\) with ethanol (>97%; both ethanolysis and hydrolysis of TiCl\(_4\) take place) at 0 \degree\text{C} followed by gelation at \text{<}100 \degree\text{C}. To reduce impurity concentrations in the product within the ppm level, they repeatedly rinsed the products several times a day for up to 60 days.

3. Titanium halide in alcohol/acid mixed solvents

Wang et al. [124] examined the reaction of TiCl\(_4\) in a mixed solvent of acetic acid and alcohol (ethanol, ethylene glycol, or glycerin); anatase (ethylene glycol)
and rutile (ethanol) are obtained. They concluded that the formation of titania is based mainly on the hydrolysis process initiated by the water generated as a result of the esterification reaction between alcohols and acid as discussed in Section 3.2.2. They also suggested that the viscosity and pressure of the reaction media influence the particle size of the product.

4. Reaction of titanium alkoxide in carboxylic acids and anhydrides

Titanium alkoxides readily react with carboxylic acid under mild conditions [125], and this process is believed to be the first step of the catalytic action of titanium alkoxides for esterification:

\[ \text{M} - \text{OR} + \text{R}’\text{COOH} \rightarrow \text{M} - \text{OCOR’} + \text{ROH} \quad (15) \]

Cozzoli et al. [90] obtained rod-shaped anatase nanoparticles by the reaction of TIP in oleic acid at 80 °C in the presence of tertiary amine or quaternary ammonium hydroxides. Kim et al. [126] reported that dumbbell-shaped nanorods were formed by the decomposition of TIP in toluene in the presence of oleic acid at 250 °C. In this reaction, ligand exchange between TIP and oleic acid followed by ester elimination reaction may take place [126].

Joo et al. [127] synthesized anatase nanorods (3.4 nm in diameter, 38 nm in length) by the reaction of TIP with oleic acid at 270 °C. The addition of 1-hexadecylamine \([\text{CH}_3(\text{CH}_2)_{15}\text{NH}_2];\text{bp} = 330^\circ\text{C}\) reduced the diameter of the nanorods, elongating them. They attributed this result to the preferential adsorption of the amine on the (101) surface of anatase. An ester was detected by GC–MS confirming the ester elimination process. Since they observed the bending vibration mode of the Ti–OH bond in the IR spectrum of TIP–oleic acid complex obtained at 100 °C, they proposed the reaction paths as depicted by the following equation:

\[ \text{M} - \text{OR} + \text{R}’\text{COOH} \rightarrow \text{M} - \text{OH} + \text{R}’\text{COOR} \quad (16) \]

They also showed that surface oleic acid can be removed by reduction with lithium triethylborohydride and the stripped anatase nanorods were shown to exhibit superior photocatalytic inactivation of *Escherichia coli* to that of Degussa P-25 [127].

However, the reaction mechanisms seem not to be so simple as suggested by these papers. Since alcohol molecules are liberated from the titanium alkoxide according to Eqn (15), the thus-formed alcohol reacts with carboxylic acid yielding ester and water because of high catalytic activity of titanium alkoxide (see Section 3.2.2). Therefore, the reaction route, Eqn (15) followed by Eqn (11), seems to be accompanied by hydrolytic reactions. The Ti–OH groups detected by Joo et al. [127] may be formed by the reaction of TIP with water.

De Marco et al. [128] reported transparent mesoporous films of anatase nanorods for dye-sensitized solar cell application. Anatase nanorods, prepared by the reaction of TIP in a mixture of acetic acid and benzyl alcohol at 150 °C for 8 h, were centrifuged without complete removal of the residual organics, and screen-printable paste was prepared by the aid of polymer binders.

Zhang et al. [129] proposed an aminolysis route (Eqn 17) for the reaction of TIP with oleic acid and oleylamine in the presence of 1-octadecene as a solvent. They injected TIP to the preheated mixture of oleic acid and 1-octadecene and evaporated 2-propanol formed by Eqn (15). After the mixture was heated at 260 °C, oleylamine was injected. Escape of 2-propanol from the reaction system seems to be the reason why the ester elimination (Eqn 11) did not take place under their reaction conditions. However, direct evidence for the formation of amide was not presented in their paper:

\[ \text{M} - \text{OCOR} + \text{R}’\text{NH}_2 \rightarrow \text{M} - \text{OH} + \text{R}’\text{COONHR’} \quad (17) \]

An interesting extension of the ester elimination route (Eqn 11) would be to generate the acyl groups in situ, as shown in the following equation [130]:

\[ \text{M} - \text{OR} + \text{R}’\text{COOCOR’} \rightarrow \text{M} - \text{O} - \text{COR’} + \text{R}’\text{COOR} \quad (18) \]

Thus, the reaction alkoxide with acid anhydride would give carboxylate, and the reaction of the thus-formed carboxylate with alkoxide would give metal oxide via ester elimination. This route does eliminate the possibility for participation of a hydrolytic reaction. However, Arnal et al. [110] found that the reaction of TIP with acetic anhydride (Eqn 18→11) hardly occurred at 140 °C and only proceeded in the presence of a Lewis acid catalyst such as TiCl₄ [110]. Therefore, Guo et al. [131] examined the reaction of TIP with acetic anhydride (or trifluoroacetic anhydride) in the presence of a catalytic amount of TiCl₄. The reaction was carried out in supercritical CO₂ at 110 °C for 20 h and the products were recovered by depressurizing CO₂. However, the products were amorphous containing significant amounts of organic groups (~40 wt%). Calcination of the product gave anatase crystals [131], but a hydroxyl-free surface of anatase is not guaranteed because of the formation of water by the combustion of the organic groups.

3.2.5. Hydrolysis of Titanium Alkoxide by Water Dissolved in Organic Solvent From Gas Phase and Solvothermal Crystallization

Since primary alkoxide of zirconium does not decompose in toluene, hydrolysis of the alkoxide with water dissolved in this solvent from the gas phase followed by solvothermal
crystallization of the hydrolyzed product was examined [132]. In this method, the alkoxide solution in an inert organic solvent is placed in a test tube, which is then placed in an autoclave as described in the footnote of Figure 1. The desired amount of water is placed between the test tube and the autoclave wall. When the autoclave is heated, the water evaporates and is dissolved into the inert organic solvent from the gas phase, where hydrolysis of the alkoxide takes place, followed by crystallization of the hydrolyzed products. This procedure can avoid the heterogeneity caused by the addition of water. Using this method, monoclinic zirconia with a large surface area was prepared [132].

Kominami et al. [133–135] further extended the application of this method and found that titania prepared by this method has a high photocatalytic activity because the product has a large surface area as well as a low number of crystal defects that can act as the recombination sites for holes and electrons generated by photoactivation. They called this method the HyCOM method [134], and applied it for synthesizing various metal oxides [136, 137, 189].

Dinh et al. [138] used the same geometry of the apparatus for synthesizing monodispersed nanocrystals of anatase with different morphologies such as rhombic, truncated rhombic, spherical, dog-bone, and bar. A mixture of titanium n-butoxide, oleic acid, and oleylamine was placed in a cup, while the azeotropic mixture of ethanol-water was placed in an autoclave. The morphology of the product was controlled by reaction temperature (140–180 °C) and titanium n-butoxide/oleic acid/oleylamine ratio. The presence of water vapor plays a critical role in the morphology of the product, and Dinh et al. believe that the initial step is the hydrolysis of the alkoxide yielding unstable hydroxylalkoxides ((RO)3TiOH), while the subsequent condensation reaction takes place in a nonhydrolytic manner [138].

Injection of titanium precursor to a solution containing water is an opposite route. Thus, Wahi et al. [139] injected the solution of titanium tetraethoxide in dry ethanol to a preheated ethanol solution containing a known amount of water in an autoclave using a pressure pipette. High-surface-area anatase nanocrystals were formed by this method.

3.2.6. Solvothermal Reaction in a Solvent Containing a Small Amount of Water

Lee et al. [140] reported that a colloidal solution containing nano-sized TiO2 (anatase) particles was obtained by solvothermal treatment of TIP in 1-butanol at 200 °C in the presence of a small amount of aqueous HCl. They found that the product particles, coated on γ-alumina, showed excellent performance for photocatalytic decomposition of CHCl3.

Lin et al. [141] examined the reaction of TIP in alcohol in the presence of ionic liquid at 210 °C. They found that the reaction proceeds via intermediate formation of titanium-oxo clusters [Ti4O2(OEt)20 in the case of ethanol solvent] and only a small amount of water (2%) was sufficient for the formation of anatase nanocrystals [141].

3.2.7. Solvothermal Crystallization of Titania From Amorphous Gel or Sol Obtained by Precipitation Method

The general procedure is as follows: Amorphous sol or gel is obtained by the precipitation method from a suitable starting material; the precipitate is dispersed in an organic solvent, and the mixture is solvothermally treated. One drawback of this procedure is that the amount of water in the solvothermal system is not known unless the water content in the precipitate is measured. Washing the precipitate before solvothermal treatment is recommended so that foreign ions are not incorporated into the product. Moreover, if ammonia is used as a precipitant of metal nitrate, ammonium nitrate is formed, which can cause an explosive reaction in organic solvents.

Wu et al. [142] examined the effects of alcoholic solvents (methanol, ethanol, 2-propanol, and 1-butanol) on the solvothermal crystallization of titania sol obtained by hydrolysis of titanium n-butoxide. The crystallinity of anatase depended on the solvent used, and they concluded that 2-propanol having a low dielectric constant and thus a lower solubility of the product accelerates the crystal growth of anatase. We, however, think that the activity of water is the important factor; the activity of water increases with decreasing the polarity of the solvent, thus accelerating the crystal growth of the products. This argument is based on the solvothermal reaction of gibbsite in various alcohols [143]. For a detailed discussion, see our previous review article [144].

Hydrothermal synthesis of metal oxides from the corresponding hydroxides usually requires high temperatures because of the limitations of equilibrium:

\[ \text{M(OH)}_n \rightleftharpoons \text{MO}_{n/2} + n/2\text{H}_2\text{O}. \]

Solvothermal dehydration can avoid this limitation, and dehydration may proceed at a temperature much lower than that required by the hydrothermal reaction. However, thermal dehydration may compete with the solvothermal reaction. When dehydration of hydroxides starts, water formed by dehydration of the starting material may facilitate hydrothermal transformation of the starting material. Therefore, complicated reactions may occur simultaneously.

3.3. Rare Earth Garnets

Rare earth (RE) garnets, especially yttrium aluminum garnet (YAG), are now widely used in lasers and phosphors. The thermodynamic stabilities of the garnet phases depend on
the ionic size of the RE element, and RE aluminum garnets (REAGs) were reported to be thermodynamically stable for all the RE elements from Tb to Lu [145], while GdAG is a metastable phase. Synthesis of single-phase REAG by the ceramic process from mixed alumina and RE oxide powders usually requires a temperature higher than 1600 °C with a prolonged heating period [146]. Homogeneous mixing of aluminum and RE atoms in the starting materials can lower the crystallization temperature of REAG [147–150], but these materials still require calcination temperatures higher than 800 °C to crystallize the REAG phase. Hydrothermal synthesis of single-phase REAG requires higher temperatures (350–600 °C) and pressures (70–175 MPa) [151,152], although Mill reported that the lowest temperature for the formation of YAG was 280 °C [151]. He also reported that with an increase in the ionic size of the RE element, the lowest temperature increased. The REAG with the largest RE ion size that has been hydrothermally prepared so far is TbAG [151], and it was reported that 420–450 °C was required for the formation of this garnet.

### 3.3.1. Glycothermal Reaction

Since aluminate anion (>Al–O–) is expected to be formed by the decomposition of the glycoxide intermediate derived from aluminum alkoxide and 1,4-butanediol (see Figure 5), the presence of metal cation that gives basic oxides would give M–O–M’ bonds. According to this working hypothesis, Inoue et al. examined the reaction of aluminum isopropoxide (AIP) with yttrium acetate (Y(OAc); Ac = CH₃CO) in 1,4-butanediol at 300 °C and found the formation of crystalline yttrium aluminum garnet (YAG) [6]. The hydrothermal reaction of pseudoboehmite (microcrystalline boehmite, the hydrolyzed product of AIP) with yttrium acetate at 300 °C yielded boehmite together with a small amount of YAG. Single-phase YAG was not obtained, even with prolonged reaction time [6,153]. The difference between glycothermal and hydrothermal reaction can be attributed to the different stabilities of the intermediate phases, that is, the glycol derivative of boehmite vs. well-crystallized boehmite (see, Figure 4). Similarly, the reaction of the stoichiometric mixture of AIP and RE (Gd–Lu) acetate (RE(OAc)₃) gives the corresponding rare earth aluminum garnet (REAG) in single phase [153]. Therefore, all the thermodynamically stable REAGs together with metastable GdAG were prepared by the glycothermal method.

The glycothermal reaction of RE(OAc)₃ alone yields RE(OH)₂(OAc), REO(OAc) (two polymorphs), and RE(OH)(OAc)₂, depending on the ionic size of the RE ion [154,155]. The acetate ions are not completely expelled from the coordination sites of the RE ion. However, in the presence of aluminum alkoxide as the starting material, acetate ions are fully eliminated from the product. Therefore, the aluminate ion facilitates the cleavage of the bond between acetate and RE ions.

It is interesting to note that the reaction of Sm(OAc)₃ or Eu(OAc)₃ with AIP produced SmAG or EuAG, although the product was contaminated with RE acetate oxide (REO(OAc)) [153]. EuAG was recently prepared by a sol–gel method followed by calcination at 800–850 °C [156], but SmAG has never been prepared by other methods. These results can be interpreted by the instability of the intermediate phase(s); it is possible that all the metastable phases having free energies lower than the intermediate can be formed from the thermodynamical point of view.

The reaction of RE(OAc)₃ (Nd–Lu) with gallium acetylacetonate (Ga(acac)₃) in 1,4-butanediol at 300 °C yielded the corresponding RE gallium garnets (REGGs) [26,157]. The garnet phases were reported to be thermodynamically stable for all the RE elements from Sm to Lu [158], and all of the stable REGGs were prepared by the glycothermal method (Figure 7).

The reaction at 270 °C also gave phase pure REGGs for Sm–Lu, but the reaction at 250 °C gave amorphous products except for the reaction of yttrium, which gave yttrium gallium garnet (YGG). The hydrothermal reaction of Ga(acac)₃ with RE acetate under conditions identical to the glycothermal reaction (except reaction medium) gave γ-Ga₂O₃ together with a small amount of the garnet phase [26].

The reaction of Pr(OAc)₃ and Ce(OAc)₃ with Ga(acac)₃ gave carbonate hydroxide (RECO₃(OH)) as the sole crystalline product. When the reaction was carried out in the presence of gadolinium garnet (GGG) seed crystals, garnets (PrGG and CeGG) were crystallized in spite of the fact that cell parameters of these garnets are much larger than those of GGG. [26]. These results suggest that under glycothermal conditions, nucleation is the most difficult process and that, once nucleation takes place, crystal growth proceeds easily.

![FIGURE 7](image-url) Synthesis of monodispersed particle of rare earth gallium garnets.
The particles of gallium garnets with large RE ions are spherical (0.5–2 μm) and the surface of the particles is smooth. On the other hand, the surface of the particles (0.1–0.3 μm) of the gallium garnets of Tb and RE elements having ionic sizes smaller than Tb is rough, with apparent polycrystalline outlines. However, high-resolution images of the latter type of particles show that a whole particle is covered with a single lattice fringe, indicating that each particle is a single crystal grown from only one nucleus. Inoue et al. concluded that the latter type of morphology is formed by quite rapid crystal growth [26]. For these garnets, monodispersed particles were obtained (Figure 7). Monodispersed particles can be prepared if a burst of nuclei is formed at the early stage of the reaction and if nucleation does take place during the crystal growth stage [159]. Once nucleation occurs in the glycothermal synthesis of the latter class of gallium garnets, the concentration of the intermediate in solution decreases, which is determined by the balance between the dissolution rate and the rate of consumption of the reactants in the solution by crystal growth. Since the crystal growth of the garnet with small RE ions proceeds rapidly, the concentration of the reactants in the solution becomes low, prohibiting nucleation during the crystal growth stage, thus leading to monodispersed particles.

The morphology of the products suggests that defects are frequently incorporated into the products. This point seems to be closely connected with an important feature of solvothermal products. In hydrothermal reactions, dissolution–deposition equilibrium takes place. Dissolution of adatoms on the surface occurs selectively, while preferential adsorption of ions at the vacancies of the surface proceeds. Therefore, a nearly perfect surface is formed by the hydrothermal reaction. On the other hand, under solvothermal conditions, dissolution of oxide materials into the organic solvent barely takes place, and therefore the product usually contains various types of crystal defects.

The defect structure of the glycothermally prepared REAG was examined recently [160]. The unit cell parameters of REAGs were significantly larger than the values reported in JCPDS cards. Rietveld analysis of the products indicates the presence of Al vacancies in the 24d sites and oxygen vacancies in the 96h sites. Partial substitution of Al ions in the 16a sites with RE ions (antisite defects) was also suggested, which contributes to the enlargement of lattice parameter. The occupancy of RE ions in the 16a sites increased with the decrease in the ionic size of the RE element [160].

Allison et al. [161] prepared YAG:Ce (30 nm size) by the glycothermal method: Ytrium acetate, cerium acetate, and aluminum tris(2-methoxyethoxide) prepared by the reaction of AIP with 2-methoxyethanol were used as the starting materials, and the reaction was carried out in 1,4-butanediol in the presence of oleic acid and diphenyl ether at 305 °C for 12 h. The effects of the starting material and additives on the physical properties of the products were not clarified; however, Allison et al. reported that the fluorescence decay lifetime of the thus-prepared nanophosphor varied with temperature [161].

Nishi et al. [162] prepared YAG:Erions by the glycothermal method and examined the fluorescence properties in the optical-telecommunication band. The emission intensity and lifetime increased with the increase in the heat treatment temperature and time, which is attributed to the removal of organic matter remaining in the glycothermal product and the decrease in the specific surface area. They concluded that Er ions in the surface region with a thickness of about 10 nm do not work well as luminescence centers due to the defects in the host material [162].

Kasuya et al. [163] prepared YAG:Ceions by the glycothermal method and examined the photoluminescence properties of the product. Contrary to the case of YAG:Erions, the photoluminescence intensity of this product decreased by calcination [163,164]. The internal quantum efficiency significantly increased when poly(ethylene glycol) (PEG) was added to the glycothermal system and PEG was found to be preferentially adsorbed on the phosphor particles preventing the oxidation of Ceions to Ceions. Kasuya et al. also showed that the addition of PEG increased the tetrahedral/octahedral ratio of Alions, which also contributes to the enhancement of photoluminescence intensity [163]. Photoluminescence intensity is also increased by prolonging the reaction time, and this result is attributed to the homogeneous incorporation of Ceions in the lattice of YAG, since the atomic composition of the product was not altered [165]. Interestingly, Kasuya et al. [165] showed that tetrahedral Alions are not present near the surface of the YAG:Ceions phosphor by comparison of single-pulse and cross-polarization MAS 27Al NMR spectra. The same group also reported that the addition of citric acid to the glycothermal system significantly decreased the particle size of YAG:Ceions phosphor and increased its photoluminescence intensity [166].

Kasuya et al. also reported the synthesis of a transparent color-conversion thick film of YAG:Ceions prepared by the glycothermal method. The colloidal suspension of the product was placed in a refrigerator for 2 weeks to settle the coarse particles. To the supernatant, ethanol/methanol (95/5) was added and the mixture was centrifuged yielding a transparent yellow paste, which was placed inside the gap between the slide and cover glasses [167]. This group also used repeated glycothermal treatment so that the nanoparticles grow with the increase in the crystallinity. In this procedure, the particles formed by the first glycothermal reaction act as the seed crystals for the second glycothermal reaction as mentioned previously. Thus, they succeeded in suppressing the photobleaching of the YAG:Ceions phosphor caused by oxidation of Ceions [168]. Biological
application of the glycothermally prepared YAG:Ce$^{3+}$ phosphor was also reported [169,170].

Nyman et al. [171] found that the YAG:Ce$^{3+}$ phosphor prepared in 1,4-butanediol at 225 °C has a higher photoluminescence efficiency and explained the result by suggesting that the phosphor is surface-passivated with the glycol derivative of boehmite formed as the intermediate. They also explored various glycols for synthesizing the phosphor at that temperature: 1,4-Pentanediol and 1,5-pentanediol give amorphous products, and the product obtained in 1,3-butanediol does not emit the yellow-green luminescence, while diethylene glycol results in non-luminescence precipitates [171]. The Ce$^{3+}$ emission of (Y,Tb)AG:Ce and (Y,Gd)AG:Ce prepared by a similar glycothermal method was reported to be red-shifted due to the increase in crystal field strength around Ce$^{3+}$ ion [172].

3.3.2. Solvothermal Crystallization of Precipitated Gel

Chinese groups use the co-precipitation—solvothermal route for the preparation of garnets. A stoichiometric mixture of rare earth nitrate and aluminum nitrates was precipitated with a suitable precipitant such as aqueous ammonia [173,174] and ammonium hydrogen carbonate [175,176], and the precipitates were washed. Then, the precipitates were dispersed in aqueous ethanol [173,177] or aqueous ethylenediamine [174,178] and then solvothermally treated. It was reported that YAG was crystallized by solvothermal treatment at 280 °C, while the hydrothermal treatment of the same starting material at 300 °C resulted in the formation of a mixture of YAG, γ-Al$_2$O$_3$, YOOH, and YAIO$_3$ [173]. Using these techniques, these groups reported the formation of YAG [173,179], YAG:Ce [175], YAG:Eu [176], YAG:Tb [178], and LuAG [174].

In this method, the synthesis of homogeneous precipitate seems to be the most important point, as the authors noted that the reverse-strike method (i.e., addition of the metal salt solution to the precipitant solution so that the pH of the solution is kept alkaline during the course of the precipitation) has the advantage of higher cation homogeneity [178]. Based on close examination of the TEM images of the products reported [175–178], however, I think that the reproducibility of this method would be rather low.

3.3.3. NonHydrolytic Sol–Gel-Like Method

Su et al. [180] report the synthesis of YAG:Ce nanophosphor with a size of 250 nm by the thermal decomposition of the solution of mixture of metal isopropoxides at 250 °C in the presence of TOPO and oleic acid stabilizers. However, the details of the experimental procedure and characterization of the product were not reported. The same group also reported the synthesis of the phosphor of 5 nm size by the thermal decomposition of metal carbonate–oleate complex at 300 °C; thus, a mixture of yttrium, cerium, and aluminum carbonates prepared by precipitating from the mixed chloride solution with ammonium carbonate was dissolved in oleic acid with vigorous stirring. To the solution, oleylamine was added and the mixture was solvothermally treated for 6 days [181].

3.3.4. Thermal Decomposition of Metal Nitrates in Alcohol

Metal nitrates (Gd, Y, and Ce) and AIP dissolved in 2-propanol were thermally treated in an autoclave at 230 °C for 5 h. This procedure gave an amorphous product, and (Gd$_0.9$Y$_{0.1}$)$_3$Al$_5$O$_{12}$:Ce was prepared by calcination (>1000 °C) of the product [182]. Using a similar method, YAG:Eu [183], GdAG:Dy [184], (Gd$_0.9$Y$_{0.1}$)$_3$Al$_5$O$_{12}$:Tb [185], and GdAG:Tb [186] have been prepared.

In this method, metal alkoxide may be formed according to the following equation:

$$M(NO_3)_3 + 3(CH_3)_2CHOH \rightarrow M(OCH(CH_3)_2)_3 + 3HNO_3$$

Alkoxides may be hydrolyzed by water formed by thermal dehydration of the secondary alcohol, 2-propanol. Because of the difference of the hydrolysis rate of the alkoxides, the homogeneity of metal cations in the product is not attained, and high calcination temperature is required, while an undesired phase such as orthorhombic phase may be contaminated in the products [182,184,186]. Although metal nitrate itself decomposes endothermically, the presence of organic compound in the reaction system may cause an exothermic reaction (see, Section 2.2). Therefore, the reaction of metal nitrates should not be carried out with an organic compound in a sealed vessel.

4. CONCLUDING REMARKS

Metal oxides can be synthesized by various solvothermal techniques: solvothermal dehydration of metal hydroxides, solvothermal decomposition of metal alkoxides, solvothermal synthesis of mixed oxides, solvothermal crystallization of amorphous oxides, solvothermal ion exchange or intercalation, and solvothermal oxidation of metals. Although some of the reaction products may not be oxides, but hydroxides, oxyhydroxides, acetates, or organic—inorganic hybrid materials, oxides can be prepared by calcination of these products. Most of the solvothermal products have fine particle sizes and therefore the oxides derived from these products have physical properties significantly different from those of metal oxides obtained by conventional methods. In this article, solvothermal routes for synthesizing oxides are reviewed, addressing the differences and similarities between solvothermal and hydrothermal reactions.
As the hydrolysis of metal alkoxides proceeds exothermically, the reaction would appear to be irreversible. However, metal alkoxide and water are in equilibrium with hydroxide and alcohol [99], and this equilibrium can be attained at high temperature with high-boiling-point alcohols:

$$\text{M(OH)}_n + n\text{ROH} \rightleftharpoons \text{M(OR)}_n + n\text{H}_2\text{O}$$

The glycothermal reaction of aluminum alkoxide yields aluminum glycoxide by alkoxyl exchange reaction and the thermal decomposition of the intermediate yields the final product, i.e., the glycol derivative of boehmite. The structure and physical properties of the product are essentially identical to those obtained by the reaction of aluminum hydroxide in the same glycol. This is achieved by the above-mentioned equilibrium. This means that partially hydrolyzed metal alkoxide may be used in the glycothermal reaction. However, fully hydrolyzed alkoxide may give completely different results.

Solvothermal reaction usually gives nanoparticles, but the toxicity of the nanoparticles is unknown. Since nanoparticles may pass through the cell membranes and become internalized within the cell, impairing cellular functions, the toxicity of nanoparticles is completely different from micrometer-size particles or bulk materials [187,188]. If nanoparticles are stabilized by a capping agent, the agent may cause the cytotoxicity, because of its presence at a high level on the surface of nanoparticles [188].

REFERENCES

[22] Yu S-H, Yoshimura M. Shape and phase control of ZnS nanocrystals: template fabrication of wurtzite ZnS single-crystal nanosheets and ZnO flake-like dendrites from lamellar molecular precursor ZnS-(NH2CH2CH2NH2)0.5. Adv Mater 2002;14:2963—900.
[28] As the hydrolysis of metal alkoxides proceeds exothermically, the reaction would appear to be irreversible. However, metal alkoxide and water are in equilibrium with hydroxide and alcohol [99], and this equilibrium can be attained at high temperature with high-boiling-point alcohols:

$$\text{M(OH)}_n + n\text{ROH} \rightleftharpoons \text{M(OR)}_n + n\text{H}_2\text{O}$$

The glycothermal reaction of aluminum alkoxide yields aluminum glycoxide by alkoxyl exchange reaction and the thermal decomposition of the intermediate yields the final product, i.e., the glycol derivative of boehmite. The structure and physical properties of the product are essentially identical to those obtained by the reaction of aluminum hydroxide in the same glycol. This is achieved by the above-mentioned equilibrium. This means that partially hydrolyzed metal alkoxide may be used in the glycothermal reaction. However, fully hydrolyzed alkoxide may give completely different results.

Solvothermal reaction usually gives nanoparticles, but the toxicity of the nanoparticles is unknown. Since nanoparticles may pass through the cell membranes and become internalized within the cell, impairing cellular functions, the toxicity of nanoparticles is completely different from micrometer-size particles or bulk materials [187,188]. If nanoparticles are stabilized by a capping agent, the agent may cause the cytotoxicity, because of its presence at a high level on the surface of nanoparticles [188].


[118] Vuki...[117


[143] Inoue M, Kitamura K, Tanino H, Nakayama H, Inui T. Alcoho-


Chapter 11.1.5

Supercritical Hydrothermal Synthesis

Tadafumi Adschiri,* Seiichi Takami,** Toshihiko Arita,** Daisuke Hojo,* Kimitaka Minami,† Nobuaki Aoki* and Takanari Togashi†

† World Premier International Research Center-Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan, ‡ Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan, † New Industry Creation Hatchery Center, Tohoku University, 6-6-10 Aoba, Aramaki, Aoba, Sendai, Miyagi 780-8577, Japan,

Chapter Outline

1. Introduction 949
2. Basic Principles of Supercritical Hydrothermal Synthesis 950
3. Apparatus 953
4. NP Synthesis by Supercritical Hydrothermal Synthesis 956
5. Supercritical Hydrothermal Synthesis of Organic—inorganic Inorganic Hybrid Nanoparticles 957
6. Self-assembly of Hybrid Organic—inorganic Nanoparticles 965
7. Hybrid Nanomaterials 972
8. Summary 974
References 974

1. INTRODUCTION

The nanomaterials are known for their unique mechanical, chemical, physical, thermal, electrical, optical, electronic, magnetic, and specific surface area properties. Various inorganic and organic compounds have been fabricated in nanoscales as tubes, ribbons, spheres, wires, belts, dots, fibers, dendrites, mushrooms like, etc. Most commonly used methods are physical vapor deposition, colloidal chemistry approach, mechanical alloying, sol—gel, mechanical grinding, hydrothermal method, laser ablation, chemical vapor deposition (CVD), electrodeposition, plasma synthesis, microwave techniques, etc.

In nanotechnology, hydrothermal processing has an edge over the other conventional processes, because it facilitates the issues like simplicity, cost effectiveness, energy saving, use of large volume of equipment (scale-up process), better nucleation control, pollution-free (since the reaction is carried out in a closed system and in “green” water), higher dispersion, higher rate of reaction, better shape control, and lower temperature operation in the presence of an appropriate solvent, etc.

The hydrothermal technique has a lot of other advantages such as the following: it accelerates interactions between solid and fluid species; phase pure and homogeneous materials can be achieved; reaction kinetics can be enhanced; the hydrothermal fluids offer higher diffusivity, low viscosity, and facilitate mass transport and higher dissolving power. Most important is that the chemical environment can be suitably tailored. However, the high dissolving power is a drawback for the nanoparticle (NP) synthesis, because not only to obtain higher supersaturation degree but also to suppress crystal growth, lower solubility is preferred.

To solve the problems for the synthesis of NPs, Adschiri and Arai have proposed to use supercritical state for the hydrothermal synthesis [1—4]. A fluid is supercritical when its temperature and pressure are higher than their critical point values (Tc, Pc). Most of the interesting applications of supercritical fluids occur at 1 < T/Tc < 1.1 and 1 < P/Pc < 2. Under these conditions, the fluid has some of the advantageous properties of both liquid and gas phases. Further, supercritical hydrothermal synthesis is known for making new materials under mild conditions, since they increase the chemical reactions, improve the mass transfer, and greatly assist in the stability of selected or desired products. The supercritical hydrothermal method with water as a solvent is highly suitable for the NP fabrication of a wide range of high melting-point inorganic compounds and also for hybrid organic—inorganic NPs.

One of the promising applications of NPs is for hybrid materials with polymers. The organic—inorganic hybrid

Handbook of Advanced Ceramics. http://dx.doi.org/10.1016/B978-0-12-385469-8.00051-4
Copyright © 2013 Elsevier Inc. All rights reserved.
materials have a tremendous application potential as sensors, biological tags, catalysts, bioelectronics, solar energy conversion, mechanics, membranes, protective coatings, transparent pigments, electromagnets, cosmetics, etc. The possibility of combining properties of organic and inorganic materials is an old challenge, which began from the time of industrial era. Some of the earliest and most well-known inorganic—organic representatives are those where inorganic pigments or fillers are dispersed in the organic components such as solvents, surfactants, plastics, polymers, etc., to improve optical or mechanical properties of materials. It is to be noted that a major challenge in the nanomaterial fabrication of whether organic, inorganic, or composite NPs is the preparation of highly dispersible unagglomerated NPs with an accurate control over the size and shape, which in turn is directly linked with the nanomaterial processing method. Owing to the higher surface energy of the particles in nanoscale, there is a greater tendency for NPs’ agglomeration and aggregation. At the same time, it is difficult to break down the aggregation of NPs into individual particles size. To disperse the NPs into polymers, surface energy should be reduced by increasing the affinity between NPs and polymers. Thus, surface modification of NPs is of great importance. Therefore, a new strategy for an effective solution to this problem has to be worked out. In this context, the supercritical hydrothermal method can be very effective to fabricate the nanomaterials, especially for hybrid materials, since organic modified inorganic NPs can be synthesized in situ nucleation in supercritical hydrothermal synthesis [3,4].

In this chapter, first the basic principles of supercritical hydrothermal synthesis are introduced. Designing the reactor, especially for mixing method, is explained. After summarizing some specific features of supercritical hydrothermal synthesis method, commercialization of the process is introduced. Then, organic—inorganic hybrid NP synthesis by using supercritical condition is explained. This provides the fabrication of “superhybrid nanomaterials”.

2. BASIC PRINCIPLES OF SUPERCRITICAL HYDROTHERMAL SYNTHESIS

Figure 1 shows the density of water around the critical point [5]. The critical temperature and pressure of water are respectively 374 °C and 22.1 MPa. Above the critical point, the density of water varies greatly with a little change in temperature and pressure. Because of the drastic change in density, all the fluid properties change greatly around the critical point, including the dielectric constant, which is a controlling factor of reaction rate, equilibrium, and solubility of metal oxides. Therefore, it is necessary to

FIGURE 1 Density of water around the critical point. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 2 Dielectric constant of water around the critical point. For color version of this figure, the reader is referred to the online version of this book.
understand the basic principles, which insist upon the understanding of the properties of water, including density, dielectric constant (Figure 2) [6], and ion product, varying greatly around the critical point of water and result in a specific reaction atmosphere. The dielectric constant of water at room temperature is 78, and with a raise in temperature at a constant pressure, this value decreases greatly, and around critical point of water, it is nearly equal to the dielectric constant of polar organic solvents.

Due to the variation in the properties of water, phase behavior changes greatly around the critical point. Since supercritical water is of high-density steam, light gases like oxygen or hydrogen from a homogenous phase with supercritical water, as shown in Figure 3. Figure 4 shows the critical loci for binary systems of organic compound and water [7]. In the right-hand side of these curves, a homogeneous phase is formed at any composition. The organic compounds and water are not miscible at a low-temperature range but forms a homogeneous phase at higher temperatures, which is because of the reduced dielectric constant of water. It should be noted that when the pressure is very high (i.e. the density of fluid is as high as of liquid), phase separation occurs, just like liquid water—liquid oil-phase separation.

The reaction rate, equilibrium, phase behavior, solubility of metal oxides, and distribution of soluble chemical species change greatly at the critical point range. Various models have been proposed to describe the variation of reaction rate or equilibrium over a range of supercritical state [8,9].

The knowledge on the solubility of metal oxides formed and the identification of the ionic species are essential in order to understand the supercritical hydrothermal processes. For example, the main equilibrium reactions for AlOOH dissolution in aqueous solution are given in Table 1. For the estimation of metal oxides, the equilibrium constant K for each reaction is required. There

<table>
<thead>
<tr>
<th>TABLE 1 Reactions Related to AlOOH Dissolution [3,4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlOOH + H₂O = Al³⁺ + 3OH⁻</td>
</tr>
<tr>
<td>AlOOH + H₂O = Al(OH)²⁺ + 2OH⁻</td>
</tr>
<tr>
<td>AlOOH + H₂O = Al(OH)⁺ + OH⁻</td>
</tr>
<tr>
<td>Al³⁺ + OH⁻ = Al(OH)²⁺</td>
</tr>
<tr>
<td>Al(OH)²⁺ + OH⁻ = Al(OH)₃</td>
</tr>
<tr>
<td>Al(OH)₃ + OH⁻ = Al(OH)⁴⁻</td>
</tr>
<tr>
<td>Al(NO₃)³⁻ = Al³⁺ + 3(NO₃)⁻</td>
</tr>
<tr>
<td>NaNO₃ = Na³⁺ + NO</td>
</tr>
<tr>
<td>NaOH = Na⁺ + OH</td>
</tr>
<tr>
<td>HNO₃ = H⁺ + NO₃</td>
</tr>
<tr>
<td>H₂O = H⁺ + OH⁻</td>
</tr>
</tbody>
</table>
are many models to estimate this Helgeson-Kirkham-Flowers (HKF) model is very widely used for $K$ at high temperature/pressure, including supercritical state. The following equation is a concept for the estimation of $K$:

$$
\ln K_{T, \rho} = \ln K_{T_0, \rho_0} - \frac{\Delta H^0_{T, \rho}}{R} \left( \frac{1}{T} - \frac{1}{T_r} \right) + \beta (1.0 - \rho^*)^{2/3} + a \frac{\Delta \omega_{T, \rho}}{R} \left( \frac{1}{T} - \frac{1}{T_r} \right) - \frac{\Delta \omega_{T, \rho}}{RT} \left( \frac{1}{\epsilon} - 1 \right) + \frac{\Delta \omega_{T, \rho}}{RT} \left( \frac{1}{\epsilon_{T, \rho}} - 1 \right)
$$

$T_0 = 298.15$ K
$a = 6.385 \times 10^{-5}$/K
$\rho_0 = 0.997$ g/cm$^3$

$$
\beta = \lambda_1 (\Delta C^0_{\rho, T, \rho} + \lambda_2 \Delta \omega_{T, \rho}) + \lambda_3
$$

where $K$ is the equilibrium constant, $T$ is the absolute temperature, $T_0$ is the room temperature, $R$ is the gas constant, $\epsilon$ is the dielectric constant, $\omega$ is a parameter determined by the reaction system, $\rho$ is the density, and $\rho_0$ is the density at the ambient conditions.

The solubility and the distribution of ionic species in solution are also important for the understanding of supercritical reaction atmosphere. Solubility and distribution of ionic species can be predicted by calculating $K$ for each reaction (Table 1), mass balance, and charge balance. There are several models available today to estimate the solubility of the species of interest, based on the dissociation reaction in a given system. Even by using the value estimated from the ionic radius reported in the literature, the solubility behavior can be precisely estimated.

Unlike the conventional hydrothermal process, the solubility around the critical point shows unique behavior around the critical point [10]. Figures 5 and 6 show how significantly the solubility of metal oxides and kinetics of hydrothermal synthesis change around the critical point due to the variation in the properties of water. The solubility of CuO in water over a wide range of temperature is shown in Figure 5. The solubility first increases with an increase in temperature and then decreases. This decrease in solubility is due to the lowering of the density and dielectric constant values. An estimation method of metal oxide solubility is described in detail by Sue et al. [11]. Figure 6 shows the Arrhenius plot of the first-order reaction rate constant of hydrothermal synthesis, which was evaluated under supercritical hydrothermal conditions, using a flow-type of apparatus [4,12]. The Arrhenius plot of kinetic constant shows a straight line below the critical point but, above the critical point, increased two orders of magnitude. This is because of the decrease of dielectric constant, and the details of such a mechanism has been discussed elsewhere [13].

These two results imply an important concept of NPs’ synthesis. If the aqueous solution is rapidly heated up to the supercritical state, rapid hydrothermal reaction takes place to form metal oxides; since the solubility of metal oxide under the condition is very low, extremely high supersaturation degree is obtained, which leads to nanoparticle synthesis. However, the rapid heating cannot be achieved by conventional autoclave reactors. To solve the problem, Adschiri and Arai proposed to use a flow-type system in which two flows are mixed to achieve rapid heating. Thus, Adschiri and Arai have pioneered the supercritical hydrothermal synthesis of metal oxides in detail during 1990s, based on this concept [1-4,12-15].
3. APPARATUS

3.1. Basic Concept of Flow-type Reaction System for Rapid Heating

Figure 7 shows the typical experimental setup used in this study. An aqueous metal salt solution is prepared and fed into the apparatus in one stream. In another stream, distilled water is pressurized and then heated to a temperature above the desired reaction temperature. The pressurized metal—salt solution stream and the pure supercritical water stream are then combined at a mixing point, at which rapid heating of the metal—salt solution and subsequent reactions in the reactor occur. After the solution leaves the reactor, it is rapidly quenched. In-line filters are used to remove larger particles. The pressure is controlled with a back-pressure regulator. Fine particles are collected in the effluent. The supercritical hydrothermal flow reactor provides NPs with desired shape, size, and composition in the shortest possible residence time. Figure 7b shows the semipilot plant flow reactor (ITEC Co., LTD, Japan).

3.2. Microreactor

For the rapid mixing/heating followed by the reactions, nucleation and particle growth, designing the geometry of the mixing point is of great importance [16]; some researchers have developed microreactor reaction systems for the supercritical hydrothermal synthesis.

Microreactors are miniaturized reactors that contain microchannels of characteristic dimensions in the order of submillimeters to a few millimeters [17]. Some micro-mixers, which are mixers containing submillimeter mixing chambers, can also be used as a reactor especially for reactive mixing [18,19]. The reactor miniaturization provides improved mass- and heat-transfer rates owing to a high surface-area-to-volume ratio and thus enables us to proceed with reactions under conditions controlled more precisely as compared with conventional macroscale reactors. This advantage results in improved yields and selectivities of desired products [20]. Enhanced mixing performance in microreactors is also effective to produce desired products in high yield and selectivity. The selectivities of desired products for very rapid multiple reactions have been improved using microreactors [21]. Thus, microreactors enable syntheses with higher efficiency and less by-products than those of conventional macroscale reactors. The efficiency of microreactor can lead to the development of more compact and environmentally benign processes. In this section, hydrothermal syntheses using microreactors are reviewed.

The combination of straight tubes and T-shaped mixers such as union tee is an example of simple microreactor system. Dr. Ikushima’s group performed organic syntheses using supercritical water (scH$_2$O), based on such microreactor system [22–24]. The schematic diagram of the system and temperature profile in the system is illustrated in Figure 8. The tubes and the union tee are made of Hastelloy-C-276. The inner diameter (i.d.) of the tubes and the tee is 300 μm, and the outer diameter (o.d.) of the tubes is 1/16 inch (1.59 mm). In organic syntheses, a desired product is often unstable under the reaction condition. The products tend to change into by-products in quite a short time. For increasing the selectivity of the desired product, the reaction should be quenched immediately after the formation of it. To realize such reaction operation, a reactor that enables rapid mixing and heat transfer and short residence time is required, and the features of microreactor are suitable for this purpose. As shown in Figure 8(b), the scH$_2$O microreactor system can heat up instantaneously an ambient substrate solution to the supercritical state in a molten salt bath and can then quench rapidly to sufficiently low
temperatures after the reaction in a water/ice bath. Meanwhile, in the conventional macroscale system (Figure 8(c)), the temperature slowly changes up and down, and this causes the decomposition of substrates and side reactions. By using this microreactor system, the authors successfully controlled the reaction to achieve extremely high selectivity and yield for organic synthesis reactions.

T-shape micromixers are also used for continuous hydrothermal synthesis of NPs. In such synthesis, in which an aqueous metal-salt solution at a normal temperature and supercritical water are directly mixed, the mixing process of the two fluids is crucial for obtaining a highly supersaturated crystallization field and controlling the growth process of fine metal oxides. Sue et al. synthesized Fe$_2$O$_3$, NiO, and CuO NPs from starting solutions of metal nitrates at 673 K and 30 MPa, using a microreactor [25]. A T-shaped micromixer (0.3-mm i.d.) connected by tubular microreactor (0.5-mm i.d., 50-mm length and 1.7-mm i.d., 4-mm length) was used for rapid heating of the starting solutions. The residence time was varied from very short time of 0.002 s with changing the length of the 1.7-mm i.d. channel. In the Fe$_2$O$_3$ synthesis, the conversion became 97.9% at shortest residence time of 0.002 s, and the average particle size slightly increased from 4.0 to 6.7 nm with increasing residence time. For NiO and CuO syntheses, the same tendency of the particle size was obtained. They also confirmed the rapid heating in the reactor through computational fluid dynamics simulation, using FLUENT (ANSYS, Inc.). This group also synthesized ferrite NPs from Fe(NO$_3$)$_3$ and Me(NO$_3$)$_2$ (Me = Ni, Cu, Zn) aqueous solutions, using the microreactor system [26].

Moreover, this tubular microreactor system can be used for higher temperature reactions (450 to 650 °C and 250 bar) [27,28]. A Hastelloy-C-276 tube with a volume of 0.912 cm$^3$ (2 m long, 1.6 mm (1/16 inch) of o.d., 762 μm (0.03 inch) of i.d.) served as the tubular microreactor. From the experimental results, the kinetics of this reaction, which is essential for the reactor design, was determined. Thermal decomposition of HCO$_2$H was also carried out to generate H$_2$ in high-temperature pressurized water (15–33 MPa, 250–400 °C), using 1/16-inch tubular reactor made of stainless steel AISI 316 [28]. The kinetics determined accurately is a critical information for understanding metallic NP synthesis in supercritical conditions.

The rate of temperature rise in tubular microreactors have been investigated experimentally [29] and numerically [30,31]. Wakashima et al. [29] achieved heating from room temperature to 673 K within 0.05 s under the pressure of 23–45 MPa and the flow rate of 0.6–4.6 kg/h. A microchannel made of Ni-base Alloy 625 (i.d.: 0.258 mm, o.d.: 1/16”) was used. This heating corresponds to the heating rate of $10^4$–$10^5$ K/s and an overall heat transfer coefficient of 34,500 W/(m$^2$·s).

Microreactors designated for hydrothermal synthesis have also been developed. Mae et al. developed a microreactor designed for the use under 50 MPa and 600 °C and
consists of a 5-way junction with a moveable needle [32]. Figure 9(a) shows the overview and Figure 9(b) the mixing part of the developed micromixer. From one of the five channels in the junction, fluid A flows into the mixing part, and from four of five channels, fluid B flows. To change the degree of mixing, the edge of needle is moved down within the range of 0 to 1 mm (denoted by \(D\) in Figure 9(b)) from the top position of the 5-way junction by rotating the adjustment screw. The device consists of upper and lower parts, which are assembled with a \(\Delta\)-ring sealing to suppress the heat transfer from lower parts to upper parts.

The flow of reactant fluids and heat transfer fluids in the mixing part is illustrated in Figure 9(b). The mixing principle is based on the collision of split fluid segments and confluence proposed in the development of the K-M (Kyoto University-MCPT; Research Association of Micro Chemical Process Technology) mixer [33]. The mixing performance is expected to be enhanced by turbulence of mixed flow. Fluid A flows in the annular clearance around the needle. The width of the clearance is 200 \(\mu\)m. At the edge of the needle, fluid A is mixed with fluid B (preheated supercritical water). The needle edge enables confluence of the two reactant fluids, and the inside of the needle acts as a channel for heat transfer fluid (see next paragraph for detailed explanation). The outlet channel diameter of the mixing part is 800 \(\mu\)m. For using this reactor in industrial production with high throughput, the channel sizes were designed to be of the order of several hundred micrometers. Moreover, this micromixer includes a structure for the heat insulation to suppress the heat transfer from lower parts to upper parts. In the upper parts of mixer, fluid A is kept at room temperature by heat exchange through coolant and cooling fin. These designs enable us to mix the reactants rapidly while keeping each temperature of reactant.

The mixing performance of the micromixer was examined, using the boehmite NP synthesis, with supercritical water under 400 °C and 30 MPa. Boehmite (AlOOH) NPs having the average diameter of 20 nm were successfully obtained under high total flow rates (90 g/min) through the strict control of mixing, heating, and reaction time within 10 ms. In addition, the proper selection of needle position and flow rate allows the control of fast crystal growth.

A swirling micromixer as a replacement of a conventional union-tee mixer for improved mixing of reactant fluids was also developed [34]. The developed micromixer was featured with segmented introducing of supercritical water, swirling and convergent nozzle flow. The inner diameter of the three fluid-inlet channels is 1.0 mm. The mixing chamber has a conical shape with an apex angle of 8°. Supercritical water is divided into two equivalent streams and directly mixed with metal—salt solution at room temperature at across angle of 60°. This mixer was fabricated from a Ni-base alloy block, using high-precision machining. The average size of boehmite NPs from an aqueous solution of aluminum nitrate (Al(NO₃)₃·9H₂O) was 59.7 nm for the swirling micromixer and 126.2 nm for the conventional 1/16-inch union tee (0.8 mm i.d.).

Because of the requirement of high-temperature and high-pressure resistances, the choice of reactor material is also important for the reactor design. Stainless steel and Ni-based alloys are mainly used for the reactor material. For optical access and flow visualization, glass is also selected. Instances on the evaluation of materials of microreactors operated under near-supercritical conditions are then introduced. Javad et al. fabricated microtubular reactors composed of Inconel 625 support, TiO₂/Ti intermediate layer, and thin catalytic metal film (Pd, Pd–Cu alloy, Pt, Rh, and Au), using continuous electroless plating technique [35]. The reactors withstood high temperature (400 °C) and high-pressure (30 MPa) water flow without the loss of catalytic metals.

Trachsel et al. fabricated silicon/glass microreactors to safely handle high pressure and provide optical access into
the reaction channel for flow investigations [36]. Around the inlet and outlet holes, chromium, nickel, copper, and gold were subsequently deposited by electron-beam evaporation (Pfeifer Vacuum Classic 500). Gold is used as the top layer to provide complete wetting with the tin–silver solder. Copper prevents the formation of an oxide layer on the nickel layer. Nickel itself acts as a diffusion barrier. Chromium is the actual undercoating on the silicon surface. The mechanical analysis of the microreactor and its microfluidic connections showed a mechanical strength exceeding operating conditions of the temperature of 80 °C and the pressure of 140 bar.

Tiggelaar et al. examined the design, fabrication, and high-pressure performance of several in-plane fiber-based interface geometries to microreactor chips for high-pressure reactor operation [37]. Pressure experiments revealed that the optimal inlet geometry for microreactor chips for high-pressure chemistry is a tubular structure that is etched with HF and suitable for fibers with an outer diameter of 110 μm. These inlet/outlet geometries can withstand pressures up to 690 bar.

4. NP SYNTHESIS BY SUPERCritical HYDROTHERMAL SYNTHESIS

4.1. Metal Oxide Nanoparticle

By employing the above discussed principles of supercritical hydrothermal synthesis with the well-designed flow-type reaction system, a wide range of metal oxide NPs have been obtained; the applications of those NPs cover not only in ceramics, coatings, catalysts, sensors, semiconductors, magnetic data storage, solar energy devices, and ferrofluids but also in medical fields such as hyperthermia, bioimaging, cell labeling, and special drug delivery systems. Researchers like Adschiri, Arai, Johnston, Lester and Poliakoff, etc. have contributed extensively on the synthesis of metal oxides and proposed a systematic mechanism [1,2,38]. Metal oxides like Al₂O₃, Ga₂O₃, In₂O₃, SiO₂, GeO₂, ZnO₂, V₂O₅, TiO₂, CeO₂, ZrO₂, CoO, α-Fe₂O₃, γ-Fe₂O₃, NiO, Co₃O₄, Mn₃O₄, γ-MnO₂, Cu₂O, CoFe₂O₄, ZnFe₂O₄, ZnAl₂O₄, Fe₂CoO₄, BaZrO₃, BaTiO₃, BaFe₁₂O₁₉, LiMn₂O₄, LiCoO₂, La₂O₃, etc., have been prepared by the above method. A complete list of the materials obtained under supercritical hydrothermal conditions is available in the works of Byrappa and Adschiri (2007) and Reverchon and Adami (2006) [39,40].

Basically hydrothermal synthesis method is available for the metal oxides by conventional hydrothermal synthesis method. The point to be noted here is that, under supercritical hydrothermal conditions, nanometer size metal oxides could be synthesized, and crystallinity of the NPs would be much higher when compared with the metal oxides obtained under conventional hydrothermal conditions, wherein bulk single crystals are formed. This sometimes leads to the specific characteristics of the products. Phosphor NPs show high luminescence without heat treatment, which is usually necessary for the products obtained by low-temperature wet method [15,41]. NPs of LiMn₂O₄ with a particle size ranging from 10 to 20 nm, synthesized from LiOH, Mn(NO₃)₂, and H₂O₂ at 400 °C and 30 MPa, do not show decay of its capacity even after the charge–discharge cycles, which has been considered as a major breakthrough point of these solid electrolyte materials [12,42]. Kanamura et al. (2001) have discussed this mechanism in detail and concluded that these particles are single crystals of LiMn₂O₄ and are totally different from those obtained by other methods [43]. In the synthesis of LiMn₂O₄ NPs, oxidizing reaction atmosphere has to be controlled by regulating oxygen gas partial pressure in the system. In this case, Mn²⁺ of Mn(NO₃)₂ should be oxidized into Mn³⁺. In order to achieve this, H₂O₂ was fed into the system. H₂O₂ decomposes at supercritical conditions into oxygen gas, which forms a homogeneous phase with supercritical water to provide an excellent oxidizing atmosphere [12,42,43].

The reduction atmosphere can be also controlled by using supercritical phase. Ni NPs could be synthesized from nickel acetate and formic acid at 400 °C and 30 MPa. In this case, HCOOH was introduced as a reducing agent with nickel acetate solution [44,45]. In supercritical water, HCOOH is decomposed into hydrogen and carbondioxide. An important point is that these gases and supercritical water form a homogeneous phase, and this mixture of gas (H₂ and CO₂) shows higher reducing ability than H₂ gas, as was reported in the literature [46,47].

4.2. Commercialization of the Process

A Korean company, Hanwha Chemical, has developed a continuous process in a pilot plant for supercritical hydrothermal synthesis of CeO₂, CeₓZr₁₋ₓO₂, LiCoO₂, LiNi₀.₅Mn₀.₅O₂, LiNi₀.₅Mn₀.₅O₂, LiFePO₄, CaMn₂O₄, and so on. The pilot plant, with a capacity of 4 kg of dry powder/h, was operated to collect scale-up information and basic design data for a full-scale commercial plant.

For CeO₂, ceria particles in a wide size range (15–45 nm) are produced for shallow trench isolation chemical mechanical polishing. High pattern selectivity and oxide removal rate was obtained for the products. The CeO₂ products obtained at 25 MPa and 673 K within 30 s showed high-performance oxygen storage capacity (OSC). An engine test showed similar performance in the conversion of CO, hydrocarbon, and NOₓ, and a chassis test showed more than 20% lower emissions of hydrocarbon and NOₓ [48,49]. They have developed a new method for the treatment of wastes from the process, as well. In the reactor, nitric acid generated during the hydrolysis of metal...
nitrate reacted with ammonia to yield nitrogen gas. This reaction occurs only above the critical point, making the process cleaner.

LiFePO$_4$ is a promising cathode material of Li-ion battery. Hanwha Chemical constructed the first commercial plant for the supercritical hydrothermal synthesis of LiFePO$_4$ at the end of 2010. This commercial plant continuously produces LiFePO$_4$ and has a capacity of 1000 tons/year [48,50].

The charge/discharge curves obtained for LiFePO$_4$ produced by supercritical hydrothermal synthesis are between 2.0 and 4.4 V (versus Li/Li$^+$) at 298 K and has a capacity of 162.7 mAh/g, which is close to the theoretical capacity [48].

5. SUPERCRITICAL HYDROTHERMAL SYNTHESIS OF ORGANIC–INORGANIC HYBRID NANOPARTICLES

5.1. Organic–Inorganic Hybrid Nanoparticles

There is a significant need for the organic–inorganic hybrid materials that show the tradeoff functions between polymers and inorganics (light and high mechanical strength, high thermal conductivity and electrical resistance, transparent and flexible electroconducting films, etc.). So far, tremendous efforts have been made to fabricate NP-dispersed polymer, but it has been considered a difficult task to disperse the NPs in organic solvents or in polymers, especially for the particles synthesized under hydrothermal conditions. This is because metal oxide particle surface is hydrophilic, and for the case of NPs, it shows extremely high surface energy, which leads to the formation of aggregates.

In order to resolve the problem of particle aggregation, to increase the affinity between the organic molecules and NPs’ surface, namely, organic surface modification of NPs is required. There are several approaches for the surface modification of NPs, including silane coupling method and solvothermal or hot soap method with capping agents. However, for the surface treatment, large quantity of organic solvents is required, which is not environmentally benign. As described above, hydrothermal process is regarded as the most environmentally friendly route to synthesize NPs, but for the surface modification, organic ligands should be mixed with the aqueous solution, which is contradictory to the purpose; a new processing strategy has been proposed by Adschiri and group, utilizing the supercritical hydrothermal technology [51–56].

Figure 10 shows a schematic representation of highly effective strategy for the synthesis of metal oxide nanocrystals in the organic-ligand-assisted supercritical hydrothermal technology [52]. The method yields perfect hybrid organic–inorganic nanocrystals with very high dispersibility and a precise control over the size and shape of the NPs. The organic components are introduced into the system during the hydrothermal synthesis, and in situ surface modification is obtained with an ultrathin layer of organics surrounding the inorganic, unlike the case of silane coupling on the metal oxides. The organic ligands and supercritical water form a homogeneous phase, and it is known that, under these conditions, water molecules themselves work as acid or base catalyst for various organic reactions. Depending upon the applications of NPs, one can select a suitable functional group to introduce hydrophobicity or hydrophilicity property to the surface of the modified NPs.

**FIGURE 10** Concept of the hybrid organic–inorganic NPs’ synthesis in supercritical water [52]. For color version of this figure, the reader is referred to the online version of this book.
There are several reports on the synthesis of hybrid organic—inorganic NPs under supercritical conditions like TiO₂, ZnO, Co₃Al₂O₄, etc. [53,58,71]. It is to be noted that the surface-modified organic—inorganic hybrid NPs, synthesized under supercritical hydrothermal conditions, can be easily recovered by extraction with organic solvent from the water suspension phase and well-dispersed in organic solvents.

Interestingly, this method can be applied not only for small capping agents but also for polymers. Taguchi et al. [57] succeeded in making CeO₂ NPs with hydrophilic polymers attached to the surface by the supercritical hydrothermal synthesis method, as shown in Figure 11. Hybrid (NPs) like these are especially useful, as they allow the physical properties of the metal oxide to be combined with the desirable tunability of the organic molecules. Ceria NPs are used in many applications from three-way catalysts to oxygen ion conductors, and also in cosmetics, for their UV shielding.

This approach can control the size and morphology of the particles, as shown in Figure 12. When decanoic acid (molar ratio to ceria precursor 6:1) was added to the reaction system, the resulting nanocrystals were nanocubes with an average size of 6 nm. The ceria nanocrystals unmodified were (truncated) octahedral, which corresponds with the results of the earlier workers [58]. However, when decanoic acid was introduced into the system under supercritical hydrothermal conditions, the organic ligand molecules become miscible with water, and they contribute to the change in morphology of ceria nanocrystals from (truncated) octahedral to cube, which is mainly because of the suppression of the crystal growth on the (001) face. As the CeO₂ (001) surface is less stable than the (111) surface, the organic ligand molecules were likely to interact preferentially with the (001) surface, thereby reducing the crystal growth along the [001] direction. If the amount of decanoic acid was further increased (molar ratio to ceria precursor, 24:1), the ceria nanocrystal size decreased to about 5 nm, and the shape of the nanocrystals also changed drastically because of the growth inhibition not only along the [001] direction but also along the [110] direction. Figure 13 shows the schematic diagram of the shape control of ceria NPs under supercritical hydrothermal conditions in the presence of organic ligand molecules [52].

Exposing (100) facet is of great importance to obtain high catalytic activity for CeO₂, a three-way catalyst for cleaning automobile emission. It was known that the catalytic activity can be tuned by controlling the exposed surface. However, it has been considered a difficult task to control the exposed facet for nanocrystals. Always, the most reactive surface grows faster, leaving the unreactive facet exposed to outside. However, the in situ capping method could solve the problem and succeeded in exposing the most reactive facet with controlling the size to be around several nm. OSC is a measure of catalytic reactivity, which is directly related to the oxygen mobility and its total amounts. Usually, mobility of oxygen atom in metal oxides is observed at above 300 °C. In other words, catalyst works at higher temperatures. However, CeO₂ nanocube shows high mobility even at lower temperature, as shown in Figure 14. This result suggests not only the high catalytic reactivity but also the high potential of this nanocube for environmental catalyst that works at very low temperature, possibly even at room temperature [59].

We believe this method opened a door to fabricate a new nanocatalyst with high catalytic activity. This approach can be applied for the other catalysts, as well. In this experiment of CeO₂ nanocube synthesis, we used carboxylic acid as a capping agent. However, by selecting other capping agents, or by controlling reactivity/stability/affinity of each face of crystals by changing pH/density/temperature, we may be able to control the exposed surface. We are now challenging on making new methodology of catalyst fabrication.

The supercritical method shows another advantage for metallic NPs’ synthesis [44,45,60–62]. It has been already
explained that Ni NPs could be synthesized in supercritical hydrothermal condition by utilizing the homogeneous phase for H\textsubscript{2}-H\textsubscript{2}O system. Obviously, a green route for metal NPs is the advantage. However, since water can catalyze oxidation of metals and water itself can oxidize metals, it may be unrealistic to synthesize metal NPs in aqueous media. As predicted by the free energy change of reduction of metal oxides by hydrogen, platinum, gold, silver, palladium, and copper can be synthesized. However, synthesis of aluminum, zinc, and iron should have extremely high partial pressure of H\textsubscript{2}, and even though these metals are formed under the supercritical conditions, these metals should be oxidized easily with water after cooling. In order to suppress the oxidation, in situ surface modification may be an effective method. Figure 15 shows the result for iron NPs' synthesis with 0.1-M hexanoic acid (surface modifier). As a result, the surface modification increased the yield of iron, up to 7.6%. The dispersion of Fe NPs synthesized did not exhibit color even after 20 days of storage in Ar atmosphere, whereas Fe NPs synthesized without hexanoic acid did exhibit yellowish color just after only a few days of storage, as shown in Figure 16. X-ray diffraction (XRD) patterns of Fe NPs, synthesized with 0.1-M hexanoic acid, was also shown in this Figure 15(d). The XRD diffraction pattern shown in Figure 16(a) was taken just after synthesis and Figure 16(b) was taken 20 days after the synthesis. There was no change in XRD pattern, and this implies the Fe NPs were not oxidized. Thus, the in situ surface capping effectively suppressed the surface oxidation.

We also performed sc-hydrothermal synthesis of FePt NPs. Iron alloys are more easily synthesized and more oxidation resistant than pure metal iron. As mentioned above, sc-hydrothermal synthesis, using formic acid as reducing agent and oleic acid as surface modifier, succeeded to synthesize small amount but actually iron NPs. By applying the same method, we succeeded to synthesize oleic acid-capped FePt NPs. The synthesized FePt NPs
showed quite stable dispersion in several organic solvents e.g. tetrahydrofuran (THF) and chloroform (see Figure 16). The FePt NPs was synthesized with 3-M formic acid and 0.2-M oleic acid. As a reference, the diffraction patterns of JCPDS file for FePt (No. 21-1272) and Pt3Fe (No. 29-1423) were also displayed in the electron diffraction (ED) pattern. The images of the THF dispersions of the unmodified and oleic acid-modified FePt NPs were also shown in this figure. Left-hand side image (d) was taken just after the preparation and right-hand side image (e) was taken 3 days after the preparation. Thus, this figure also demonstrates the stable surface capping by this method.

5.2. Mechanism

As explained in Figure 11, the key point of the in situ surface modification during the supercritical hydrothermal synthesis is the homogeneous phase formation for organic capping agents and aqueous solutions. The direct observation of the reaction atmosphere is critical to elucidate the reaction mechanism. Thus, a batch reaction system with optical view windows was employed. The optical cell was made of Hasterloy-C equipped with two sapphire windows of 15 mm in diameter and xxx mm of thickness. The inner volume of the cell was 1.4 mL. A half of the cell volume was filled with cerium hydroxide water slurry, and then 500 µL of octanoic acid was added. The sample solution was heated up to 390 °C with 30 °C/min by using four cartridge heaters in the cell; the phenomena inside the cell was recorded with charge-coupled device (CCD) recorder.

Figure 17 shows the result of the direct observation. As shown in figure (a), the under phase in the cell was water phase in which cerium hydroxide existed, and the upside phase was organic phase of octanoic acid. As figure (b) shows, color of the upside phase changed to yellow around 200 °C and then became a uniform phase. These results suggest that the cerium hydroxide was transferred to the organic phase, probably with forming cerium octanoate. Above 330 °C, when water and organic phases become uniform, cerium octanoate was hydrolyzed to form NPs, but at the same time, surface modification occurs with the octanoic acids [63].

FIGURE 15 (a) TEM image of Fe NPs synthesized with 0.1-M hexanoic acid. (b), (c) Image of the dispersion of Fe NPs synthesized with and without hexanoic acid, and (d) XRD pattern of the surface-modified Fe NPs, [62] respectively. For color version of this figure, the reader is referred to the online version of this book.
Adschiri and group have worked out in detail for the theory and mechanism of the hybrid organic— inorganic NP formation under supercritical hydrothermal conditions. In aqueous systems, metal oxide particles are hydrated and $\text{M} - \text{OH}$ groups cover their surface. The $\text{M} - \text{OH}$ sites on the surface of particles can react with $\text{H}^+$ or $\text{OH}^-$ ions from dissolved acids or bases, and positive ($\text{M} - \text{OH}^2+$) or negative ($\text{M} - \text{O}^-$) charges develop on the surface, and type of the reaction depends on pH of the solution. Thus, metal oxides have a characteristic pH, where the net surface charge is zero isoelectric point (iep). At pH lower than iep, the pure metal oxide surface is positively charged, while it has negative charge above it [64–66]. Normally, the modifiers attach onto the surface of the NPs either by physisorption or chemisorption through strong hydrogen bond. If we consider $\text{R-COOH}$ as modifier reagents in a highly acidic pH of the reaction medium, the dissociation of modifier does not occur, and the conjugation is due to the strong hydrogen bonding between hydroxylic groups on the surface of NPs and functional groups of modifier [67]. If the pH of the reaction medium is in the range of modifier dissociation, another type of interaction between the modifier and the metal oxide surfaces is expected [68].

Figure 18 shows the case of CeO$_2$ NP capped with oleic acid: oleic acid has a pKa value 5.02 [69], and iep of cerium oxide is 6.8 [70]. By calculating [RCOO$^-$] from low pH up to the base conditions, it was found that amount of dissociated part of modifier increases with pH, and the maximum point for surface modification is in the pH range around 5.5, somewhere between pKa and iep, in which the surface charge is negative and chemical reaction occurs between NPs and dissociated part of the modifier as RCOO$^-$.
However, it is necessary to consider such calculations at elevated temperature and pressure or the experimental pressure and temperature conditions. Figure 18 shows the concentration of [RCOO$^{-}$] and amount of surface modifier on the particle’s surface versus pH of the reaction media. The optimum surface modification is seen in between iep and pKa. Using the above discussed novel approach, the experimental studies on the preparation of some selected hybrid organic—inorganic NPs under supercritical hydrothermal conditions are discussed here for the benefit of the readers.

Similar interaction is expected even for amines or alcohol or aldehyde as the organic ligands in the system. The addition of surface modifiers also helps to inhibit the crystal growth that facilitates the smaller particles size with a narrow particle-size distribution.

### 5.3. Bonding and Surface Nature of the Hybrid NPs

Several methods are commonly employed to understand the surface bonding or interaction between the inorganic and organic molecules. The most commonly used techniques are Fourier-transform infrared (FTIR) and thermal analysis in addition to the specific characteristic properties like luminescence, magnetic, etc. Depending upon the type of inorganic particle and its properties, an appropriate
technique is to be selected. Figure 19 shows the FTIR spectrum of ceria NPs modified with decanoic acid under supercritical conditions. Bands in the 2800–2960 cm⁻¹ region were attributed to the C–H stretching mode of methyl and methylene groups. The bands at 1532 and 1445 cm⁻¹ corresponded to the stretching frequency of the carboxylate group, which suggests that the carboxylate group from decanoic acid was chemically bonded to the surface of the ceria nanocrystals, and the other hydrocarbon groups were oriented outward [72]. This is probably the result of reactions forming chemical bonds between the nanocrystal surface and the organic-ligand molecule in the unique reaction conditions of supercritical water, which are essential for the perfect dispersion of nanocrystals in the organic solvents.

Similarly the thermogravimetric studies assist in estimating, more or less precisely, the quantity of organic molecules surrounding the inorganic metal oxide surfaces. For the modified iron oxide NPs, by increasing the temperature to 100 °C and keeping it constant for several minutes, all the water molecules and hydroxyl groups on the surface will be evaporated and removed. By increasing the temperature until boiling temperature of the modifier, the weight loss corresponds to the decomposition of modifier from surface of iron oxide NPs. Keeping temperature constant at 270 °C indicates that weight loss is continuing at this temperature. Such studies are available for all other hybrid organic–inorganic NPs in the literature [53–56,58,71].

As described above, the in situ surface modified NPs have strongly bound and dense modifier layer on the surface. However, structure of surface modifier, such as self-assembly monolayer (SAM) or crystalline phase, is not clarified yet. SAMs can affect surface property and electrostatic potential of NPs so that its properties as well as stabilities in dry and solution states provides an essential information on the structure of surfaced modifiers.

For this analysis, purification of surface modifier is essential. We employed antisolvent crystallization method that is usually used for the purification of medical products, followed by dispersion–centrifugation repetition in ethanol and toluene.

The grafting density of 3D SAM was determined from Thermogravimetry (TG) weight loss and averaged size of the CeO₂ NPs measured by TG and TEM, respectively (see Figure 20). For instance, the grafting density of a purified decanoic acid ((a) curve in Figure 20(c)) was 4.85 chains/nm², which suggests highly packed structure like SAM or lie a quasicrystalline state. Figure 21 shows Differential Scanning Calorimetry (DSC) curve for the decanoic acid-modified CeO₂ NPs. No thermal phase transition was observed, which suggests the quasicrystalline state.

After purifying the modifier layer on NPs, the structure and physical property of the modifier layer was studied. As
shown in Figure 21, in trans-decalin, there is no small exothermic peak in the DSC cooling curve, whereas small exothermic peak is seen in cyclohexane. This implies cyclohexane swelled the SAM, and the cyclohexane trapped in the SAM could not come out when the cyclohexane outside the SAM was frozen. In the case of trans-decalin, because of larger molecular size of trans-decalin, the SAM was not swelled. Therefore, no small peak is seen in the DSC cooling curve.

The swelled SAM had large osmotic pressure because of the extraordinarily high cyclohexane concentration. Therefore, cyclohexane inside of the SAM could not come out from the SAM, even though the cyclohexane outside the SAM was frozen. The strong osmotic pressure inside the swollen SAM may have key importance to disperse alkanoic acid-modified CeO$_2$ NPs in organic solvents. Actually, alkanoic acid-modified CeO$_2$ NPs that have quasicrystalline state SAM on the surface showed good dispersion in nonpolar organic solvents (Figure 22).

5.4. Dispersion of NPs in Solvents

Normally transparent solutions are obtained for the well-crystallized and very small hybrid NPs, when dispersed. Through the dynamic light scattering (DLS) studies also, it can be confirmed that the particles are not aggregated but suspended as individual NPs. Figure 23 shows the perfect dispersion of various NPs such as Ni, CoAl$_2$O$_4$, Fe$_3$O$_4$, TiO$_2$, and CeO$_2$ in the organic solvents [77 Green Chem].

However, we also found a phenomenon for understanding of the dispersion of NPs in solvents: when we...
mixed equal portions of perfectly (transparently) dispersed CeO$_2$ NPs solution in cyclohexane with CeO$_2$ NPs solution in cyclohexane with white turbidity, the turbidity was cleared and the mixture showed stable dispersion (see Figure 24). The concentration of the not well-dispersing CeO$_2$ NPs was higher than the solubility limit for perfect dispersion of the not well-dispersing CeO$_2$ NPs.

This kind of phenomena is often observed for a molecular phase behavior: by the addition of third component, phase behavior is shifted, because of the change of quasi-solvent nature, or by the balance among three components. Especially for increasing solubility of the poorly solubilized solute, the third component that shows high affinity both for the solvent and the solute increase solubility of the solute. It seems similar phenomena occurs also for the NPs systems.

In the last section, the modifier’s bonding and surface natures were described. Due to the affinity difference, the state of surface modifier is changed, which reflect the interaction between the solvent and NPs. Obviously, when the affinity is high, molecules are swollen, which makes the interaction better to show higher solubility.

The dispersion of alkanoic acid-modified NPs drastically changed. By examining the relation between the dispersion of NPs and the behavior of SAMs, we found that quasicrystalline (polymer brush-like) SAMs on NPs required for the stable dispersion of NPs at a high concentration.

A typical relation between the solution-state behavior of SAMs on NPs and dispersion of NPs in organic solvents was depicted in Figure 25.

6. SELF-ASSEMBLY OF HYBRID ORGANIC—INORGANIC NANOPARTICLES

6.1. 2D Self-assembly of NPs on Substrate

Self-assembly and self-organization of particles in a solvent is considered to be a powerful process for building patterns up to nanoscopic level through multiple interactions among the components of the system under consideration. Ordered assemblies of nanometer size particles represent interesting nanomaterials that provide exceptional potentials for a wide variety of applications. These structures would be useful for various applications like displays, sensors, data storage, photonic band gap materials, etc. owing to their exceptional physical, chemical, and electronic properties [77–80].

There are several reviews published on this aspect [81,82]. However, it is to be noted that the self-assembly of hybrid organic—inorganic NPs synthesized through supercritical hydrothermal routes are seldom found in the literature. The mechanism of making self-assembly structures involves two steps: first step is the positioning of

![Figure 23](image-url)  Transparent dispersion of NPs in organic solvents. For color version of this figure, the reader is referred to the online version of this book.

![Figure 24](image-url)  When we mixed equal portions of 5 wt% C$_{10}$ acid-modified CeO$_2$ NPs in cyclohexane (a) and 2 wt% C$_{12}$ acid-modified CeO$_2$ NPs in cyclohexane (b), the mixture (c) showed much better dispersion than 1 wt% C$_{12}$ acid-modified CeO$_2$ NPs in cyclohexane (d) [74]. For color version of this figure, the reader is referred to the online version of this book.
perfectly dispersed particles (in solvent) on the substrate by controlling interactions between substrate and particles surfaces (hydrophilicity and hydrophobicity). In the second step, capillary forces between the particles and the surface laterally displace the particles during drying. Steric repulsion between particles because of capping agent on the surface prevents aggregation of NPs. After complete evaporation in the third stage, an irreversible reorganization of the particle—substrate interface occurs, which prevents displacement of NPs [83].

As already shown in Figure 26, 2D self-assembly of NPs were obtained for CeO₃, Co₃O₄, and Fe₃O₄. For Co₃O₄ and Fe₃O₄, the self-assemblies of Co₃O₄ and Fe₃O₄ NPs were obtained under supercritical hydrothermal conditions in the presence of organic ligand molecules C9COOH and C17COOH, respectively. The decanoic acid has a carbon chain length of 1.4 nm, and it can be seen that the distance between self-assembled NPs is about 2.8 nm. Similarly, oleic acid has a carbon chain length of 1.9 nm, but the distance between the oleic acid-modified magnetite particles is about 3.8 nm. Thus, by choosing different types of modifiers with different carbon chain length, the superlattice of the self-assembled NPs can be monitored precisely. Similar superlattice structures have been obtained for modified titania, zinc oxide, boehmite, cobalt spinels, etc. [53–56]. This area of self-assembly of the hybrid organic—organic NPs is fast growing and it has several advantages in obtaining desired superlattice structures with different packing arrangement for specific applications.
As explained above, the self-assembly is affected by the interactions between particle—particle, particle—solvent, and particle—substrate. Thus, by changing the balance of the interactions, self-assembled structure is varied. The plate-like particles are likely to be stacked on each other and stand perpendicular to the TEM grid. However, for the modified particles that are hydrophobic, the particles show affinity for the TEM grid. Thus, well distributed NPs lying on the TEM grid can be seen, which suggest successful surface modification of boehmite particles [56].

6.2. Fixing NPs on the Substrate

Highly crystalline metal oxide films have a high potential for use in active catalysts, high refractive index films, and even high-permittivity insulating materials. However, such high-quality films are fabricated by subjecting them to high-temperature treatments, which damage heat-sensitive substrates such as plastic polymers. If the process of fabricating a thin film can be divided into the processes—synthesis of highly crystalline NPs at high temperature and assembly of the nanocrystals on substrates at room temperature—highly crystalline metal oxide thin films can be fabricated on such heat-sensitive substrates.

For achieving this process, in the first stage, particles should be dispersed for coating/printing processes, which require high affinity between NPs and the solvent. After the coating/printing, during the drying process, NPs should be self-assembled on the substrate through capillary force or vdw interaction, etc., for which interaction between the NPs should be stronger than that between NPs and substrate. In the final stage, for the interaction between the NPs and substrate, the stronger the better. Thus, the control of interaction among NPs-solvent—substrate is the critical issue.

To solve this problem, a new, simple, and efficient method for assembling cerium oxide nanocrystals, modified with decanoic acid two-dimensionally on the surface by modifying substrates with 3,4-dihydroxycinnamic acid (DHCA), was proposed [84]. In order to chemically anchor high-dispersible decanoic acid—cerium oxide hybrid nanocrystals, on-site ligand exchange between carboxylic group of decanoic acid, a modifier of hybrid nanocrystals, and catechol group of DHCA terminated on the Si substrate was used, as shown in Figure 27.

For the fabrication of two-dimensional structures or monolayer-controlled thin films of NPs on the surface, the interactions between NPs should be decreased, whereas the interactions between NPs and the substrate should be carefully designed [85]. The interactions between NPs and the substrate can be controlled by the formation of some covalent bonds [86–91], or electrostatic attraction [92,93], selective adhesion between NPs and the substrate can be attributed to the chemical bonding or electrostatic forces.

This method has at least four advantages: (1) catechol group coordinates with solid surfaces of nanocrystals in exchange with carboxyl group of carboxylic acid even at room temperature due to a high affinity of catechol to various metal oxides, such as Fe3O4, TiO2, or ZrO2 [94–100]. (2) Nanocrystals with a high affinity for solvents can be prepared. No process of attaching free DHCA onto nanocrystals as linkers is included in this method. (3) A comparison with the process using free DHCA shows that it is favorable to use DHCA chemically bonded with a substrate in terms of chemical equilibrium. Decanoic acid attached to the hybrid nanocrystals can be exchanged with DHCA chemically bonded with the substrate without any harmful effect of free decanoic acid in the solvent. (4) This method is highly versatile and can be divided into three steps—synthesis of highly dispersible nanocrystals at high temperature, modification of the substrates, and the assembly of the nanocrystals on these substrates at room temperature. Nanocrystals modified with carboxylic acid can be used in this method regardless of their cores since a high affinity of catechol group to various metal oxide NPs was used.

The surface treatment of silicon was carried out so that catechol group is exposed to outside. In this experiment, first, 3-aminopropyltriethoxysilane (3APTS) was attached on the surface to produce amine group termination. Secondly, the samples were immersed in N,N-dimethylformamide containing DHCA, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, a condensation agent, and N,N-dimethyl-4-aminopyridine, a catalyst, for the condensation of carboxyl group of DHCA with amine group on the substrate.

![Figure 27](image-url) Two-dimensional nanocrystals with heterogeneous ligand structures are fabricated on the surface by modifying substrates with DHCA. When cerium oxide nanocrystals modified with decanoic acid are adsorbed on the DHCA-terminated surface, decanoic acid at the bottom of the nanocrystals is locally exchanged with DHCA, and the nanocrystals covalently bond with the substrate. For color version of this figure, the reader is referred to the online version of this book.
Monolayer-controlled CeO$_2$ nanocrystal films were fabricated from solution, in which large amount of the hybrid nanocrystals were dissolved and subsequently rinsed with cyclohexane, using a silicon substrate modified with DHCA. To obtain dense hybrid nanocrystals in cyclohexane, the solvent was gradually evaporated at a pressure of 500 hPa, regulated by an evaporator. The scanning electron microscope (SEM) image presented in Figure 28(a) shows as-adsorbed multilayers of hybrid nanocrystals on the substrate. Three-dimensional self-assembly structures of hybrid nanocrystals were fabricated on the modified substrate. Monolayer-controlled structures of nanocrystals could not be obtained by mere self-assembly of nanocrystals in solution unless the density of nanocrystals in the solvent was carefully regulated. The top layer of this self-assembly structure was not fully covered with nanocrystals. Island structures from incomplete coverage of the top layer were clearly seen in a topographic image, obtained by atomic force microscope (AFM), as in inset of Figure 28(a). The SEM image presented in Figure 28(b) shows that a sufficiently high-coverage monolayer-controlled hybrid nanocrystal film was fabricated by subsequent rinse (coverage of 80.3 ± 1.5%). Hybrid nanocrystals on the hybrid nanocrystals were washed away leaving monolayer structures of hybrid nanocrystals on the modified surface. Selective adhesion between nanocrystals and the substrate arises from uniform modification of the substrate with DHCA and the on-site ligand exchange. In inset of Figure 28(b), the topographic image obtained by AFM is presented. From topographic analysis performed by AFM, the residual nanocrystals remained on the nanocrystal monolayer was sufficiently low (~10% from the image). Those nanocrystals appear as bright spots in the AFM image. The heights of those nanocrystals were evaluated to be ~8 nm in the image.

6.3. Controlled Assembly of Nanocrystals

Recently, a trend has been set in the processing of a new class of coordination polymers, in which isolated metal ions or clusters are linked into arrays by organic ligands to make M–O–M connectivity in 1-, 2-, or 3-dimensions. Such hybrid frameworks (MOF: Metal Organic Framework) represent a new type of materials that can harness the advantages and versatility of both organics and inorganics. The new class of organic–inorganic hybrid materials have a tremendous application potential such as sensors, biological tags, catalysts, bioelectronics, solar energy conversion, mechanics, membranes, protective coatings, transparent pigments, electromagnets, cosmetics, etc. [101,102]. We are exploring a new type of hybrid framework with the organic modified NPs, as an extension of the concept of MOF.

Recent progress in synthesis methods of inorganic NPs facilitates their application as components for miniaturized devices [103]. A strategy to construct such devices is the controlled preparation of two-dimensional or three-dimensional superlattices, [104,105] that is, periodic arrays of NPs. Various studies have demonstrated superlattices of metal or compound NPs through van der Waals, electrostatic, or chemical bonding. Most of the lattices comprise simple spherical NPs, but various periodic structures have been prepared using binary components [105e–h] or designed chemical interaction [105i].

NPs can work as a metal center in MOF to form 2D or 3D structure. We believe that the use of polyhedral NPs can diversify composed superlattices, thereby enabling production of advanced nanomaterials and nanodevices. The polyhedral shape might enable anisotropic interparticle interaction, leading to a unique coordination number and controlled assembly of NPs. In this section, we show the one-step synthesis of cubic nanoassemblies composed from octahedral CeO$_2$ NPs [106]. Hexanedioic acid (HOCO(CH$_2$)$_4$COOH), which has two carboxylic groups, was added to reactants to adhere the NPs. Octahedral coordination of the primary CeO$_2$ NPs plausibly led to quasi-bcc packing and formed cubic nanoassemblies, which had a different shape from that of the primary NPs.
Experiments were performed using a lab-scale plug flow reactor [107]. The flow of an aqueous solution of 0.01 mol/L Ce(NO₃)₃ was mixed with another flow of heated water to synthesize cerium oxide NPs hydrothermally. Hexanedioic acid was dissolved in and supplied with the Ce(NO₃)₃ solution to bind the NPs during synthesis. The typical reaction temperature, time, and pressure were, respectively, 250 °C, 1.9 s, and 25 MPa.

Figures 29(a) and (b) show SEM images of the cerium oxide NPs that were synthesized (a) with and (b) without hexanedioic acid. The products that were synthesized with hexanedioic acid had a cubic shape with rugged surfaces. In contrast, the products synthesized without hexanedioic acid were octahedral with smooth surfaces. The X-ray diffraction spectra showed that both products’ crystalline structure resembled that of CeO₂, suggesting that Ce³⁺ ion was

**FIGURE 29** (a) and (b) Depict SEM images of CeO₂ products that were synthesized (a) with and (b) without hexanedioic acid (HOCO(CH₂)₄COOH); Figures, 29(c)–(e) show magnified TEM images and a reduced SEM image of the CeO₂ products that were synthesized with hexanedioic acid (the same as Figure 29(a)), respectively.
oxidized by NO$_3^-$ species at the reaction conditions [108]. They had the CaF$_2$ structure, which favors octahedral crystalline shape with energetically stable [110] surfaces. Because of CeO$_2$ NPs’ practical importance, many studies have been performed to synthesize them [109,110]. Among them, several groups have reported that the synthesis of CeO$_2$ in the presence of an organic surface modifier produced cubic CeO$_2$ nanocrystals whose (100) surfaces were stabilized by capping organic molecules [111].

However, in our study, the transmission electron microscopy (TEM) image of the cubic products (Figure 29(c)) showed that they were composed from smaller octahedral primary NPs with average size of ca. 7 nm. In addition, the primary octahedral NPs assembled together, arranging their orientation (Figure 29(d)). The cubic nanoassemblies maintained their structure even when they grew to several hundred nanometers, as shown in Figure 29(e). Based on these results, we concluded that the small octahedral CeO$_2$ NPs assembled in an arranged manner to produce the cubic nanoassemblies.

This structural change of the CeO$_2$ products was caused by the presence of hexanedioic acid during hydrothermal synthesis. We have shown that the surface of metal oxide NPs was chemically modified by carboxylic acids or amines during their hydrothermal synthesis from inorganic salts of metal ions [28,51,56,76,112a]. Similarly, basic CeO$_2$ might react with carboxylic acid during hydrothermal synthesis. We measured FTIR spectra to verify that hexanedioic acid reacted with and existed on the surface of the CeO$_2$ products.

A comparison of Figure 30(a) and (b) indicates that the CeO$_2$ cubic nanoassemblies that were synthesized with hexanedioic acid exhibited larger peaks at around 2933 and 2864 cm$^{-1}$. These peaks correspond respectively to the asymmetric and symmetric modes of C–H stretching. The sample also had major peaks at around 1531 and 1421 cm$^{-1}$. These peaks were respectively assigned to the asymmetric and symmetric modes of carboxylate, i.e. COO$^{-}$ groups showing that the carboxyl group of hexanedioic acid bound with a Ce atom on the surface of the CeO$_2$ NPs.

The primary NPs in the cubic nanoassemblies (Figure 29(c) and (d)) were much smaller than the octahedral NPs that were synthesized without hexanedioic acid (Figure 29(b)). The capping of the growing surface of the octahedral NPs by hexanedioic acid might have shrunk the primary NPs.

In contrast, the CeO$_2$ cubic nanoassemblies that were synthesized with hexanedioic acid showed no peak at around 1800–1700 cm$^{-1}$, thereby indicating the absence of unreacted carboxyl group as well as ester. Therefore, a hexanedioic acid molecule is bound with two CeO$_2$ NPs, which means that the octahedral CeO$_2$ NPs were connected by hexanedioic acid to have octa-nearest neighbors.

In this study, unlike previous studies using polyhedral NPs, [110b,113] the primary NPs assembled to form a secondary structure with a different shape from that of the primary NPs. Based on the above discussion, the primary CeO$_2$ NPs are inferred to have formed the cubic nanoassemblies instead of agglomerates without distinct shapes because octahedral CeO$_2$ NPs might have acted as octa-coordinated atoms in the body-centered cubic (bcc) lattice. However, as indicated by the truncated octahedral form of the Wigner-Seitz cell of the bcc crystal, the bcc arrangement of octahedral NPs does not completely fill the three-dimensional space. Perhaps for that reason, the synthesized nanoassemblies have some irregularity, as discussed in detail elsewhere [114].

A similar approach has been applied for Fe$_3$O$_4$ NP synthesis. In this case, the Fe$_3$O$_4$ NPs-assembled clusters were synthesized by using 3,4-dihydroxycinnamic acid (DHCA) (inset of Figure 31) that have catechol and carboxylic group as surfactant and interparticle linker [115]. The precursor used was ferric sulfate—and potassium hydroxide with the concentration of 0.05 molar each—and the temperature was 280 °C. Figure 31 shows the field emission secondary electron microscopy (FESEM) image. Tiny particles were assembled into structures with a raspberry-like morphology whose insides were filled. The primary particles with a diameter of approximately 20 nm were observed at the surface of the larger particles. This size was almost the same as that of the crystallite evaluated by XRD peaks using Scherrer’s equation. The NP cluster size increased from 50 to 400 nm with increasing reaction time, as shown in Figure 31. Based on the FT-IR measurement, we conclude that the primary particles were linked together through the DHCA with the COOH and the catechol groups. Since the outer surface of the cluster has those hydrophilic functional groups, the clusters were well-dispersed in water. The clusters were attracted by a magnet, but the dispersed state could be recovered easily by
removing the magnet (Figure 32), since the sizes of the Fe₃O₄ NPs are so small as to have superparamagnetic property. The response sensitivity could be controlled by changing the cluster size.

The surface of Fe₃O₄ NPs-assembled cluster was covered by COOH groups, and therefore, biomolecules such as proteins and DNA, or drugs could be conjugated. Therefore, the prepared Fe₃O₄ clusters should be especially

---

**FIGURE 31** TEM image of the synthesized Fe₃O₄ NP-assembled cluster after reaction times of (a) 10, (b) 30, (c) 60, (d) 90, and (e) 120 min. For color version of this figure, the reader is referred to the online version of this book.

**FIGURE 32** Dispersions of Fe₃O₄ nanoclusters (1 wt%) in water (a) without a magnetic field, (b) with a magnetic field for 1 min, and (c) after the magnetic field was removed. The size of Fe₃O₄ nanoclusters is 50 nm (left) and 400 nm (right). For color version of this figure, the reader is referred to the online version of this book.
useful in the biomedical field as tags or carriers of magnetic manipulators.

7. HYBRID NANOMATERIALS

7.1. Needs for NPs-Hybrid Polymers and Breakthrough Point

Hybrid materials of polymer-inorganic NPs (NPs) that show the incompatible multifunctions (light but mechanically strong, thermally conductive but flexible, high refractive index but flexible) are required in many applications. So far, tremendous efforts were required to fabricate hybrid nanomaterials, but in most cases, the dispersion of inorganic NPs in organic solvents or polymers was problematic. To achieve these incompatible multifunctions, controlling the interface between NPs and the solvents/polymers is essential.

The silane coupling method is widely used to combine organic molecules with inorganic materials. However, for treating NPs via this method, they should be first dispersed in an organic solvent, which is possible only in the case of dilute systems. Another drawback of this method is the silica layer formation on the surface of NPs, which sometimes changes the original properties of NPs, such as reduction of refractive index and thermal conductivity.

To solve the problem, we proposed a new method to synthesize Organic molecule modified NPs by using supercritical water [48,51,76]. Organic capping agents and inorganic precursors form a homogeneous phase in supercritical water and the water molecules serve as acid or base catalysts for various organic reactions. To obtain high affinity between NPs and the polymer matrix, development of a new monomer that possesses two functions, namely affinity for NPs and polymerization, is required.

In this study, we propose a method to fabricate hybrid materials and demonstrate the performance of these “superhybrid nanomaterials” that show incompatible multifunctions.

The surface-modified NPs synthesized by this method can be easily recovered by extraction with organic solvents from the water suspension phase and well-dispersed in organic solvents. Transparent solutions were obtained for the products; the solution of decanoic acid-capped ceria NPs dispersed in cyclohexane is shown in Figure 33. DLS analysis confirmed that the solution contained no aggregates but only individual particles having concentrations up to 60 wt%.

The viscosity of the nanocolloidal solution was measured by a viscometer (Brookfield). In general, nanocolloidal solutions show extremely high viscosities at high concentrations, because of the NP cluster formation. However, the surface-modified NPs show low viscosity (8 cP) even at 60 wt% concentration. This unusual phenomena may possibly be attributed to the high surface modification achieved using the supercritical in situ synthesis method. The well-dispersed NP solutions suggest a high affinity between NPs and the solvent, while the low viscosity implies a poor solvation structure around NPs. Further study is necessary to elucidate the mechanism.

7.2. Superhybrid Materials

Recently, the size of the devices has reduced while power supply has increased. Thus, how to release the heat from the devices is the bottle neck in developing future power devices. Thermal conductive materials that have high thermal conductivity, electrical insulation, adhesiveness, and fabricability are required. Polymer-ceramic (BN, AlN, SiN or Al2O3) composites are now being developed. To increase the thermal conductivity of the hybrid materials,
the filler content should be increased. However, when the filler content is increased above 50 wt%, the viscosity of the hybrid materials increases drastically, which reduces the fabricability. Furthermore, void formation occurs, which decreases the electrical insulation as well as thermal conductivity. Thus, because of these incompatible functions, creating high heat-transfer hybrid materials is a difficult task.

The surface modification of BN particles by the supercritical method improves the affinity for the polymers, resulting in high BN content of the hybrid materials (more than 90 wt%) without any void formation. The thermal conductivity is high at 40 W/m/K, which is one order of magnitude higher than the currently used materials (4 W/m/K). The devices made from these BN particles have high fabricability, electrical insulation, and flexibility, as shown in Figure 34 [48].

Controlling the refractive index while maintaining the fabricability of the polymer matrix is needed in optical materials. By transparent dispersion of high refractive-index NPs, such as TiO$_2$ and ZrO$_2$, in the polymers, the refractive index of the polymers can be increased. Also in this case, surface control is the key to disperse NPs. The organic surface modification of NPs increases the affinity for the polymer matrix and affords the fabrication of flexible polymer films that have high refractive-index transparency [48].

### 7.3. Hybrid Materials of NPs and Biomolecules

In Section 6, self-assembly of NPs with organic capping agents are introduced. Oriented surface modification on polyhedral NPs leads to the formation of superlattice structure. It is well known that biomolecules such as DNA, peptides, proteins, and polysaccharides forms self-assembly structure, as well. Now, our approach is to combine these two assembling functions.

We have succeeded in synthesizing cyclic diphenylalanine (cyclo-FF (cFF), Figure 35(a)) crystals from l-phenylalanine (l-Phe), with a high yield by hydrothermal treatment at 220 °C and 22 MPa. Especially, excess amount of cFF synthesis caused the spontaneous assembly of cyclo-FF crystals with a rod-like morphology with widths of 200 nm and lengths of 2–10 μm, (Figure 35(b)). Cyclo-FF (cFF) nanorods grew along the x-axis, and intermolecular hydrogen bonds of diketopiperadine are formed through amide group linkages along the x-axis (Figure 35(c)). These intermolecular hydrogen bonds of the amide groups may serve as a driving force in the formation of anisotropic cFF nanorods [116].

Under the reaction atmosphere for forming the cFF nanorods, we introduced the iron oxide precursor ferric sulfate (Fe$_3^{2+}$), with controlling pH to be around 9. When the l-Phe concentration is lower, cFF nanorods were not formed but needle-shaped Fe$_3$O$_4$ particles modified with l-Phe are formed. On the other hand, at the higher l-Phe concentration, interestingly, needle-shaped Fe$_3$O$_4$ particles are stacked on the surface of cFF nanorod. The conjugated l-Phe and synthesized cFF have an aromatic ring at their side chain. These aromatic rings possibly undergo $\pi-\pi$
interactions with each other to allow the self-assembly of NPs on to the nanorods [116], as shown in Figure 36.

8. SUMMARY

Supercritical hydrothermal synthesis was invented by Adschiri in 1989 [117]. Since then, gradually, this method is getting used all over the world, not only by the researchers but also in industries. In 1994, first commercial plant has been constructed, and now 600 t/year of process is running.

This method has both of the specific features of "hydrothermal synthesis" and "supercritical fluid". Hydrothermal synthesis is regarded as a most widely used method to fabricate materials, especially for metal oxides or sulfates. The conditions to synthesize materials are similar to the deep underground where most of the minerals are formed. Any materials that can be found on the earth can be synthesized under the conditions. Hydrothermal condition provides high solubility for these materials to be crystallized, and controllable atmosphere for the crystal growth. On the other hand, for NP synthesis, lower solubility is required. Supercritical condition provides the lower solubility atmosphere by reducing pressure or increasing temperature, with keeping the availability of hydrothermal synthesis for variety of materials. Furthermore, under the condition, hydrogen gas or oxygen gas can be mixed together to control the oxidizing/reducing atmosphere.

One of the problems of using NPs is the handling of NPs. In most cases, NPs are dispersed in solvent before the application such as spraying, coating, or mixing with polymers. However, it has been considered a difficult task to disperse the NPs in solvent because of the higher surface energy. The key is the surface modification of NPs. However, in hydrothermal condition, organic modification is limited only for the smaller organic compounds. Under the supercritical state, a homogeneous phase is formed so that we can expect the in situ organic—inorganic reaction at a high concentration. This was the key to open a door of fabricating NPs-polymer composites.

Synthesis of NPs—NPs superlattice structure or biomolecule-directed self-assembly structure is also succeeded. We believe this method can provide the new area of metamaterial fabrications.

This chapter summarized the state of art of the supercritical hydrothermal synthesis method. For the development of supercritical fluid technology, establishment of reaction engineering, mechanism elucidation, and thermodynamics of NPs systems will be required.

REFERENCES


Shin Y-H, personal communication.


Chapter 11.1.5 Supercritical Hydrothermal Synthesis


(c) Cabañas A, Darr JA, Lester E, Poliakoff M. Continuous hydrothermal synthesis of inorganic materials in a near-critical water flow reactor; the one-step synthesis of nano-particulate Ce$_{1-x}$Zr$_x$O$_2$($x = 0–1$) solid solutions. J Mater Chem 2001;11:561–8;
(g) Yu TY, Joo J, Park YI, Hyeon T. Large-scale nonhydrolytic sol–gel synthesis of uniform-sized ceria nanocrystals with spherical, wire, and tadpole shapes. Angew Chem Int Ed 2005;44:7411–4;
1. INTRODUCTION

Powders, or their constituent particles, are the fundamentals of various industrial and technological applications, including powder metallurgy, catalysis, luminescence, biological engineering, and so forth. They may also serve as the building blocks of hyperstructures, such as photonic crystals, to attain additional functionalities. Among many others, controlled chemical purity, phase structure, particle size/size distribution, and crystallinity are the most important aspects of the powders toward practical applications. Solution processing (typically chemical precipitation) has been the most widely adopted technique in the laboratory and industry for powder production. While having its own advantages, the synthetic methodology frequently yields precursor products of low crystallinity or an amorphous nature, and thus, a subsequent annealing is required to yield the target material or to promote further crystallization. This, in turn, would unavoidably lead to appreciable coarsening and sintering of the particles, especially when the particle size goes down to the nanometer regime. To save this problematic annealing stage, thermal plasma proves to be a very efficient and attractive powder processing technology. Thermal plasmas, either generated in DC arc torches or radiofrequency (RF) induction torches, possess the following unique properties [1]: (1) high temperatures of up to approximately 15,000 K. As the thermal plasmas are generated at relatively high pressures close to the ambient one, they have high densities and are considered to be at equilibrium (local thermal equilibrium). That is, the temperature of the heavy particles (atoms, molecules, and ions) is almost equal to that of the light species (the electrons); (2) high concentrations of chemically reactive radicals that facilitate chemical reactions; (3) superfast quenching (~10^6 K/s) in the tail flame region. With these, well-defined nanoparticles of interesting structural and morphological features can be generated in one step for various materials forms (metallic, ceramic, and ceramic/metallic composites) and a wide range of chemical compositions. Examples include the TiN/Ti5Si3 and TiN/M (M = Ag, Fe, Co, Ni, etc.) composite nanoparticles produced via Ar–H2–N2 DC arc plasma ablation of the Ti–Si and Ti–M parent alloys [2–6]. Some of the composite nanoparticles are shown in Figure 1, from which it can be perceived that the finely decorated TiN particles of excellent dispersion and ultrafine sizes are hardly obtainable in one step by any other processing technique. Further, the TiN/M(Si) combination can be readily controlled by varying the Ti/M atomic ratio of the parent alloy; thus, a wide range of chemical compositions may be obtained for the composite nanoparticles toward practical applications. Besides their interesting H2 adsorption/desorption properties, the composite nanoparticles may be consolidated into bulk ceramics at reduced temperatures for unique electrical, thermal, and tribological properties.

Compared with DC arc discharge, RF induction plasma may present some more merits. First, the RF plasma is inductively heated by an RF electrical power of several tens of kilowatts and a frequency of megahertz order and the plasma torch may have a diameter of 30–40 mm, and thus,
the plasma volume is relatively large. Second, the axial plasma velocity may be as low as 10 m/s, and thus, the residence time of the generated chemical species in the high-temperature region of the plasma flame may be as long as some tens of milliseconds, long enough for chemical reactions to proceed. Furthermore, the generation of RF induction plasma does not use electrodes and thus avoids the possible contamination of the product. Finally, the chemical properties of the RF plasmas may be tuned by adopting a wider range of plasma generating or sheath gases and can be oxidative, reductive, or neutral. With these, the RF thermal plasma provides a unique field for a wide range of materials processing. As a heat source, the RF thermal plasma has been used in the industries for the spheroidization and purification of materials with high melting points and of relatively large particle sizes (~50 to 100 μm). For example, 30 kg of refractory tungsten metal particles can be treated per hour, corresponding to 1 ton per week [7]. Spheroidization of irregularly shaped commercial TiC particles can similarly be attained utilizing the heating effects of the RF thermal plasma [8]. As a reaction field, RF thermal plasma provides several unique advantages in the synthesis of well-controlled ultrafine particles: (1) its high processing temperature allows the use of materials of various states (solid, liquid, and gas) as the precursors for powder synthesis, and (2) its superfast quenching at the tail flame enhances doping and solid-solution formation and thus may extend the functionalities of the pristine materials via doping. In this chapter, we discuss the synthesis of ultrafine TiO₂ particles using the two typical types of precursors of solid TiN and titanium alkoxide solutions, with an emphasis on the reaction mechanism, phase structure manipulation, particle size/size distribution control, and the doping of TiO₂ by various metallic and nonmetallic elements for extended/enhanced functionalities.

The plasma apparatus used for TiO₂ synthesis mainly consists of four parts: a water-cooled induction plasma torch (Model PL-50, TEKNA Plasma System Inc., Quebec, Canada), a 2-MHz RF power supply system (Nihon Kosho Co. Ltd., Kanagawa, Japan), a water-cooled cylindrical reactor with a cone-shaped bottom, and a stainless-steel filter connected to the reactor through an outlet tube. Details of the plasma reactor may be found elsewhere [9]. The initial TiN powder (~28 μm, Nippon New Metals Co., Osaka, Japan) was injected from the top of the plasma torch through a water-cooled probe into the center of the plasma plume by using powder carrier gas. The feeding rate of TiN was controlled at about 1.5 g/min. The Ar/O₂ plasmas were generated by mixing O₂ in the Ar sheath. Pure Ar was used as the carrier gas to deliver the TiN precursor particles. The flow rates of the central (Ar gas), sheath, and carrier gases were set at 30, 90, and 5 L/min, respectively. The O₂ flow rate in the plasma sheath was varied to investigate its effects on the characteristics of the resultant powders. The plate power was 25 kW and the reactor pressure was controlled at approximately 500 Torr. The generated TiO₂ particles are collected from the inner wall of the reactor and from the filter.

**2. CONTROLLED RF THERMAL PLASMA PROCESSING OF TiO₂-BASED ULTRAFINE PARTICLES**

2.1. Particle Synthesis with Solid TiN Particles as the Precursor

TiN is well known as a highly refractory inorganic compound. It, however, can be readily oxidized by the Ar/O₂ thermal plasma to yield TiO₂ [9]. The particles produced by the in-flight oxidation of TiN mainly deposit at the inner wall of the water-cooled reactor under low O₂ inputs while at the filter under high oxygen additions,
suggesting that the products have different characteristics under different processing conditions. Morphology observations (Figure 2) indicate that the as-synthesized powders typically contain particles of two distinctly different sizes and shapes. The coarse (micron size or larger) and the fine particles powders can be effectively separated via sedimentation from ultrasonically dispersed alcoholic suspensions of the total powder. This operation allows one to perform more detailed formation mechanism and phase structure studies of the products.

Figure 2 shows typical morphologies of the starting TiN particles, the total powder after plasma treatment, and the separated coarse and fine particles. The irregularly shaped TiN particles apparently undergo spheroidization and size reduction on Ar/O2 thermal plasma treatment, and the extent of morphological changes increases with increasing the O2 input in the plasma sheath. At 20 L/min of O2 flow, the large particles turn into almost perfect spheres with a maximum diameter of approximately 5 μm, a size significantly smaller than that of the starting TiN particles (~28 μm on average). The separated fine particles all have sizes in the range of approximately 50–300 nm, showing relatively weak dependence of the particle size on the O2 flow rate.

The coarse particles, separated from the total powder synthesized under various O2 flow rates, are observed via optical microscopy (Figure 3). It is also clearly seen that the particles steadily get smaller and more rounded with an increasing O2 input. Cross-section views (the insets) revealed a composite nature of the particles produced at low

**Figure 2** FE-SEM morphologies of the starting TiN precursor powder and the oxidation products.

**Figure 3** Optical micrographs showing morphologies of the coarse particles separated from the total powder obtained under various O2 inputs in the plasma sheath, with (a) 2.5 L/min, (b) 5.0 L/min, and (c) 10 L/min. The insets are the cross-section views of the particles. For color version of this figure, the reader is referred to the online version of this book.
O₂ flow rates up to 5.0 L/min. The golden and pinkish purple colors observed in the cores (Figure 3a and b) suggest the presence of TiN arising from incomplete oxidation. The black, blue, or yellow colors of the shells may imply that they are primarily composed of the rutile phase. It is known that the rutile crystallites exhibit dark blue or black color depending on the amount of oxygen vacancies in the crystal lattice, and may turn pale blue or yellow on reoxidation. Increasing the oxygen flow rate to 10 L/min results in an almost complete oxidation, and the golden color corresponding to TiN is no longer observed in the particle cores (Figure 3c). Besides, the whole cross-section appears white, suggesting a decreased amount of oxygen vacancies in the crystal structure. The TiO₂ microbeads shown in Figure 3c, having over 80 wt% of the anatase phase and a particle size of approximately 5 μm, are believed to be hardly obtainable by other synthetic techniques.

Phase structure analysis via XRD of the products indicated that, under the lowest O₂ input of 2.5 L/min, the large particles (Figure 2) are mainly composed of TiN and rutile, along with the presence of trace anatase and a Ti₄O₇ suboxide, whereas the fine particles are a mixture of the rutile and anatase phases free from contaminations by either TiN or Ti₄O₇. The amount of TiO₂ in the coarse particles steadily increases at a higher O₂ flow rate, due to the enhanced oxidation, and TiN disappears from the big particles at 10 L/min of O₂ flow or above, conforming to the observations shown in Figure 3. The resultant TiO₂ is a mixture of the anatase and rutile phases in each case, but anatase exhibits a steadily increasing content at a higher O₂ input and finally becomes the dominant phase in the powder produced at 17.5 L/min of O₂ input. The fraction of fine particles in the total powder also increases toward a higher O₂ input, and the highest value attained is approximately 86 wt%. The anatase content of the fine particles shown in Figure 2 keeps constant at approximately 60 wt% up to ~5 vol% of O₂ in the plasma stream, after which it increases gradually to approximately 90 wt% at 13 vol% of O₂. Anatase is not detected in the coarse particles up to approximately 2 vol% of O₂ in the plasma, but its content similarly increases on raising the O₂ input in the plasma sheath, finally reaching approximately 81 wt% at 13 vol% of O₂. The preceding results are summarized in Figure 4.

Based on the above microstructure and phase structure analyses, the formation and phase selection of TiO₂ particles on Ar/O₂ thermal plasma oxidizing TiN might be depicted in Figure 5 and discussed later.

The two types of particles (coarse and fine ones) observed in Figure 2 suggest their different formation pathways. The oxidation of TiN would start at the peripheries, and particularly the corners, of the irregularly shaped TiN particles on plasma treatment. Due to the relatively low melting point of TiO₂ (~1900 °C), the oxidation layer is expected to undergo subsequent melting and evaporation. Fine particles are then formed as a result of condensation from the evaporated species. At lower O₂ inputs (such as 2.5 and 5.0 L/min), the aforementioned processes are largely limited to the surface layers because of the short duration time of the particles in the thermal plasma, thus leading to a preferential rounding-off of the corners and edges and hence the formation of more rounded particles with a size only slightly smaller than that of the initial TiN. Coarse composite particles consisting of TiN cores and TiO₂ shells are thus formed as a result of incomplete oxidation. For the same reason, only a small amount of fine particles is produced due to the limited evaporation. Increasing the oxygen input promotes the oxidation of TiN, and the oxidation layers are impelled to vaporize due to the additional heat arising from the exothermic TiN oxidation,
leading to a rapid decrease in the size of the coarse particles and an increased portion of the fine ones. Under the O₂ input rate of 10 L/min, the TiN particles are completely oxidized to TiO₂ melts. Rapid quenching of the melts at the tail flame would have yielded the nearly perfect TiO₂ spheres of greatly decreased diameters.

It is well known that rutile is the most stable polymorph of TiO₂ and that anatase converts to rutile on being annealed in the temperature range of approximately 500–1000 °C [10,11]. The direct production of almost phase-pure anatase despite the high plasma temperature implies that the phase selection of TiO₂ in thermal plasma processing is different from that in conventional annealing, and may be interpreted as follows. As mentioned earlier, the surfaces of the TiN precursor particles are rapidly oxidized and melted in the thermal plasma, leading to the formation of composites consisting of molten TiO₂ shells and TiN cores. The TiO₂ shells tend to solidify as rutile due to the oxygen deficiencies arising from the low O₂ input and the inward oxygen diffusion from the shells to the TiN cores. There has been plenty of evidence, either derived from the phase transformation studies of anatase via conventional annealing or from flame processing of TiO₂, showing that oxygen vacancies tend to stabilize the rutile crystal structure [12,13]. Higher O₂ inputs diminish the oxygen deficiency, thus leading to more anatase in the coarse particles. Besides, a higher O₂ input improves the thermal conductivity of the plasma and thus yields more deeply undercooled melts at the plasma tail. This would also promote anatase nucleation [14]. Nanoparticle formation via gas-phase condensation is also affected by the plasma composition and, similarly, is characterized by more anatase in an O₂-rich atmosphere and more rutile in an O₂-lean atmosphere. The above oxygen-deficiency-dependent phase selection is supported by the fact that, if adopting Ar/O₂ mixture as the carrier gas and Ar/H₂ mixture as the plasma sheath, the fine particles condensed from the gas phase become pure rutile under proper processing conditions, as shown in Figure 6. It has been widely observed that the usually nanosized anatase crystallites undergo rapid coarsening on phase transformation to rutile. That is, ultrafine rutile crystallites are hardly obtainable via conventional annealing of anatase. The Ar/O₂/H₂ plasma generated rutile particles, however, are mostly well below 100 nm in diameter and are well dispersed (Figure 7), demonstrating again the advantages of thermal plasma processing.

### 2.2. Thermal Plasma Processing of TiO₂-based Nanoparticles with Liquid Precursors

In view of the fact that when solid precursors, such as TiN, are used for powder synthesis, the insufficient evaporation of molten droplets leads to micron-sized spheres; liquid precursors are thus adopted to produce much finer TiO₂ particles of more uniform sizes. In this case, the liquid precursor is delivered into the center of the plasma plume through an atomization probe, at the tip of which the liquid is nebulized into mists by the carrier gas. The tiny precursor drops then undergo an instantaneous and more sufficient oxidation by the high temperature of the Ar/O₂ thermal...
plasma, followed by greatly enhanced evaporation of the oxidation products. Ultrafine TiO$_2$ particles are then formed via nucleation, growth, and coagulation, as schematically shown in Figure 8.

Titanium tetrachloride and alkoxides are the most common titanium compounds used in TiO$_2$ synthesis. The chloride is highly corrosive and highly sensitive to air moisture while the latter, besides the moisture sensitivity, has increasing viscosity toward longer carbon chains. These make the titanium compounds inconvenient for plasma processing. The problem, however, can be solved by employing diethanolamine (DEA) as a chelate for the alkoxides. For example, the solution obtained by reacting 0.1 mol of titanium tetra-n-butoxide (TTBO) and 0.4 mol of DEA is highly stable against water under alkaline conditions[15]. On the other hand, citric acid reacts with most metal cations to form chelates that are stable under high pH values. A combination of the above two strategies allows one to obtain liquid precursors of various chemical compositions and proper viscosity for the thermal plasma processing of doped TiO$_2$ ultrafine particles. Examples are shown in Figure 9 for some typical combinations. It is seen from the figure that the TiO$_2$ particles indeed turn much finer and more uniform by employing liquid precursors for synthesis. The rounded particles of approximately 100 nm are believed to be formed via the solidification of molten droplets while the others coagulated from the vaporized species. The two different formation pathways are owing to the different trajectories, and thus different thermal histories, of the oxidation products in the plasma field. The advantages of thermal plasma processing become more evident when one wants to achieve an efficient doping of TiO$_2$, an excellent semiconductor host, for extended functionalities. For example, TiO$_2$ can be doped with much larger lanthanide (Ln) cations (the size difference between Ti$^{4+}$ and Ln$^{3+}$ is well above 30%) via thermal plasma oxidation of liquid precursor mists containing TTBO and Ln nitrate [16,17], although the dopant concentration seems to be limited to 0.5 at%. Even higher dopant contents would lead to the formation of Ln$_2$Ti$_2$O$_7$ pyrochlore impurities in the final products. Despite the low achievable doping level, lattice-site doping of TiO$_2$ with Ln$^{3+}$ is almost impossible via the traditional solution-based processing techniques of chemical precipitation and hydrothermal treatment [17]. Optical spectroscopy studies confirm that the properties of the TiO$_2$ semiconductor host and the Ln$^{3+}$ are indeed integrated, and the TiO$_2$:Eu$^{3+}$ crystallites exhibit an efficient nonradiative energy transfer from the TiO$_2$ lattice to the Eu$^{3+}$ activators [16]. As a result, the TiO$_2$:Eu$^{3+}$ nanoparticles show the typical $^5D_0 \rightarrow ^7F_j$ transition emissions of Eu$^{3+}$ either by exciting the TiO$_2$ host lattice or by directly exciting the intra 4f$^6$ transitions of Eu$^{3+}$. More importantly, much stronger emission can be obtained via host–lattice excitation through the energy transfer. Similar energy-transfer-enhanced luminescence is also observed from the Eu$^{3+}$ doped TiO$_2$ nanoparticles synthesized with thermal plasma [17]. TiO$_2$ crystallites doped with transition metal cations may find potential applications as dilute magnetic semiconductors (DMS) and as efficient photocatalysts [18,19]. With alcoholic solutions containing TTBO and ferrocene (Fe(C$_5$H$_5$)$_2$) as the liquid precursor, thermal plasma processing allows a significantly higher amount (up to 20 at%) of iron to enter the TiO$_2$ lattice [20]. Fe$^{3+}$ doping was found to significantly promote rutile crystallization (from ~20 wt % at 0 at% of Fe$^{3+}$ to 80 wt% at 5 at% of Fe$^{3+}$), owing to...
the increased amounts of oxygen vacancies created to balance the charge difference between Fe$^{3+}$ and Ti$^{4+}$. Successively lowered direct and indirect bandgaps of TiO$_2$ were observed on Fe$^{3+}$ doping. Coupled characterizations via Raman spectroscopy and HR-TEM indicate that homogeneous TiO$_2$:Fe$^{3+}$ solid solutions only exist up to approximately 2 at% of Fe$^{3+}$, above which the oxygen vacancies and the Fe$^{3+}$ dopants tend to aggregate to form extensive shear defects [20]. Photocatalytic studies via bleaching methyl orange solutions under visible light illumination showed that the catalytic performance does not steadily improve toward a higher Fe$^{3+}$ incorporation, despite the successively narrower bandgaps, and an optimal Fe$^{3+}$ concentration exists for a certain particle size [21].

The supersaturated TiO$_2$:Fe$^{3+}$ solid solutions decompose by conventional annealing to form a mixture of pseudo-brookite (Fe$_2$TiO$_5$) and rutile in air and a mixture of rutile and ilmenite (FeTiO$_3$) or, depending on the Fe$^{3+}$ content, a mixture of rutile, ilmenite, and pseudobrookite in high vacuum. The amount of Fe$^{3+}$ that remains in the rutile lattice after annealing is around 2 at%, in accordance with the value determined for TiO$_2$:Fe$^{3+}$ homogeneous solid solutions. Magnetic investigations showed that the TiO$_2$:Fe$^{3+}$ crystallites are primarily of paramagnetic in nature under room temperature, possibly due to the poor coupling ability of the 5$d$ electrons of the Fe$^{3+}$ dopant [20]. Cobalt-doped TiO$_2$ (TiO$_2$:Co$^{2+}$) is widely viewed as one of the most promising type of DMS materials. Its origin of ferromagnetism, however, has long been under debate. Because most of the materials that show good ferromagnetism at room temperature are prepared under reducing conditions and by employing suitable liquid precursors. For example, TiO$_2$ nanopowders doped with 0.3 wt% of N can be produced via thermal plasma oxidation of TiN, using pure O$_2$ as the carrier gas, Ar as the central gas (plasma generation gas), and N$_2$ as the plasma sheath. The thus-made powders exhibit red-shifted band edge and enhanced absorptions of the visible light, and thus better performances in the bleaching of methyl blue solution under visible light irradiation, as shown in Figure 10.

With an aqueous solution of titanium trichloride as the precursor, doping TiO$_2$ with much bigger chlorine anions is achieved via Ar/$O_2$ thermal plasma oxidation [25]. Analysis via thermal desorption spectroscopy confirmed that, unlike the surface adsorption of chlorine species on the P25-type TiO$_2$ particles pyrolyzed from TiCl$_4$, the chlorine anions are located at the lattice sites of the plasma generated TiO$_2$ crystallites. Accordingly, the indirect bandgap of the chlorinated TiO$_2$ powder is effectively lowered from a typical value of approximately 2.90 eV to 2.65 eV. Chlorine doping enhances both the ultraviolet (UV) and visible light (Vis) absorptions of TiO$_2$, and thus better catalytic performances than P25 were observed for the plasma generated powders in the bleaching of methyl orange solutions either under UV or Vis irradiation. Further studies showed that the O$_2$ input in the plasma sheath affects the catalytic efficiency of the TiO$_2$ product. Higher O$_2$ inputs (up to 90 L/min) tend to eliminate surface Ti$^{3+}$ species, which act as recombination centers for the photoexcited electrons and holes, and thus yields better TiO$_2$ nanocrystallites for photocatalytic applications.

![FIGURE 10](image-url) A comparison of the light absorption behavior and catalytic efficiency of the TiO$_2$ powders with and without N-doping. For color version of this figure, the reader is referred to the online version of this book.
2.3. Size and Size Distribution Control of the TiO$_2$ Particles via Quench Gas Injection

As discussed earlier, the employment of misted liquid precursors for thermal plasma processing allows one to obtain particles of finer sizes and narrower size distributions. Due to the different trajectories, and accordingly thermal histories, of the mist droplets in the plasma field, the TiO$_2$ particles are formed via the two different routes of liquid solid and gas solid. As a result, the final resultant powders comprise nanoparticles and sub-micron-sized ones. Morphological uniformity of the plasma-generated particles can be further improved by injecting quench gases to influence the temperature distribution of the plasma field and the flow pattern of the plasma stream [26]. The quench gas may be injected from the top shoulder of the plasma reactor in the transverse (swirl) mode or from the bottom of the reactor in the counterflow mode, as shown in Figure 11a and b, respectively.

Modeling is performed to better understand the effects of quench gases, and the calculated streamlines and temperature fields in the reactor are illustrated in Figure 12. The plasma flow and temperature field are simulated using a two-dimensional k–ε turbulence fluid model [27]. The conservation equations for the mass, the momentum, the thermal energy, the turbulent kinetic energy and its dissipation rate, as well as the vector potential are solved using the SIMPLER algorithm [28]. The dots in the figure indicate the predicted trajectories of a test particle released at the tip of the central probe. Modeling is made by assuming that the test particle has negligible mass and infinite melting point and without considering the variations of local plasma velocity and temperature. In this case, the time-dependent velocity and the temperature of the particle can be obtained analytically [29]:

\[
\begin{align*}
    u_p & = u_g + (u_{p0} - u_g) \exp \left(-\frac{18 \mu_g}{\rho_p d_p^2} t \right) \\
    T_p & = T_g + (T_{p0} - T_g) \exp \left(-\frac{12 k_g}{\rho_p d_p^2 c_p} t \right)
\end{align*}
\]

in which \( u \) is the velocity and \( T \) is the temperature. \( u_p \) and \( T_{p0} \) denote the initial velocity and temperature of the particle, respectively; \( t \) is the time; \( \mu_g \) and \( k_g \) are the viscosity and the thermal conductivity of the plasma, respectively; \( \rho_p, d_p, \) and \( c_p \) are the corresponding particle density, diameter, and specific heat. The subscripts \( p \) and \( g \) refer to the properties of the particle and the plasma, respectively. For a nanosized particle, \( d_p \to 0 \), Eqn (1) reduces to \( u_p = u_g \) and Eqn (2) to \( T_p = T_g \), indicating that the nanosized particle moves with the plasma at the same

![FIGURE 11](image1.png)  
**FIGURE 11** Schematic illustrations showing the two different ways of quench gas injection.

![FIGURE 12](image2.png)  
**FIGURE 12** Streamlines and temperature distribution for: (a) no quench gas, (b) transverse injection of Ar at 100 slpm, (c) counterflow injection of Ar at 100 slpm, and (d) counterflow injection of He at 100 slpm. For color version of this figure, the reader is referred to the online version of this book.
velocity and temperature. The calculated particle trajectory and temperature history are expected to qualitatively describe the effects of the quench gas. Comparing Figure 12a and b, one can find that transverse injection of Ar does not significantly alter either the temperature field of the plasma or the trajectory of the test particle, though the flow pattern of the plasma stream is affected considerably as evidenced by the decreased size of the recirculation zone. When Ar is injected counter to the plasma plume, it profoundly affects the temperature field and also the trajectory of the particle (Figure 12c). The greatly decreased size of the plasma plume allows a shorter duration of the particle in the high-temperature zone and hence favors finer particle formation. It can also be seen from Figure 12c that the plasma plume is redirected toward the wall of the reactor, creating an additional low temperature \((T_g < 2000 \text{ K})\) recirculation zone in between the plasma plume and the injection point. This recirculation zone tends to trap nanosized particles due to the small masses of the nanoparticles, and thus consequently increase the growth time of the particles. Comparing Figure 12c and d, it can be seen that, when both injected counter to the plasma plume from the bottom of the reactor, the Ar quench gas reaches a higher position than the He gas does, owing to the larger atomic weight of Ar and hence its higher inertia, so the plasma plume is much shorter in the case of Ar injection. Although the low temperature recirculation zone is also seen in the case of He injection (Figure 12d), the test particle assumes a different trajectory, turning away from this recirculation region and thus avoiding the additional particle growth mentioned in the case of Figure 12c. Figure 13 presents the temperature histories of the test particle for the transverse and counterflow quench gas injections. It is found that the particle temperature in the case of counterflow injection is much lower than that in the transverse injection, indicating that counterflow injection has a much better quenching effect and finer particles can be expected. Moreover, in both the injection configurations, the particle temperatures, especially in the 1500—300 K region within which the nucleation and growth of TiO\(_2\) particles largely take place, are lower for He injection compared with Ar injection. From the numerical analyses one may conclude that counterflow quench gas injection is an effective way for the control of particle size and size distribution, and He is more effective than Ar.

The effects of quench gas injection on the size and size distribution of the plasma generated TiO\(_2\) particles are experimentally investigated via Ar/O\(_2\) thermal plasma oxidizing an aqueous solution of TTBO (stabilized with

![FIGURE 13](image13.png)  
**FIGURE 13** Temperature histories of the test particle with (a) transverse quench gas injection and (b) counterflow gas injection.

![FIGURE 14](image14.png)  
**FIGURE 14** FE-SEM micrographs showing morphologies of the TiO\(_2\) particles synthesized with different quench gas injection conditions. (a)–(d) have their denotations as in Figure 12.
DEA) sent into the center of the plasma through an atomization probe. Figure 14 shows typical FE-SEM morphologies of the TiO\(_2\) powders obtained under different synthetic conditions. It is clearly seen that an effective control of the particle size and size uniformity can be achieved by optimizing the type of quench gas and the mode of quench gas injection. The particles shown in (a)–(d) have average sizes of approximately 55 ± 22 nm, 53 ± 24 nm, 34 ± 14 nm, and 20 ± 6 nm, respectively. The experimental results thus comply well with the predictions of the numerical analysis (Figures 11 and 12). That is, counterflow injection is better than swirl flow, and He is more effective than Ar in the generation of particles of finer sizes and narrower size distributions.

3. SUMMARY

Thermal plasma provides an efficient reaction field for the one-step processing of ceramic nanoparticles of unique morphological and compositional features. Upon Ar/O\(_2\) plasma oxidation of large TiN particles, TiO\(_2\) are formed via two significantly different pathways: nanoparticles via direct condensation from the gas phase and microbeads via solidification of the molten droplets. Liquid precursor input through an atomization probe leads to an instantaneous combustion and a complete combustion of the raw material and thus significantly more uniform and finer particles. Another advantage of liquid precursor is that it allows an easy combination of different chemicals to achieve the doping of TiO\(_2\) for more colorful functionalities. Despite the type of precursor used, phase selection is largely influenced by the oxygen content of the species that leads to TiO\(_2\) formation and an oxygen-rich atmosphere favors anatase. For this, a higher oxygen input in the plasma promotes anatase crystallization while the introduction of reducing gases or doping TiO\(_2\) with subvalent cations yields more rutile. With these, almost phase-pure anatase and rutile crystallites can be selectively synthesized. Quench gas injection appreciably affects the plasma flow pattern and temperature field. It is shown that counterinjection of quench gas is better than transverse injection, and He is more effective than Ar in the production of finer particles of uniform sizes. Enhanced doping of TiO\(_2\) lattice with a variety of metallic and nonmetallic elements is achievable with the high central temperature and the superfast quenching at the tail flame region of the RF thermal plasma. Depending on the dopant, the resultant nanoparticles exhibit interesting photo-functionalities and magnetic properties.

REFERENCES


Chapter 11.1.7
Development of Easy-Handling Ceramic Nanoparticles: Present and Future
Noriya Izu, Toshio Itoh, Ichiro Matsubara and Woosuck Shin
National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Moriyama, Nagoya, Aichi 463-8560, Japan

Chapter Outline
1. Introduction 991
2. Experimental 992
3. Core–Shell-Type Ceria/Polymer Nanoparticles 992
4. Control of Particle Size 995
5. Change of Raw Materials 995
6. Formation Mechanism 996
7. Application of Nanoparticle 996
8. Core–Shell-Type Other Oxide 997
9. Future 998
10. Summary 998
References 998

1. INTRODUCTION
Dispersions (sols) of metal oxide nanoparticles are of interest in various scientific and technological fields. For example, sols of dielectric nanoparticles are used to make dielectric devices [1], sols of silica nanoparticles are used to make colloidal crystals [2,3], sols of MFe₂O₄ (M = Fe, Co, Mn) have been used in the production of magnetic nanodevices and biomedicines [4], and sols of cerium oxide nanoparticles are used for chemical mechanical polishing (CMP) [5-8]. Many sols of metal oxide nanoparticles are studied in various fields and applied to many applications. In industrial process, oxide nanoparticles are hardly used as powder but usually used as sol. Therefore, sols of oxide nanoparticles which have good dispersibility are needed.

In the normal process for making sols of nanoparticles, nanoparticle powder is prepared first, and then the powder is dispersed in a dispersion medium. In this case, however, the nanoparticles agglomerate strongly in the preparation of nanoparticles (Figure 1). Afterward, it is very difficult to disperse the agglomerated nanoparticles in a dispersion medium. The degree of agglomeration increases with decreasing size of nanoparticles. To prepare sols from dried nanoparticle powder, therefore, various steps are needed. This is the reason why sols of nanoparticle are very expensive. It is not rare that the place of a sol of nanoparticles is 10 times as much as that of a powder of the nanoparticles.

Since the two-step process which is a synthesis step of nanoparticles followed by a step to give dispersibility may be expensive for industrial production even if the sols are very good, we think that we need to carry out a synthesis step of nanoparticles and a step to give dispersibility at the same time; that is, we think that we should prepare nanoparticles that have good dispersibility in a one-step process. A similar concept is reported by some research groups [9-11].

We have used cerium oxide in order to test the above concept. Cerium oxide itself exhibits a range of useful functions, including high oxygen storage capacity [12-16], high diffusion coefficient for the oxide ion [17,18], high refractive index [19,20], and high ultraviolet (UV) absorption [5,21]. Therefore, it is used in a range of applications: as a promoter in three-way catalysts [12-16], in resistive oxygen gas sensors [22-26], in photonic crystals [27], and in oxide ion conductors [28-30]. A number of devices incorporating cerium oxide are now in practical use. Nanoparticles of cerium oxide have been the focus of substantial research, and it has been revealed in a number of studies that the nanoparticles are spherical, with a particle size of 50-500 nm and a narrow size distribution [31-35].

A sol of cerium oxide nanoparticles is used not only in CMP but also as a UV absorbent. There are some papers about sols of cerium oxide nanoparticles [5-8,21]. However, there appear to be no reports on the preparation of stable sols of such cerium oxide nanoparticles, which would be useful for the synthesis of high-performance
colloidal crystals (i.e., photonic crystals) in addition to the applications above. Colloidal crystal is expected to be a next-generation material for optical devices and many researchers study it recently\[36e38\]. Stable sols of such cerium oxide nanoparticles have potential for raw materials to make promoters for three-way catalysts, resistive-type oxygen sensors, and oxide ion conductors, and the process to make devices may change dramatically by using such stable sols. This is the reason why we selected cerium oxide to test the process concept mentioned above.

Although the addition of alkaline substances is the common method\[5,39\], a polyol method has been recently reported as a method for precipitating cerium oxide nanoparticles\[31\]. The polyol method is a well-known method to prepare fine particles of metal or metal oxide\[40,41\]. In this method, fine particles are obtained by heating a solution adding a precursor into a polyol, which has a high boiling point, at a temperature less than the boiling point under atmospheric pressure (1 atm). Fine particles are also obtained by glycothermal method in which glycol is used. The broader concept of glycothermal method is the solvothermal method. In this method, fine particles are obtained by heating a solution with a precursor at a temperature more than the boiling point of the solution under a pressure more than atmospheric pressure using a pressure vessel. In cases where the solvent is water, it is called hydrothermal method, while it is called glycothermal method in the case of glycol. In both cases, a pressure vessel is used in the process; that is, the reaction proceeds in high pressure. In the case of the polyol method, however, no pressure vessel is needed. Fine particles can be obtained only by heating a solution.

We noted the simplicity of the polyol method and prepared dispersions (sols) by polyol method to apply a process that prevents particle agglomeration simultaneously with the synthesis of particles. As a result of optimization of process condition, it is found that core—shell-type ceria/polymer nanoparticles are obtained with good dispersion of spherical particles of 50—120 nm in size with a narrow size distribution \[42\]. It is also found that the dispersion sol is very stable and easy to handle \[42\]. In this article, we give an outline of the unique easy-handling nanoparticles. Since the nanoparticles have easy-handling property and monodispersed spherical structure, we carried out ink-jet printing and fabrication of photonic crystals. We also give an outline of the results of these experiments in this report.

**2. EXPERIMENTAL**

The experimental details are described in Ref. \[42\]. Polyvinyl pyrrolidone (PVP) and cerium nitrate Ce(NO$_3$)$_3$·6H$_2$O were added to ethylene glycol and stirred. The concentrations of PVP and cerium nitrate in the solution were 120 kg/m$^3$ and 0.600 kmol/m$^3$, respectively. PVP additives with average molecular weights of 4350—18,000 were tested. Molecular weights were determined by gel permeation chromatography calibrated against polyethylene glycol. The mixture was heated under reflux to the boiling point of ethylene glycol for 15—20 min, yielding a white sol. A powder of the as-prepared sol was prepared by removal of unreacted material and excess PVP by centrifugation (10,000—18,000 rpm), followed by washing with water and ethanol and drying at 80 °C.

**3. CORE—SHELL-TYPE CERIA/POLYMER NANOPARTICLES**

**3.1. Characterization of Nanoparticles**

The X-ray diffraction (XRD) pattern of the powder of core—shell-type ceria/polymer nanoparticles \[42\] displays diffraction peaks identical to that given by card 34-0934 of the Joint Committee on Powder Diffraction Standards (JCPDS), confirming that the powder includes cerium oxide. The crystallite size determined from the full-width at half-maximum of the diffraction peaks on the Williamson—Hall plot is approximately 3 nm.

Thermogravimetry (TG) analysis \[42\] reveals a gradual weight loss of 5% with heating to 150 °C and a sudden weight loss of 15% with further heating to 300 °C. Heating the powder to 900 °C resulted in a total weight loss of 21%. These results indicate that the powder of core—shell-type ceria/polymer nanoparticles is composed of not only cerium oxide but also other materials that are decomposed by heating.

Scanning electron microscopy (SEM) image (Figure 2(a) \[42\]) of the nanoparticle powder reveals spherical particles with uniform size. The average particle size determined from the SEM images is 114 nm with
a coefficient of variation of 0.120 (standard deviation divided by average). In transmission electron microscopy (TEM) image of one of the 110 nm particles, the gray rim and black core of the particle were observed (Figure 2(b) [42]). From this image, the particles observed are aggregates of much smaller crystallite particles with a size of approximately 3 nm. High-resolution TEM images of the core of the core−shell-type ceria/polymer nanoparticles are shown in Figure 3 [43]. The core particle, which is a secondary particle, forms from nonporous cerium oxide primary particles of 2−4 nm in diameter. The image reveals that the core particle is not porous, but has a dense structure. A lattice image with a (200) plane distance of the fluorite structure is shown in Figure 3(b) [43], indicating that the primary particles have good crystallinity.

Fourier transform infrared (FTIR) spectrum [42] contains not only peaks from cerium oxide but also unknown peaks. Therefore, the unknown peaks are thought to be attributable to the shell. The FTIR spectrum for shell is similar to that of PVP but not the same. From $^{13}$C cross-polarization/magic angle sample spinning (CP/MAS) nuclear magnetic resonance (NMR) study [44], it is revealed that the shell of the core−shell-type ceria/polymer nanoparticles forms cross-linked structures (formation of cross-linked structure due to reaction with ethylene glycol and formation of cross-linked structures due to opening of pyrrolidone ring and reamidation) during the synthesis of core−shell-type ceria/polymer nanoparticles. The rigidness of the structures is thought to ensure the stability of the shell of the core−shell-type ceria/polymer nanoparticles.
The shell is not removed even if the wash treatment is carried out a few times. This strong adhesion between the shell and core also may be attributable to the cross-linked structures.

Pore-size distributions of the as-prepared and calcined (600 °C) nanospheres [43] were obtained from nitrogen adsorption isotherms of samples degassed at 373 K and measured at 77 K. In the pore-size distribution, there are two peaks at 40 and 100 nm for the as-prepared core—shell-type ceria/polymer nanoparticles, while there are three peaks at 3 nm or less, 30 nm, and 100 nm for the nanoparticles calcined at 600 °C. The specific surface area and total volume of the as-prepared nanospheres are 82.8 m²/g and 0.211 cm³/g, respectively. Calcination of the nanoparticles at 600 °C led to an increase in the specific surface area. The increased specific surface area is due to the appearance of nanometer-sized pores on the surface of core (Figure 3 (a)) by pyrolysis of the shell (organic polymer). The monodispersity of nanoparticles may give an ordered arrangement on the 100 nm scale. In the case of ordered arrangement of a closed packed structure, openings with a size of 40% diameter of nanoparticles form. The size of the nanoparticles was approximately 100 nm; therefore, the opening size is almost 40 nm. This is consistent with the peak of 40 nm in the pore-size distributions.

The core diameter and shell thickness are 92 nm and 11 nm, respectively, assuming that the weight loss of 15% until 200 °C is caused by combustion of the shell. These results suggest that the shell thickness is 10% of the diameter of the core—shell particles. According to Isra-elahvili et al., the secondary minimum of the DLVO theory is 5–7 nm [45]. The shell thickness is larger than the secondary minimum; therefore, the shell may cause steric hindrance that prevents the particles’ agglomeration. This is considered to be the reason for good dispersibility of the core—shell nanospheres, described in detail later.

The average density of the core—shell-type ceria/polymer nanoparticles was obtained using the two methods from the core diameter and shell thickness and from the density of the sols of the core—shell-type ceria/polymer nanoparticles. The density of the core—shell-type ceria/polymer nanoparticles is calculated to be 4.2 or 4.6 g/cm³ by the former method. From the latter method, a density of 4.7 ± 0.2 g/cm³ is obtained. The densities obtained by the two different methods are consistent, which indicates that the value of the density is reliable. In the former method, it is assumed that the densities of cerium oxide and PVP were 7.22 and 1.66 g/cm³, respectively, and the results are indirect evidence that this assumption is true. A core density that is the ideal density of cerium oxide is consistent with the dense core structure shown in the high-resolution TEM image.

3.2. Characterization of Sols

The size distribution of nanoparticles in the as-prepared dispersion sol, which was measured by dynamic light scattering (DLS), is shown in Figure 4 [42]. The average particle size is 110 nm, with a coefficient of variation of 0.065, consistent with the results obtained by SEM observations. This result demonstrates that the spherical particles are completely separated in ethylene glycol. This good dispersion may be due to the shell as mentioned above. This is because the PVP is lyophilic in dispersion media, and/or that the polymer coating acts as steric hindrance.

The DLS particle size of the sols prepared by redispersing the dried powder in various media is shown in Table 1 [42]. The results show that dispersion is very good even in the case of redispersion. The size of cerium oxide nanoparticles in the redispersed sol after 1 day is approximately equal to that after 8 days, indicating that aggregation does not occur even after 7–8 days. Figure 5 shows photographs of a redispersed sol of core—shell-type ceria/polymer nanoparticles in terpineol after 3, 18, and 39 days. No precipitation or transparent layer is observed in these samples, demonstrating that the redispersed sols are stable over extended periods. Redispersion in water or ethanol resulted in the appearance of a transparent layer on the surface of the sol after 7 days, indicating that sedimentation of the nanoparticles occurred. The transparent layer was not observed in the case of redispersion in terpineol or ethylene glycol. The transparent layer appears due to the ready sedimentation in water and ethanol, which have lower viscosity (ca. 1 mPa·s) compared to terpineol or ethylene glycol (10–40 mPa·s). The opaque layer in all cases was still soft, and uniform dispersion could be
achieved with small vibration. The ability for redispersion is attributed to the shell on the cerium oxide nanoparticles.

4. CONTROL OF PARTICLE SIZE

Synthesis experiments were carried out using a range of molecular weights of PVP (4350, 4790, 13,900, and 18,000) to determine the effect on particle size [42], where the molecular weight of PVP was determined by gel permeation chromatography calibrated against poly(ethylene glycol) (PEG). In the experiments, PVP weight was constant. This means that the number of monomers is the same in the experiments. The average particle size decreases with increasing molecular weight of PVP in the range of molecular weight from 4350 to 18,000 calibrated against PEG. The coefficient of variation for all samples is 15% or less, indicating monodispersal of particles in all cases. The particle size of the as-prepared dispersion sols, as determined by DLS measurements, is consistent with that of the powder determined from SEM images. Therefore, the size of nanoparticles appears to be controllable by appropriate adjustment of the molecular weight of the polymer added to prevent aggregation.

5. CHANGE OF ROW MATERIALS

When ammonium cerium (IV) nitrate (NH₄)₂Ce(NO₃)₆ is used instead of cerium nitrite in the synthesis process, the obtained nanoparticles are not CeO₂ but Ce(HCOO)₃[50]. The morphology of the particles is not spherical but rod like, and the sol of the nanoparticles shows poor dispersion.

When hydroxypropyl cellulose (HPC) is used instead of PVP in the synthesis process, the XRD pattern of the powder obtained is the same as that of the powder by using PVP. So, it is confirmed that the powder is cerium oxide [50]. The SEM image of the powder shows that the nanoparticles have spherical structure. The average particle size and the coefficient of variation of the nanoparticles are 90.1 nm and 0.223, respectively. The precipitate in the as-prepared sol is hardly observed after the as-prepared sol is left slightly for 14 days. So, it is confirmed that the dispersibility of the nanoparticles obtained by using HPC is stable for a long time. These results imply that the shell is formed on the surface of core which is cerium oxide. Therefore, it is concluded that the core—shell-type cerium oxide/polymer hybrid nanoparticles can be obtained even if HPC is used as a polymer.
6. FORMATION MECHANISM

The formation mechanism of nanoparticles is considered to be that primary particles of cerium oxide nucleate and aggregate to grow secondary particles. Because the space of the nucleation is several nanometers [51], PVP has to be excluded in order to increase the size of the secondary particles to tens of nanometers. Although a small force is needed for the exclusion of PVP of low molecular weight, a larger force is needed in the case of high molecular weight. In the case of high molecular weight, it is considered that the growth rate of the secondary particles is small due to the large force, and the growth stops early because of the formation of PVP layer around the secondary particles. This may be the reason why the particle size decreases with the increasing molecular weight of PVP.

7. APPLICATION OF NANOPARTICLE

7.1. Ink-jet Printing

7.1.1. Flow Properties of Dispersion Sol

First, the relationship between shear rate and shear viscosity of sol of the core—shell-type cerium/polymer nanoparticles dispersed in water is mentioned here. In this experiment, no dispersing agent is used. As can be seen in Figure 6, the viscosity is about 10 mPa·s even if the volume concentration is 17.4%, and the viscosity is almost independent of shear rate. Therefore, it is concluded that the dispersion sol has good flow properties.

Next, dispersion sols for ink-jet printing are explained. The dispersion sols for ink-jet printing were prepared by mixing the core—shell-type cerium/polymer nanoparticles into a mixing solution of water and ethylene glycol with a ratio of 1:1. The core—shell-type cerium/polymer nanoparticles in sols were synthesized by using cerium (III) nitrate hexahydrate and polyvinyl pyrrolidone as starting materials. In this experiment, no dispersing agent was used.

Shear viscosity and surface tension of the dispersion sol including the core—shell nanoparticles were measured. The viscosity increases slightly with increasing concentration of nanoparticles. The shear viscosity of the dispersion sol with a concentration of 4.3 vol% is 5 mPa·s and very small. The surface tension of the dispersion sol including the nanoparticles is independent of the concentration of the nanoparticles and approximately 55 mN/m. In general, the shear viscosity and surface tension of an ink-jet ink are in the ranges of several to 10 mPa·s and of 30–60 mN/m, respectively [15]. The sols synthesized in this experiment are in these ranges.

7.1.2. Ink-Jet Experiment

The ink-jet experiment using the sols as an ink was carried out. The results are summarized in Table 2 [50]. It becomes clear that all the sols synthesized without any dispersing agent can be used for ink-jet process. It is expected that the sol with a higher concentration of the nanoparticles would be used for ink-jet process, since the surface tension is independent of the concentration of the nanoparticles and the viscosity of the sol with 4.3 vol% nanoparticles is 5 mPa·s.

7.2. Feasibility Study for Photonic Crystal

Assemblies of nanoparticles were obtained by slow and rapid evaporation of ethanol from the sols by heating at
different temperatures. Figure 7 [42] shows an optical micrograph of the resultant assemblies obtained by slow evaporation. The formation of assemblies of silica has been reported to occur through the formation of homogeneous films followed by the generation and propagation of cracks and the formation of fibers [52]. The formation mechanism of the assembly prepared by slow evaporation appears to be similar to that of silica [52]. Cracks in the lower layers of the assembly could be observed through the upper layers in the optical micrographs, indicating that this assembly transmitted visible light. The ordered structure as shown in Figure 8 [42] was observed on the surface of the assembly, although not whole.

In the case of the assembly prepared by rapid evaporation, the reflectance is constant over a wavelength range of 250–350 nm, above which the reflectance increases with wavelength from 350 to 450 nm. The slow-evaporation assembly, on the other hand, exhibits a peak reflectance near 330 nm. The wavelength ($\lambda$) of light reflected by Bragg diffraction owing to the ordered arrangement of nanoparticles can be calculated by the following equation: [2,53]

$$\lambda = 2D(n_{\text{eff}}^2 - \sin^2 \theta)^{0.5}, \quad (1)$$

where

$$n_{\text{eff}}^2 = n_1^2V_1 + n_2^2V_2. \quad (2)$$

Here, $D$ is the spacing between planes, $n_{\text{eff}}$ is the average refractive index, $\theta$ is the angle between the incident ray and scattering planes, $n_i$ is the refractive index of component $i$, and $V_i$ is the volume of component $i$. As the average size of nanoparticles is 110 nm, $D$ is 110 nm. The value of $\theta$ of the UV-visible instrument is 67°. If the structure of the arrangement of nanoparticles is assumed to be face-centered cubic (fcc), $V_1 = 0.74$ and $V_2 = 0.26$. Using values of $n_1 = 2.1$[19, 20] and $n_2 = 1$, the calculated wavelength of reflected light is $\lambda = 359$ nm, which is only 10% larger than the experimental value. The difference between the calculated and experimental values may be due to the influence of the PVP coating on the nanoparticles, as the PVP layer cannot be included in these calculations. The reflectance peak at 330 nm is thus attributed to the ordered arrangement of nanoparticles, representing a colloidal crystal (i.e., photonic crystal). This result therefore indicates that the nanoparticles underwent ordered self-arrangement during evaporation.

8. CORE–SHELL-TYPE OTHER OXIDE

8.1. Zinc Oxide

PVP and Zn(CH$_3$COO)$_2$·2H$_2$O were added to diethylene glycol (DEG) and stirred. The concentrations of PVP and

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Results of Ink-jet Process Using the Dispersion Sols of the Core-shell Type Ceria/polymer Hybrid Nanoparticles. Reproduced by permission of The Ceramic Society of Japan [50]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume concentration (%)</td>
<td>Weight concentration (%)</td>
</tr>
<tr>
<td>0.82</td>
<td>3.5</td>
</tr>
<tr>
<td>1.6</td>
<td>6.9</td>
</tr>
<tr>
<td>3.2</td>
<td>13</td>
</tr>
<tr>
<td>4.3</td>
<td>17</td>
</tr>
</tbody>
</table>
Zn(CH$_3$COO)$_2$·2H$_2$O in the solution were 120 kg/m$^3$ and 0.10 kmol/m$^3$, respectively. PVP additives with average molecular weights of 10,000 reported by supplier were tested. The mixture was heated for 20 min at a heater temperature of 180 $^\circ$C, yielding a sol. A powder of the as-prepared sol was prepared by removal of unreacted material and excess PVP by centrifugation (10,000–18,000 rpm), followed by washing with water and ethanol and drying at 80 $^\circ$C.

The average particle size of as-prepared sol is about 282 nm determined by DLS. The average particle size calculated from SEM image is 336 nm. These sizes are very similar. The XRD patterns of the powders show wurtzite structure. So, it is found that the powder includes zinc oxide. From the calculation of FWHM of the diffraction peaks, the crystallite size is 14 nm. From FTIR and TG results, it is concluded that nanoparticle has PVP or PVP-related polymer on the surface.

8.2. Cobalt Oxide

PVP, Co(CH$_3$COO)$_2$·4H$_2$O, and water were added to diethylene glycol (DEG) and stirred. The concentrations of PVP and Co(CH$_3$COO)$_2$·4H$_2$O in the solution were 120 kg/m$^3$ and 0.10 kmol/m$^3$, respectively. PVP additives with average molecular weights of 10,000 reported by supplier were tested. The ratio of water to DEG was 1000:17. The mixture was heated for 360 min at a heater temperature of 200 $^\circ$C, yielding a sol. A powder of the as-prepared sol was prepared by removal of unreacted material and excess PVP by centrifugation (10,000–18,000 rpm), followed by washing with water and ethanol and drying at 80 $^\circ$C.

In SEM image of the as-prepared powder, spherical particles are observed. The average particle size calculated from SEM image is 81.1 nm and coefficient of variation is 0.166. These results mean that the nanoparticles are monodispersed. The XRD patterns of the powders showed NaCl structure. So, it is found that the powder includes cobalt oxide, although it includes a small amount of CoOOH. From FTIR and TG results, it is concluded that nanoparticle has cobalt oxide as a core and PVP or PVP-related polymer as a shell.

9. FUTURE

The core–shell-type nanoparticles mentioned above have a lot of potential. First, the size distribution of nanoparticles may decrease by the following method: The control of particle size by adjustment of molecular weight of PVP may be useful for improvement of monodispersity of the nanoparticles, because the molecular-weight distribution of PVP used in the study is not monodispersive but polydispersive. If the molecular-weight distribution becomes narrow, we would obtain more monodispersed cerium oxide nanoparticles.

Due to good flow properties of dispersions (sols) of the core–shell-type nanoparticles, it is expected that the ink-jet printing would be carried out using a sol with high concentration of nanoparticles. In the case of dilute ink, the ink-jet printing should be carried out several times to obtain sufficient amount. On the other hand, the sufficient amount would be obtained by one time ink-jet printing when high concentration ink is used.

High functional optical devices could be fabricated if nanoparticles with very narrow size distribution are obtained. This is because ceramics such as cerium oxide, titanium oxide, and zirconium oxide have large refractive index.

Since, furthermore, it was revealed that the two polymers and three oxides can be used for making shell and core, respectively, other polymers and oxides may be applicable for making shell and core. If doping secondary element can be applicable, the number of applications will expand dramatically.

10. SUMMARY

Many ceramic nanoparticles are very attractive due to many good functions but they are difficult to handle to fabricate parts or devices, since they are easy to agglomerate. Here we presented core–shell-type ceria/polymer nanoparticles and their various properties as a typical example of easy-handling ceramic nanoparticles. As a very important concept, we showed a process that prevents particle agglomeration simultaneously with the synthesis of particles. It is expected that this concept will be widely known and various processes to make easy-handling ceramic nanoparticles will be invented. It is also expected that various kinds of easy-handling ceramic nanoparticles will be developed and used in the industrial world.

REFERENCES


Sonoprocess of Ceramic Materials

Naoya Enomoto
Department of Applied Chemistry, Faculty of Engineering, Kyushu University, 744 Motooka, Nishi, Fukuoka 819-0395, Japan

Chapter Outline
1. Introduction 1001
2. Apparatus for Sonoprocess – How to Create Sonochemical Field 1002
3. Sonochemical Effects on Ceramic Process in Practice 1003
Acknowledgments 1009
References 1009

1. INTRODUCTION

1.1. Definition and Brief History of Ultrasound

Ultrasound is defined as a sound which is not for listening. Since a human ear can hardly detect a sound higher than that, a sound of frequency above 20 kHz is regarded as an ultrasound in general. As will be explained later, there is a particular phenomenon called ultrasonic cavitation when an intensive sound wave is applied to a liquid phase.

In a historical sense, ultrasonic technology has been developed since the 1910s as a naval tool for navigation and communication. Application of power ultrasound to chemical reactions and/or chemical process began in 1927 [1-3]. Using a natural quartz transducer, they showed a variety of amazing phenomena such as heating, emulsification, atomization, agglomeration, dispersion, crystallization, acceleration of chemical reactions, cell disruption, and so forth. Theoretical support on the dynamics of cavitation was given in 1950 [4]. An epoch of ultrasonic transducers was made by the discovery of giant piezoelectricity at the morphotropic phase boundary of a PZT ceramic. Thereafter, people are able to use strong and stable ultrasound. Sonochemistry has been applied to and developed in a variety of science and technology [5-7].

1.2. Ultrasonic Cavitation – Origin of Sonochemistry

When a sufficiently strong wave of compression and rarefaction (i.e., strong sound) propagates a liquid medium, rarefied part of the liquid can be torn apart to form bubbles or cavities. The bubbles expand and shrink (in other words, oscillate in size) in the sound field at the frequency of the sound. In the stage of shrinkage, the bubbles are subjected to adiabatic compression to generate incredibly high temperature of ~5000 K and high pressure of ~100 MPa, depending on several factors of bubble dynamics. Ultrasonic cavitation implies the formation, the growth, and the collapse of the bubbles in the ultrasonic field. Figure 1 shows a schematic of the occurrence of cavitation.

Inside the bubbles, there mainly are two kinds of matter; one is the dissolved gas in the liquid and the other is a vapor of the liquid phase. The former is regarded as the crucial factor to determine the maximum temperature/pressure during the collapse of cavitation:

\[
T_{\text{max}} = T_0 \left( \frac{P_m (\gamma - 1)}{P} \right)^{\gamma / (\gamma - 1)}
\]

\[
P_{\text{max}} = P \left( \frac{P_m (\gamma - 1)}{P} \right)^{\gamma / (\gamma - 1)}
\]

where \(T_0\) is the temperature of the bulk solution, \(T_{\text{max}}\) is the temperature in the bubble at the moment of transient collapse, \(\gamma\) is the \(C_p / C_v\) of the dissolved gas, \(C_p\) is the specific heat capacity at constant pressure, \(C_v\) is the specific heat capacity at constant volume, \(P\) is the pressure in the bubble at its maximum size, and \(P_{\text{max}}\) is the pressure in the bubble at the moment of transient collapse. Since \(T_{\text{max}}\) is proportional to \(\gamma\), noble (monatomic) gases such as He, Ne, Ar, Kr, and Xe are preferred to proceed an efficient sonochemistry [8]. Considering the availability and cost, in...
general, argon is the most popular gas that should be substituted in the solution prior to sonochemical reactions. Vapor pressure of the solution is also a very important factor in sonochemistry. Vaporous materials can enter inside of the bubbles and then decompose very thermally. It should be noticed here that this thermal treatment is absolutely momentary (~μ sec). In a sense of material processing, such a rapid heating is accompanied by rapid cooling and would be useful in quenching of nonequilibrium phases.

The most popular solvent is water. Pure water can be decomposed to radicals in the collapsed bubbles as follows:

\[ \text{H}_2\text{O} \rightarrow \text{H}^\cdot + \cdot\text{OH} \]

In the presence of dissolved gas or ion (e.g., N\textsubscript{2} or Cl\textsuperscript{−}), a small amount of acid (e.g., H\textsubscript{3}NO\textsubscript{2} or HCl) can easily be formed by ultrasonic cavitation [9]. It is known that the active chemical species generated by the sonolysis of water often plays a dominant role in sonochemical processes.

Design of sonochemical reaction in the bubbles is a complicated issue. For example, let us consider the case when we increase temperature for a certain system of sonochemical process. It is a matter of course that chemical reactions are generally enhanced with increasing temperature. In a sonochemical sense, the increase in vapor pressure due to the temperature rise allows more reactant to enter the bubbles and thus more effects would be expected. However, the results could be often the opposite because the following factors can be accompanied with the temperature rise: (i) amount of dissolved gas that is to be heated in the bubbles decreases as the solubility of gas in liquid decreases, (ii) too much vapor of liquid may dampen the collapse of cavitation, and (iii) vaporization itself is a process to decrease the temperature. We should notice that it all depends on the system investigated.

### 2. APPARATUS FOR SONOPROCESS—HOW TO CREATE SONOCHEMICAL FIELD

#### 2.1. Two Types of Sonicator

There are two types of ultrasonication in general. One is called cleaner type and the other is called horn type. The former is very widely used in laboratories to clean glassware or to facilitate dissolution and dispersion of solid particles in a liquid medium. The latter was originally known as a cell disruptor. Figure 2 shows schematic illustration of the two sonicating ways. The horn-type sonicator, usually driven at 20 kHz, is good at making large amplitude of vibration and thus intensive wave energy. Additional stirring is not necessary because a strong straight flow is available in the solution. It is necessary to be very careful for (i) cooling the reactant for avoiding significant increase of the temperature due to the intensive sonication and (ii) unfavorable contamination from horn tip (mostly made of titanium alloy) when you use the horn-type sonication. The latter happens exponentially when the horn tip is eroded to increase its surface area. If the horn-tip surface becomes matte, polishing is required.

As the reactant is separately irradiated by ultrasonication, on the other hand, the experimental advantages of cleaner-type sonication are being contamination-free and convenient in controlling the atmosphere (dissolved gas) by closed system. The frequency for the clear type can be selected by selecting different transducer (40–200 kHz in general). Owing to the indirect sonication at higher frequency (= smaller amplitude), the cleaner-type sonicator is not very good at such mechanical treatments as dispersion and dissolution when compared with the horn type. However, it can effectively produce radical species when the height of solution and the position of the reaction vessel against the transducer are set correctly to form a standing wave.
2.2. Determination of Ultrasonic Power

It is not always easy to determine such a fundamental parameter as an ultrasonic power, or that can rather be a controversial issue that could be a cause of lack in experimental reproducibility. In the horn-type sonication, we can assume the averaged power applied from a horn tip into the reactant in a vessel since the solution can be well circulated due to a straight flow, even if the sonic field is quite inhomogeneous from the beneath-tip to the away-tip regions. In the cleaner-type sonication, on the other hand, it must be carefully noticed that the position of vessel or the height of liquid has a crucial influence on the sonochemical effects. There are three ways to express the power of ultrasound. First, an electric power applied to an ultrasonic transducer can be the easiest one. However, it may not always correspond to the power injected in the reactant, and, also, it may not be suitable for a global comparison. The second one is called calorimetry, where linear temperature rise given by ultrasonication for a short time is converted by the following equation, assuming that a whole power of ultrasonication is adiabatically dissipated in a liquid medium:

\[ P = \left( \frac{dT}{dt} \right) \times C_p \times M \]

where \( P \) (J/s) is the calorimetric ultrasound power, \( dT \) (K) is the increase of temperature, \( dt \) (s) is the time of sonication (favorably 10–20 sec), and \( C_p \) (J/g·K) and \( M \) (g) are the heat capacity and the amount of the liquid, respectively. Typically, a certain amount of water in a heat-insulated container is irradiated by a constant ultrasound, and the temperature rise is simply measured by a thermocouple or a thermistor. (A thermometer using the thermal expansion of liquid mercury or organic matter should NOT be used in this measurement since the ultrasonication may damage the liquids.) The third way of the ultrasonic power measurement is rather chemical and is considered as the standard evaluation of sonochemical ability[10]. An aqueous solution of 0.1 mol/L KI (potassium iodide) is sonicated for a certain time. The solution changes from clear to slightly yellow-brown due to the following equations:

\[ 2I^- \rightarrow I_2 \]
\[ I_2 + I^- \rightarrow I_3^- \]

By measuring the specific absorption of \( I_3^- \) at 352 nm with an ordinary spectrophotometer, sonochemical efficiency can be calculated (with calorimetric power) for the global standard of sonochemistry.

3. SONOCHEMICAL EFFECTS ON CERAMIC PROCESS IN PRACTICE

Very roughly speaking, every ceramic processing involving a liquid phase can adopt an ultrasonic treatment. In other words, ceramic processes in solid state and gas state (e.g., CVD) are rarely subjected to ultrasound because cavitation never happens in solid or gas. This chapter reviews various factors/elements in solution processing under ultrasonication.

3.1. Dissolution and Precipitation

It is NOT usual in practice that thermodynamic-stable state can be achieved immediately at the first stage of the process and remains forever. In most cases, the initial precipitates would dissolve into the mother solution and then reprecipitate to be in a more stable state. For example, Ostwald ripening occurs due to the difference in size and solubility. Without a single doubt, ultrasonic irradiation is a powerful tool to enhance dissolution because effective dispersion of agglomerated substances absolutely increases the surface area as well as the mass transfer in solution is greatly accelerated by micro- and straight flow. It is definite from the author’s experiences on sonoprocess that a process involving a dissolution step can be significantly improved (or influenced) by ultrasound.
Figure 3 shows the ripening behavior of aluminogel under ultrasonic irradiation [11]. In this system, amorphous aluminogel precipitated by mixing AlCl$_3$ aqueous solution and ammonia water transformed to crystalline bayerite in mother solution. Compared to a control run (M.S. denotes mechanical stirring), it is obvious that the treatment with an ultrasonic homogenizer (denoted as U.S.H.) fairly promoted the bayerite formation (appearance and final amount) at moderate intensity (1.9 W/cm$^2$), but suppressed it at high intensity (25.4 W/cm$^2$). This ripening process occurs via the dissolution of aluminogel and subsequent precipitation of the crystalline phase. Even at the lower intensity, ultrasound was found to be effective in dissolving the gel. At higher intensity, however, it was found that a pseudo-boehmite phase with lower crystallinity (i.e., poorly detected by X-ray diffractometry) was precipitated. This may be called sonochemical switching [12] from bayerite to pseudo-boehmite according to ultrasonic intensity. We may remember in general that a less-defined structure/phase is likely to form with an intensive sonication.

### 3.2. Nucleation and Growth

Nucleation and growth in ceramic processing are fairly complicated in general. In order to look at the effect of ultrasound on nucleation and growth without chemical changes, we examined crystallization of alum (ammonium aluminum sulfate hydrate) from a supersaturated solution by cooling [13]. After passing through a similar cooling history, the size of precipitated crystals was thoroughly influenced by agitation with or without ultrasound as shown in Figure 4. Huge and large crystals were often observed with mechanical stirring, while the size distribution became quite uniform with ultrasonic agitation and median size could be controlled by the ultrasonic condition. The median size decreases with increasing intensity (data not shown in Figure 4). It is considered that cavitation bubble can be a nucleation site and that the number of cavitation bubbles increases with increasing intensity of ultrasound. It is considered, moreover, ultrasonic cavitation is likely to occur at lower frequency. The results shown in Figure 4 suggest that the growth was accelerated at higher frequency.

### 3.3. Oxidation

As mentioned in the explanation on the origin of sonochemistry, a most typical solvent of water can be sonolized to OH radical, which is an oxidative. Figure 5 shows the oxidation of Fe$^{2+}$ ion at different atmosphere [14]. It is strange at a glance but reasonable as a sonochemical result that oxidation is greatly enhanced in the absence of oxygen. This is of course true as the cavitation temperature for argon is higher than that of air (N$_2$ + O$_2$) to generate the oxidative species. It is also found in Figure 4 that oxidation is promoted in air. This can cause another type of sonochemical switching [15] from a product with lower oxidation number (e.g., Fe$_3$O$_4$ and magnetite) to that with higher oxidation number (e.g., α-FeOOH and goethite) provided the conditions other than sonication are all the same. Dang et al. [16,17] successfully prepared monodispersed and stable magnetite nanoparticles by using a mixed solvent of water and ethanol with ultrasonic irradiation.

### 3.4. Reduction

When ultrasound is applied to the aqueous solution containing noble metal ions such as Au$^{3+}$ or Pt$^{4+}$, they can be reduced to form metal nanoparticles with narrow size distribution [18,19]. Reduction rate and thus particle size can be controlled by the presence of organic substances such as SDS (sodium dodecyl sulfate) or PEG (polyethylene glycol). The organics work both as a stabilizer to coat the nanoparticle surface and as a source of reductive radical after thermally decomposed by cavitation. By further control of the reduction, it is reported that a variety of nanostructures as core/shell, random alloy, and mere mixture can be fabricated in the Au-Pd system.

Another simple system is the sonochemical reduction of solid oxide [20]. By sonicating Ag$_2$O particles of several μm in ethanol, the oxide can be easily reduced to form silver nanoparticles. Since it can be executed at room temperature without any harmful reagents, this technology is expected for the anti-virus coating onto the substrates of glass, plastic, rubber, and so forth.

### 3.5. Polymerization

It has been known as sonochemical darkening for benzene derivatives such as nitrobenzene to turn yellowish when
irradiated with a power ultrasound. This phenomenon is due to the extension of the π conjugate or the combination among benzene rings. Katoh et al. reported that a trace amount of C60 or a fair amount of CNT (carbon nanotube) can be obtained by ultrasonating benzene or benzene derivatives with a homogenizer in the presence of various solid particles in an ice bath to prevent excessive heating. Jeong et al. also reported that SWCNT (single-wall CNT) was obtained from p-xylene in the presence of ferrocene catalyst and silica particles. Knowing that those nanocarbon materials are nonequililibrium, a rapid heating of the organic sources inside the bubbles and subsequent quenching would be the mechanism in such cases.

3.6. Sol–Gel (Alkoxide) Process

In the years around 1970, ceramists found that metal alkoxides are promising raw materials for fabricating advanced ceramics. Now in 2010, sol–gel process is one of the main topics in the ceramic society in the world.

Alkoxides are hydrophobic and immiscible with water which is the reactant for the their hydrolysis. Usually, co-solvent such as an alcohol is used to dissolve both alkoxide and water. Incidentally, one particular feature of ultrasound is emulsification to mix oil and water. A gel prepared by a compulsory mixing of alkoxide and water...
without using co-solvent is called sonogel [26]. It is reported that sonogels have less amount of organic residue and high density than a classical gel prepared with co-solvent.

Of course, hydrolysis of alkoxide in the presence of co-solvent is accelerated by sonication. Tomomasa et al. [27] applied high-frequency ultrasound to the preparation of mullite precursor from silicon and aluminum alkoxides. Since the hydrolysis rate of these two alkoxides are quite different, normal hydrolysis causes inhomogeneity. With the application of ultrasound, slow hydrolysis of silicon alkoxide was significantly promoted to reduce the difference, resulting in a homogeneous mullite precursor which directly crystallized to mullite without passing an intermediate spinel phase.

Monodispersed silica sphere (or Stöber silica ball) is an interesting topic for sonoprocess. Figure 6 shows the effect of ultrasonication during the synthesis of monodispersed silica [28]. Silica balls were observed to form relatively uniform and dense aggregates as if they were sintered. Considering that such agglomeration occurred through the collision among the particles and the dehydration among Si–OH groups in the sphere surface, this phenomenon may be called sonochemical polymerization that was mentioned in the previous section.

In a quite different point of view, there is a brand-new (or somewhat weird) concept in the sonoprocess named soft sonication [29,30]. It is based on the idea that a transparent solution may not be homogeneous at the molecular level and the solution structure (or clustering state) could have a time dependence through days or year. In soft sonication, neither cavitation and nor intensive vibration is required but small vibration amplitude and high-frequency mixing of molecules and clusters in solution have been noticed. In practice, very weak ultrasonic vibration at a mW level was applied to the starting solution prior to the reaction. In the case of silica ball synthesis, we need two bottles of starting solution. One is an ethanolic solution of TEOS (tetraethylorthosilicate) and the other is an ethanolic solution of ammonia water. These two solutions can be immediately prepared because both the solutes (TEOS and water) are quite soluble with ethanol. Commonly, these two are mixed to proceed with the reaction soon after the preparation. In this concept, however, these two solutions...
were aged with or without soft sonication for days, and then mixed. Figure 7 shows the experimental results; that is, the precipitation (occurrence of turbidity) was delayed, monodispersity was enhanced, and the sphere size increased to 120\textsuperscript{e}130\%. The increment of sphere size is in fact very large if we convert it to volume or mass; i.e., \(v = \frac{1}{1.3} \times 0.5\), this value implies that the nucleation using the aged starting solutions became half as much as the fresh ones, as the reactant here was thoroughly consumed. These results are quite reproducible, essential, and free from extra factors such as contamination from container or moisture in ethanol \([29]\). Even in the absence of soft sonication, the aging was observed. Soft sonication can be a novel tool to enhance the aging process or the modification of the solution structure \([31]\).

3.7. Sonoelectrochemical Process
Electrochemistry can also be stimulated by ultrasonication \([32]\). Figure 8 shows the results of anodic oxidation of titanium metal in nitric acid \([33]\). In the MS (mechanical stirring) run at [HNO\(_3\)] = 0.25 M, amorphous titania having a nanostructure of straight mesopores was formed. Ultrasonication decreased the pore size and increased the pore number density, which suggests that the initial pitting at the early stage of anodization is enhanced by sonoprocess. At higher concentration of 1 M, ultrasonic agitation was found useful to maintain the mesoporous structure.

3.8. Solidification of Molten Salt
It is known as ultrasonic metallurgy that the microstructure of a casted metal ingot is drastically improved by an ultrasonic vibration during solidification. For instance, reduction of grain size, removal of dissolved gas in a melt, and suppression of compositional segregation are reported since the dawn of ultrasonics to the present \([34,35]\). As the properties of polycrystalline materials are absolutely dependent on their microstructure, it is widely recognized that ultrasonication fairly improves the properties of casted metals. Polycrystalline ceramic materials are rarely prepared through a melting and solidification process, because there are many problems in adopting the melting process into ceramic polycrystals — undesired large grain growth, high melting point and high reactivity of inorganic melts, formation of microcracks or voids, and so forth. It is the last topic in this article to ask the following question — Can ultrasonic casting of molten salt solve those problems?

As a model inorganic, the NaNO\(_3\)–Ba(NO\(_3\))\(_2\) system \([36]\) was investigated with a horn-type sonicator equipped with furnaces as shown in Figure 9. It is not difficult to execute sonication at high temperature if only the horn is properly cooled, the horn tip and the vessel are durable enough, and suitable temperature gradient allows the solidification to occur at the bottom of the vessel. This binary system has an eutectic point at Ba(NO\(_3\))\(_2\) = 18 mass\% and \(T = 290 \degree C\). As it resembles the Al–Si binary of great importance in industry, the nitrate binary is often investigated. Figure 10 shows the casted structure at hypoeutectic at Ba(NO\(_3\))\(_2\) = 8 mass\%. In this compositional image, black and white areas correspond to NaNO\(_3\) and Ba(NO\(_3\))\(_2\), respectively. It is obvious that the large dendrite of NaNO\(_3\) was removed by the ultrasonic vibration.
FIGURE 8  Mesoporous structure of titania prepared by anodic oxidation process [33]. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 9  Schematic of ultrasonic casting equipment [36]. For color version of this figure, the reader is referred to the online version of this book.
ACKNOWLEDGMENTS

The author acknowledges Prof. Junichi Hojo, Prof. Zenbe-e Nakagawa, and many co-workers on the research in the sonoprocess of ceramic materials.

REFERENCES


FIGURE 10  Cast structure at 92%NaNO₃—8%Ba(NO₃)₂ [36]. For color version of this figure, the reader is referred to the online version of this book.


Chapter 11.1.9

Organic—Inorganic Hybrid Materials Prepared Through Supramolecular Assembly

Kiyofumi Katagiri and Kunihito Koumoto

Chapter Outline

1. Introduction 1011
2. Cerasome as a Liposomal Organic—Inorganic Nanohybrid Hollow Spheres 1011
3. Nanohybrid Coatings via Layer-by-Layer Assembly of Water-Soluble Precursor 1015
4. Organic—Inorganic Hybrid Hollow Capsules Prepared by Colloid Templating 1016
5. Summary 1021
Acknowledgments 1021
References 1021

1. INTRODUCTION

Researches on sol—gel science and technology have received both fundamental and practical interests over the past several decades. Especially, preparation of organic—
inorganic hybrid materials is one of the most attractive fields of sol—gel chemistry to this day [1]. Various types of material combinations and synthetic strategies have been developed for the hybrid materials up to the present time: (1) the encapsulation of functional organic molecules within the sol—gel-derived inorganic matrices; (2) the polymerization of functional organoalkoxysilanes; and (3) the organic functionalization of inorganic matters, such as nanofillers and nanoclays. However, the interface between the organic and inorganic components in these hybrids seems to be structurally ambiguous and difficult to control precisely. Therefore, the development of materials with organized nanostructures is a fascinating research subject in the field of hybrid materials. The technique based on supramolecular chemistry, such as the self-assembly technique, is widely used as a powerful and versatile approach to achieve the nanomaterials with well-organized structure. Recently, this approach is adopted for the development of nanocomposites of organic—
inorganic hybrid materials.

In this chapter, several approaches for the preparation of organic—
inorganic nanohybrids by supramolecular organization are reviewed. Firstly, the development of liposomal nanohybrid materials using amphiphilic organoalkoxysilanes is described. This hybrid has a unique structure since the inner structure is a lipid-bilayer vesicle, whereas the outer structure is analogous to silica particle. The hierarchical organization of these vesicular hybrids on the substrate is also introduced. Next, the nanohybrid coating by layer-by-layer assembly technique is described. The crystallization of titania was achieved at a temperature less than 100 °C in this system. Preparation of hybrid hollow capsules by layer-by-layer assembly and sol—gel method using colloid templating is also described in this chapter.

2. CERASOME AS A LIPOSOMAL ORGANIC—INORGANIC NANOHYBRID HOLLOW SPHERES

2.1. Formation of Cerasomes via Sol—Gel Reaction and Self-Assembly of Amphiphilic Organoalkoxysilane

The molecular self-assembly is a nanofabrication strategy involving designing supramolecular entities so that shape complementarity causes them to aggregate into desired structures [2,3]. Organic—
inorganic hybrid materials also
have attracted much attention from material scientists and chemists in recent years, due to their great potential for use in a wide variety of applications through fusion of individual organic and inorganic properties [1,4–7]. The sol–gel method is one of the most powerful techniques to prepare hybrid materials and has provided moderate preparative conditions for construction of inorganic oxide frameworks that are derived from hybrid materials [8–10]. Especially, preparation of novel materials with organized nanostructures is a fascinating research subject in the field of hybrid materials [11–18]. The combination of molecular self-assembly and the sol–gel chemistry successfully provides us the novel organic–inorganic nanohybrid materials. Molecular assemblies such as the templates (e.g., rod-like micelles [19–22], block-copolymers [23–25], microemulsions [26,27], organogels [28,29], and cast films of bilayer membranes [30]) were employed for the preparation of nanohybrid materials via these approaches. However, most of these materials are finally composed of inorganic components alone and the organic parts are simply employed as templates. In contrast, a novel class of layered organic–inorganic nanocomposites, composed of compounds with covalent bonds between the silicates and the organic molecules, has been developed in recent years [31–37]. These materials would offer much potential, since the hybrid precursors can form three-dimensional networks during the self-assembly process whereby inorganic layers and organic moieties are linked by stable covalent bonds.

A novel type of the organic–inorganic nanohybrids was developed through a combination of sol–gel reaction and self-assembly of amphiphilic organoalkoxysilanes [38–41]. This organic–inorganic hybrid material is composed of a spherical lipid-bilayer membrane having an internal aqueous compartment, like a liposome [42–44], and covered with a silicate framework on its surface (Figure 1). The abbreviated name “cerasome” was proposed for this vesicular assembly by merging “liposomes” with “ceramics” [38]. A proamphiphilic organoalkoxysilane (1) with a triethoxysilyl head moiety and a hydrophobic double-chain segment were designed for the preparation of cerasome. Upon applying the sol–gel reaction to the compound 1, the resulting amphiphiles with a silanol head aggregated to form morphologically stable bilayer vesicles. Another cerasome-forming organoalkoxysilane (2) having a quaternary ammonium group at the head moiety in addition to the triethoxysilyl part was also developed. Different from the compound 1, the compound 2 acts as an amphiphilic molecule; even its silanol group is capped by ethoxy group.

As for the lipidic organoalkoxysilane 1, hydrolysis of the triethoxysilyl head group converts the proamphiphile into the corresponding amphiphilic molecule to form a liposome-like bilayer membrane. In addition, condensation among the silanol groups on the relatively hydrophobic vesicular surface would proceed to develop a siloxane network. However, if the hydrolysis and followed condensation occur much quickly prior to the self-assembling,
formation of the bilayer structure may be disturbed. Therefore, control of the reaction rate in the sol–gel process, especially the hydrolysis rate, seems to be important for the cerasome formation.

The hydrolysis behavior of the organoalkoxysilanes 1 and 2 was monitored by $^1$H–NMR spectroscopy. $^1$H–NMR spectra of the aqueous dispersion of 1 at pD 7 showed a broad quartet signal at 3.75 ppm corresponding to the methylene protons in the ethoxy group of the lipid ((CH$_3$CH$_2$O)$_3$Si–). No signal assigned to hydrolyzed product was detected even after 12 h incubation at 25 °C. On the other hand, at pD 3 the broad signal for the methylene protons in the ethoxy group of the lipid gradually decreased to be replaced by a sharp quartet signal at 3.65 ppm, which is assigned to the methylene protons of ethanol as a byproduct of the hydrolysis reaction. Time course for the hydrolysis of 1, as evaluated from the peak area changes in the methylene proton signal of the produced ethanol, is shown in Figure 2. While the hydrolysis hardly proceeded at pD 7 (Figure 2a), slow hydrolysis was observed at pD 3 in the initial stage of the reaction followed by a gradual rate acceleration showing a sigmoidal curve to complete after 8 h (Figure 2b). Such behavior is owing to the change in the reaction system from a heterogeneous to a homogeneous state with time, since the organoalkoxysilane 1 has less water solubility and is initially present as oil droplets. On the other hand, hydrolysis of the organoalkoxysilane 2 proceeded rapidly to complete within 2 h under the similar pD conditions (Figure 2c). The results imply that the amphiphilic organoalkoxysilane 2 spontaneously forms bilayer vesicle prior to hydrolysis and that the hydrolysis of the triethoxysilyl head efficiently occurs on the vesicular surface.

The aggregate structures of the cerasomes were observed by transmission electron microscopy (TEM). In the slightly turbid aqueous dispersion of 1 with vortex mixing, the formation of the multilamellar vesicles (MLVs), with a lipid-bilayer thickness of ca. 4 nm and a vesicular diameter of 200 nm, was clearly confirmed by the TEM images (Figure 3a). Ultrasonication of the aqueous dispersion gave a clear solution with low surface tension. The electron micrographs showed the presence of cerasome with a diameter range from 150 to 300 nm, whose sizes corresponded to the hydrodynamic diameter (214 nm) as evaluated by dynamic light scattering (DLS) measurements (Figure 3b). Unlike conventional liposomes, ultrasonication did not alter the structure of the cerasome 1 from MLVs to single unilamellar vesicles (SUVs). The TEM images of the cerasome prepared from the lipid 2 are also shown in Figure 3. The vesicles in a diameter range from 50 to 100 nm were observed for the cerasome 2 prepared by vortex mixing (Figure 3c). An internal view of the MLVs with a bilayer thickness of ca. 5 nm was clearly confirmed. Upon sonication of the dispersion sample with a probe-type sonicator for 10 min at 30 W, smaller particles with a diameter of 20–40 nm, suggesting the SUVs of the cerasome, were observed (Figure 3d). The vesicular size was well consistent with the hydrodynamic diameter (30 nm) as evaluated by DLS.

2.2. Hierarchical Assembly of Cerasomes on Substrates via Layer-by-Layer Assembly

Liposomes have been extensively employed as biomembrane models in the field of the supramolecular chemistry. Up to the present time, it became apparent that the bilayer vesicles are effectively utilized as nanocapsules for the drug delivery or gene transfection systems [45], artificial cell membranes [46], and artificial enzymes [47]. From a point of view in the recently growing nanotechnology, however, the bilayer vesicle seems to have another potential as a fascinating candidate of the nanomaterials for designing functional supramolecular devices [48]. For such purpose it should be important to develop a new methodology to form hierarchically integrated assemblies of liposomes, since the multicellular bodies in biological system can create highly organized architectures and exhibit much more functions as compared to the unicellular ones.

Here, the three-dimensional integration of lipid vesicular nanoparticles on a substrate can be successfully achieved by the layer-by-layer assembly technique [49,50]. As the lipid-bilayer vesicles for the integration, a couple of cationic and anionic cerasomes, prepared from organoalkoxysilane, 1 and 2 were chosen [38,41]. These cerasomes perform as polyamionic and polycationic nanoparticles in neutral pH region, since the isoelectric points determined by pH-dependence of the zeta-potential were 4.3 and 12.0.

Three-dimensional integration of the cerasomes was done by employing the layer-by-layer assembly technique.
and monitored by measuring the adsorption mass changes on a 9 MHz quartz crystal microbalance (QCM) [51]. The QCM resonator was immersed in an aqueous solution of the anionic and cationic cerasome with intermediate water washing and drying under a nitrogen stream. This process was periodically interrupted to measure the QCM resonance frequency. In Figure 4 the odd- and even-number steps were indicated for the anionic and cationic cerasome adsorption, respectively. For comparison, the mass change for alternate layer-by-layer assembling of the traditional synthetic liposomes monitored by QCM was also shown in Figure 4. In this case, dihexadecyl phosphate (DHP) and dimethyl dioctadecylammonium chloride were employed to prepare anionic and cationic liposomes, respectively. The average values of the mass change for the cerasome adsorption were much larger than those of conventional liposomes, reflecting formation of the layered assembly of each cerasome. In addition, these values were larger for the anionic and smaller for the cationic cerasome layer, reflecting the size of cerasome in the aqueous solution.

In order to clarify the integrated structure of the cerasomes, the similar alternate layer-by-layer assembly of the anionic and cationic cerasomes was formed on a substrate, and the surface structure was examined by an atomic force microscopy (AFM) in tapping mode. AFM images after the first and second assembling step in Figure 4 were shown in Figures 5a and 5b, respectively. The cerasome particles closely packed like a stone pavement were clearly observed in both layers. In addition, difference in the particle size for each layer indicates that the cationic and anionic cerasomes undoubtedly formed the layer-by-layer assembly. The layered paving of the vesicular nanoparticles was observed in every layer at least up to tenth adsorption steps. It is noteworthy that the layered paving superstructure was never observed by replacement of the cerasome to other bilayer vesicles formed with phospholipids or synthetic lipids, e.g., DHP. Therefore, the morphological stability of

![FIGURE 3](image1.png)

FIGURE 3 TEM images of the cerasome of 1 (a, b) and 2 (c, d) prepared by vortex mixing (a, c) and followed by ultrasonication (b, d).

![FIGURE 4](image2.png)

FIGURE 4 Adsorbed mass change of QCM upon layer-by-layer assembly of the cerasomes and conventional synthetic liposomes: filled circle, the anionic cerasome; open circle, the cationic cerasome; filled triangle, the anionic liposome; open triangle, the cationic liposome.
cerasomes seems to be extremely higher than those of other bilayer vesicles, since the membrane surface of the cerasome is covered with the siloxane network to prevent collapse and fusion of the vesicle.

Here, a three-dimensionally-packed vesicular assembly was successfully prepared by using an organic–inorganic hybrid vesicle, the cerasome. The present nanoparticle assembly keeping a bilayer membrane structure and an inner aqueous compartment in each unit has a potential to construct artificial multicellular systems as supramolecular nanodevices. The author believes that the cerasome can open a new field in constructing organic–inorganic nano-hybrid materials.

3. NANOHYBRID COATINGS VIA LAYER-BY-LAYER ASSEMBLY OF WATER-SOLUBLE PRECURSOR

There has been increasing interests in the production of coating films comprising TiO₂ since they present useful properties such as high refractive index, excellent transmittance in visible range, high relative dielectric constant, remarkable solar energy conversion, and photocatalysis [52]. The flexible strategy that affords nanometer control over film thickness is the layer-by-layer (LbL) assembly method [53,54]. Recently, titanium (IV) bis (ammonium lactato) dihydroxide (TALH) has attracted much attention as a precursor of TiO₂ since it is relatively stable at ambient temperature in water and it can form TiO₂ nanocrystals at temperature less than 100 °C [55]. Here, the LbL assembly of TALH with several cationic polyelectrolyte and crystallization of TiO₂ in the LbL films were examined.

The LbL assembly of TALH and polycation was monitored by QCM. Poly(diallyldimethylammonium chloride) (PDDA) was employed as a polycation. Figure 6 shows the QCM frequency shifts of TALH/PDDA multilayers as a function of the number of deposited layers. The regular film growth was found. The total mass increase after 5 bilayer deposition was ca. 8 μg cm⁻².

Figure 7 shows TEM image and electron diffraction (ED) pattern of the LbL film of (TALH/PDDA)₅. The film was directly deposited on a carbon-coated Cu TEM grid.

![FIGURE 5 Tapping mode AFM images of the first layer formed with the anionic cerasomes (a) and the second layer formed with the cationic cerasomes (b). For color version of this figure, the reader is referred to the online version of this book.](image)

![FIGURE 6 QCM frequency shifts for the LbL assembly of TALH and PDDA: open square, TALH layer; open circle, PDDA layer.](image)
From this figure, it can be confirmed that TiO$_2$ nanocrystals were formed with ca. 2 nm in the diameter without annealing. The spacing of the lattice fringe was 3.45 Å corresponding to the d-value of (101) in the anatase type of TiO$_2$ crystal. The ED pattern also shows the existence of crystalline matters in the films. The obtained fringe patterns can be assigned to the pattern of anatase. In addition, the crystalline phase of TiO$_2$ nanoparticles in the obtained film is not single phase. Interestingly, the existence of a minor phase of TiO$_2$(B) was also confirmed from ED pattern.

4. ORGANIC–INORGANIC HYBRID HOLLOW CAPSULES PREPARED BY COLLOID TEMPLATING

The preparation and study of core-shell colloids are of burgeoning interest, principally because such particles can display novel and enhanced properties (e.g., electrical, magnetic, catalytic, mechanical, chemical, and optical) when compared with those of the core particles [56,57]. Microcapsules have also attracted widespread interest because of their potential application in the areas of drug delivery, agriculture, and the food and cosmetics industries [58]. The core-shell particle and the hollow capsules formed with organic–inorganic hybrid materials are possible candidates for these applications since they can show characteristics of both organic and inorganic matters. A versatile method that affords microcapsules with control over their size, stability, loading, and release properties is the LbL assembly technique [53,54,59]. On the other hand, metal oxide precursors are widely used to form thin inorganic coatings on colloid surfaces. Herein, the preparation of organic–inorganic hybrid capsules by a combination of LbL assembly and sol–gel methods is reported. Polyelectrolyte capsules were fabricated by the alternate assembly of poly(sodium 4-styrenesulfonate) (PSS) and PDDA onto melamine formaldehyde (MF) particles, followed by removal of the MF particles by dissolution with HCl. An SiO$_2$ coating was deposited on the capsules by dispersing the polyelectrolyte-coated capsules in ethanol, adding tetraethylorthosilicate (TEOS), ammonia, and water to cause hydrolysis and condensation of the alkoxide. Figure 8 shows SEM and TEM images of hybrid hollow capsules formed with polyelectrolyte multilayers and SiO$_2$. From the SEM image, the particles have a spherical structure without any folds and creases. The numerous folds, creases, and flattening of the capsules are usually observed for SEM image of the hollow capsules.

**FIGURE 7** TEM (a) and electron diffraction pattern (b) of (TALH/PDDA)$_3$ multilayered film formed on a carbon-coated Cu TEM grid.

**FIGURE 8** SEM (a) and TEM (b) images of (PSS/PDDA)$_5$/SiO$_2$ capsules.
formed with polyelectrolyte multilayer without SiO₂ [59]. Therefore, the siloxane network of surface SiO₂ strengthens the shell of the capsules. In addition, the size of the capsules well reflected the original size of the core templates, ca. 3.0 µm. The TEM image represents the hollow structure of the particles. The image also indicated that the shell has homogeneous thickness. The thickness is controllable by the reaction condition of sol–gel process. These techniques are applicable to the preparation of various kinds of hybrid capsules. Especially, the development of “smart capsules” with stimuli-responsive property is now focused in several research groups. These systems will be aimed at applications in medical, cosmetics, and agriculture fields.

Optically addressable delivery systems [60–63] have been the subject of much research in recent years, because irradiation with light promises to be a more direct means of releasing the active materials than an environmental trigger such as a change in pH, temperature, ionic strength, or enzyme concentration [64]. For example, Sershen and co-workers [65] developed a photothermally responsive macroscopic hydrogel (i.e., a copolymer of N-isopropylacrylamide and acrylamide incorporating gold–gold sulfide nanoshells). Although the hydrogel undergoes a dramatic volume collapse upon illumination with light in the near-infrared (NIR) region (700–1500 nm) where most tissues show only weak light absorption [66], the system has not as yet been translated to colloidal delivery vehicles. Recently, Tao et al. [62] and Skirtach et al. [63] reported polyelectrolyte multilayer capsules doped with congo red dye molecules and silver nanoparticles, respectively. These light-responsive colloids, however, rely on irradiation with light outside the “biological window” (i.e., non-NIR laser light) to induce release. To this end, Caruso et al. prepared optically addressable colloidal carriers that show potential for bioapplications by infiltrating light-absorbing gold nanoparticles into the shell of polyelectrolyte multilayer capsules (Figure 9) [60,61].

The gold nanoparticles within the capsule shell absorb in the NIR region, and because the pulses of NIR laser light required to induce release are short, the laser light energy is effectively confined to the capsule shell (Figure 10a). Consequently, laser-induced release may be achieved in vivo without damaging surrounding tissues and the encapsulated biomaterials. The release of various materials from polyelectrolyte/gold nanoparticle capsules upon irradiation at 1064 nm with a series of 10 ns pulses from a Q-switched neodymium–yttrium–aluminum–garnet (Nd:YAG) laser was investigated. It was demonstrated that lysozyme (which was encapsulated within the polyelectrolyte/gold nanoparticle shell via LbL assembly on the surface of lysozyme crystals) can be released on demand without a significant loss of bioactivity following irradiation with...
short pulses of NIR laser light. The enzyme activity following laser irradiation is comparable to that caused by mechanical rupturing of the capsules (Figure 10b).

Fluorescein isothiocyanate (FITC)-labeled dextran (a convenient model of DNA which was post-loaded into preformed capsules by exploiting the pH-dependence of the shell permeability) is released upon laser irradiation provided the capsules contain gold nanoparticles within their shell [66]. Electron microscopy confirmed that the laser irradiation has no apparent effect on capsules without the light-absorbing shell component (Figure 11). Based on these findings [60,61] and those of previous optical studies [67–69], Caruso et al. proposed that the laser-induced release results from the gold nanoparticle-mediated heating of the capsule shell to extreme temperatures, which produces significant thermal stresses that ultimately cause the shell to rupture (Figure 10a). It is believed that phase explosions around the gold nanoparticles and electron ejection [70] from the gold nanoparticles may also play a role in the laser-induced release.

Despite these studies, there are considerable challenges in encapsulating and releasing low-molecular-weight molecules (including drugs) from LbL-engineered capsules. The stable encapsulation of small molecules in polyelectrolyte capsules is considerably more difficult than for macromolecules, because polyelectrolyte multilayers exhibit a high
permeability to low-molecular-weight compounds [71]. This permeability can be reduced by the deposition of lipid coatings on the capsules [72], thus offering a pathway for encapsulating small molecules. However, such systems can be unstable under a variety of conditions, especially at different temperatures, as the permeability of the lipid (and hence the capsules) is highly temperature sensitive. Therefore, there is a need for stable capsules loaded with low-molecular-weight materials that can be triggered to controllably release the encapsulated materials.

Recently, organic–inorganic hybrid capsules fabricated by a combination of LbL assembly and sol–gel methods were proposed. It shows that these capsules are suitable for the stable encapsulation of a low-molecular-weight dye (phenol red), which can be released upon UV-irradiation. TiO2 was selected as the UV-responsive component since it can decompose organic materials by photocatalytic reaction. Such a reaction was used to trigger rupturing of the polymer capsule walls, releasing the dye. The binary inorganic oxide system SiO2–TiO2 was selected since anatase TiO2 nanocrystals can be obtained from SiO2–TiO2 gel films via treatment at less than 100 °C [73]. It is also shown that release of the dye can be tuned through controlling the SiO2:TiO2 composition and UV-irradiation intensity.

Figure 12 shows the procedure used for preparing the organic–inorganic hybrid capsules by LbL assembly and sol–gel chemistry. Polyelectrolyte capsules were fabricated by the alternate assembly of poly(sodium 4-styrenesulfonate) (PSS) and poly(diallyldimethylammonium chloride) (PDDA) onto melamine formaldehyde (MF) particles, followed by removal of the MF particles by dissolution with HCl [2]. The capsules, terminated by a PSS layer, were subsequently coated with a cationic lipid bilayer of dioctadecyldimethylammonium bromide (DDAB). Phenol red was post-loaded into the capsules by switching the lipid-bilayer membrane permeability through the “gel” to “liquid crystalline” phase transition. A SiO2 coating was deposited on the capsules by dispersing the DDAB-coated capsules in ethanol, adding tetraethylorthosilicate (TEOS), ammonia, and water to cause hydrolysis and condensation of the

![FIGURE 12](image_url)  
**FIGURE 12** Schematic drawing of preparation process of UV-responsive capsules.

![FIGURE 13](image_url)  
**FIGURE 13** SEM images of (PSS/PDDA)5/PSS/DDAB/SiO2 capsules (a, b) and (PSS/PDDA)5/PSS/DDAB/SiO2-TiO2 (75:25) capsules (c, d) before (a, c) and after (b, d) UV-irradiation (20 mW cm$^{-2}$, 60 min).
alkoxide. In the case of the SiO₂–TiO₂ coating on the capsules, the precursor sol was prepared from TEOS and tetra-n-butylorthotitanate (TBOT). TEOS, TBOT, and acetylacetone, as a chelating agent, were dissolved in ethanol. A previous report by Wang and Caruso has shown that the sol–gel coating of polyelectrolyte multilayers assembled on particles results in uniform organic–inorganic coatings [74]. The effect of UV-irradiation on the morphology of the capsules was examined by scanning electron microscopy (SEM). Figure 13 shows SEM images of the organic–inorganic hybrid capsules before and after UV exposure. Prior to UV-irradiation, the particles have a spherical structure for both SiO₂-polymer and SiO₂/TiO₂-polymer capsules. After UV-irradiation, the SiO₂-polymer capsules remain spherical and intact, whereas the SiO₂/TiO₂-polymer capsules completely decomposed. In TiO₂, electron–hole pairs are generated through excitation with UV light, and the holes residing in the valence band of TiO₂ have high oxidative activity [75]. In the current study, the polyelectrolyte capsules are decomposed by the photocatalytic effect of the TiO₂ in the inorganic layer.

The effects of UV-induced release of the encapsulated dye molecules were examined. Phenol red was selected for use in these experiments since it has been widely examined in release studies from colloidal materials, including liposomes. Phenol red can act as an indicator, changing color at different pH values [76]. The dye was encapsulated in the internal aqueous phase of the lipid-coated capsules at pH 7.0 above the phase transition temperature of the lipid membrane (ca. 40°C). At this condition, the molecular permeability of the membrane is high. After cooling to 20°C, the dye was sealed into the capsules because of the reduced permeability. The capsules were then washed and coated with SiO₂/TiO₂. The external aqueous phase was changed to pH 9.4. When the dye was released to the outer phase of the capsules, the color changed to purple. Therefore, the release behavior can be detected by an increase in absorbance at 550 nm (maximum absorbance of phenol red at pH 9.4). The control experiment was examined by using SiO₂-coated capsules. Figure 14 shows the changes in absorbance (550 nm) with UV-irradiation time. These plots are directly comparable since the same amount of

FIGURE 14 Release of phenol red from the organic–inorganic hybrid capsules, monitored by changes in dye absorbance at 550 nm as a function of UV-irradiation time. (a) Capsules coated with SiO₂ and SiO₂–TiO₂ (75:25 and 95:5) were examined. The UV intensity was fixed at 20 mW cm⁻². (b) Several UV intensities (1, 5, 10, and 20 mW cm⁻²) were used to irradiate capsules coated with SiO₂–TiO₂ (75:25). (c) Schematic illustration of color change of the dispersion containing phenol red loaded capsules by UV-irradiation. For color version of this figure, the reader is referred to the online version of this book.
encapsulated dye was employed in each experiment. The SiO₂/TiO₂-coated capsules retained the dye, with no dye release observed over a period of 24 h. When the SiO₂/TiO₂-coated capsules (SiO₂:TiO₂ = 75:25 and 95:5) were irradiated with UV light, the dye was released (Figure 14a). The dye release rate was controllable through controlling the SiO₂/TiO₂ composition in the capsules, with quicker release observed for the capsules containing a higher TiO₂ content. In contrast, there was negligible increase in the absorbance after irradiation for 60 min for the SiO₂-coated capsules, reflecting virtually no dye release. These results indicate that the presence of TiO₂ in the capsule shells decomposes the polyelectrolytes, causing release of the dye. The release rate of the dye was also tunable by changing UV intensity. When UV intensities greater than 5 mW cm⁻² were used to irradiate the SiO₂/TiO₂ (75:25)-coated capsules, the encapsulated dye was released with 60 min. In contrast, a UV intensity of 1 mW cm⁻² resulted in considerably slower release (ca. 150 min) (Figure 14b).

5. SUMMARY

This chapter has highlighted some of the recent works in the area of organic—inorganic nano-hybrid materials prepared via supramolecular organization. Especially, hollow spheres with organic—inorganic hybrid structures have been described in detail. Cerasomes are lipid-bilayer vesicles with siloxane networks on their surface. The inorganic surface provides morphological stability of supramolecular structure. Therefore, this system was inspired from cell walls of biomembranes. Due to this high morphological stability, the artificial multicellular assembly was achieved via LbL assembly. Cerasomes can be expected as functional hybrid hollow spheres to create bioinspired nanodevices. Stimuli-responsive hybrid hollow spheres have also been developed by the colloid templating technique. These spheres should be applicable for the controlled delivery systems. The target external stimuli for the trigger of release can be controlled by selection of inorganic components of spheres. Gold nanoparticles can work as responsive agents for the near-infrared laser. TiO₂ can be used for UV-responsive system by its photocatalytic activity. The release rate of this system is tunable by adopting SiO₂—TiO₂ binary oxide system. The author anticipates that, with further development, these systems will make an important contribution not only to the development of organic—inorganic hybrid materials but also in the future application of bioinspired processing in nanoscience and nanotechnology.

ACKNOWLEDGMENTS

The author is grateful to the co-workers whose names appear in the references for their contributions to this work. This work was partially supported by a Grant-in-Aid for Scientific Research on Innovative Areas of “Fusion Materials: Creative Development of Materials and Exploration of Their Function through Molecular Control” (No. 2206) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

REFERENCES

Handbook of Advanced Ceramics


Chapter 11.1.10

Precursor-Derived Ceramics*

Markus Weinmann, Emanuel Ionescu, Ralf Riedel and Fritz Aldinger

H.C. Starck GmbH, Im Schleeke 78-91, 38642 Goslar, Germany, Technische Universität Darmstadt (Darmstadt University of Technology) - Institut für Materialwissenschaft (Institute of Materials Science), Petersenstraße 23, 64287 Darmstadt, Germany, Former Affiliate: Max-Planck-Institut für Metallforschung (Max-Planck-Institut für Metallforschung) (at present Max-Planck-Institut für Intelligente Systeme), Germany.

Chapter Outline

1. Introduction 1025
2. Precursor Synthesis 1026
3. Polymer-to-Ceramic Transformation 1050
4. High Temperature Properties 1065
5. Applications 1080
6. Conclusions 1092
References 1092

1. INTRODUCTION

The condensation of organometallic compounds into merely inorganic materials by a proper thermal treatment under controlled atmosphere is a unique and fairly simple process of producing new types of ceramics, which are due to their origin—referred to as precursor- or polymer-derived ceramics (PDCs).

This procedure allows for an easy control of elemental composition, chemical homogeneity, and material architecture on an atomic scale. It provides an inimitable access for controlling and adjusting the design and the microstructure of ceramic materials that cannot be achieved using conventional processing techniques such as melting or sintering. In recent years, this process scheme has therefore created substantial interest, both scientifically and for practical purposes.

The synthesis and processing of PDCs usually do not require any additives. PDCs thus display the intrinsic chemical and physical—chemical properties of the constituting pure phases, for example, an exceptional oxidation and corrosion stability as well as crystallization and creep resistance up to very high temperatures. Multinary PDCs frequently possess even superior properties compared with their constituting binary or ternary subsystems.

Several review articles [1–5] and books [6,7] have been published on this topic, as well as special issues in the Journal of the American Ceramic Society ("Ultrahigh-Temperature Polymer Derived Ceramics") [8], Journal of the European Ceramic Society ("Polymer-derived Ceramics") [9], and Soft Materials ("Preceramic Polymers") [10], which focused on the synthesis and properties of PDCs.

The overall process for the formation of PDCs is outlined in Figure 1.

This consists of three major steps. At first, suitable monomers are transformed into polymeric macromolecules referred to as precursors, which are subsequently cross-linked at moderate temperatures into preceramic networks

FIGURE 1 Flow diagram for the preparation of PDCs.
with defined rheological behavior to provide proper processing capabilities. The networks are finally transformed into inorganic materials by a thermal heat treatment at temperatures usually ranging from 1000 to 1400 °C.

The general idea behind the process concept is to generate preferred structural features in the organometallic precursors (this can mostly be realized by comparably simple preparative organometallic chemistry) and to subsequently transform the precursors into ceramics under retention of the especially designed building blocks. Thermolysis must thereby include a series of controllable condensation steps, which are aimed for the specific architecture of the desired ceramic material. As a consequence of being mostly reaction controlled, thermolysis is especially valuable for the densification of covalently bonded inorganics.

Conventionally, low diffusion capability of nonoxide ceramics even at very high temperatures is compensated by the use of sintering aids. Such additives are mostly of the oxide type and therefore degrade the otherwise unique properties of nonoxide ceramics such as thermal, chemical, and mechanical stability over wide temperature ranges. In contrast, thermolysis of ceramic precursors provides a means to realize the production of nonoxide ceramic thereby keeping up their excellent properties without compromises.

One remarkable consequence of the low atomic mobility is that as-thermolyzed PDCs can be formed which remain amorphous up to rather high temperatures. The reason for this is that substantial thermal activation is required for nucleation, grain growth, and the formation of crystalline phases. Crystallization thereby very much depends on the chemical/elemental composition, the atomic structure and the microstructure of the materials, and the atmosphere applied. Under certain conditions, crystallization can be retarded even up to 1800 °C providing novel glass-like structural materials with superb thermomechanical properties. However, also crystalline materials are of great interest because their microstructure formation can be controlled during devitrification to a large extent providing, for example, a means for stabilizing nanosized morphologies.

This chapter intends to review the fundamentals of synthetic approaches and the processing of silicon-based PDCs, to reveal the possibilities of microstructure development and to give hints on their potential application. It has been shown that the basis for the microstructure design of PDCs is the type of macromolecule, that is, its chemical composition and molecular structure. Therefore, at first synthetic procedures, which deliver different types of organometallic, silicon-based ceramic precursors are evaluated. In the following, thermolysis reactions will be considered, because the structure of the ceramic materials is also widely determined by the structure of the molecular precursors. Further on, phase reactions controlling the high temperature stability of these materials will be discussed and finally recent trends in technologically relevant applications be reported.

2. PRECURSOR SYNTHESIS

2.1. General Comments

The most intensively investigated preceramic polymers for the preparation of PDCs are polysilanes, polycarbosilanes (PCSs), polysiloxanes, as well as polysilazanes and polysilylcarbodiimides [11–13]. Since the Yajima process to synthesize silicon carbide fibers using PCSs, significant development in the synthesis and processing of PDCs has been achieved. Thus, using silicon-based polymers, technologically important ceramic components such as complex-shaped monoliths, fibers, coatings or infiltrated porous media, and powders can be prepared.

The molecular structure and the type of preceramic polymer influence not only the chemical composition but also the microstructure and the phase composition of the final ceramic [14]. In this way, the chemical and physical properties of PDCs can be varied and adjusted to a great extent by the design of the molecular precursor. Thus, the synthesis of preceramic polymers is one of the key issues in the PDC field.

2.2. Precursors to Silicon Carbide

Silicon carbide is the only (solid) binary compound of silicon and carbon. It is used for multifold purposes. The most important feature is its outstanding mechanical behavior. Silicon carbide is one of the most frequently used ceramic hard materials [15].

Silicon carbide has an extremely low self-diffusion coefficient, even at a very high temperature. Processing, that is, compaction of silicon carbide powders thus requires sintering aids, which may—to a certain extent—unfavorably determine the properties of the processed ceramic materials [16–19].

In contrast, precursor thermolysis of suitable organo-silicon compounds is a method, which delivers sinter aid-free silicon carbide. As will be shown in the next sections, suitable precursors for this purpose are, for example, polysilanes, PCSs, polyalkenylsilanes, polyalkynylsilanes, or their mixtures (copolymers or polymer blends). In the following, the most important of the abovementioned organosilicon polymers will be reviewed chronologically, starting with polysilanes, which are composed of alternating [SiR₂] units, followed by polysilaethylenes, which are composed of [SiR₂—CR₂] building blocks. Finally, PCSs with alternating [Si—C—C] units will be discussed.
2.2.1. Polysilanes

Polysilanes are well-studied polymers [20–22]. However, although many research groups all over the world have extensively investigated synthetic approaches, especially with the goal to obtaining polysilanes as precursors to phase-pure silicon carbide, there are only a very limited number of procedures described in the literature that deliver near-stoichiometric silicon carbide.

Polysilanes, also referred to as poly(silylene)s, usually have a linear polymer backbone consisting entirely of silicon atoms to which two organic substituents R are bonded. When sufficiently purified, poly(diorganosilane)s, \([\text{SiR}_1\text{R}_2]^n\) (R \(_1\), R \(_2\) = singly bonded organic group) are stable in oxidative and hydrolytic environments and thus can be handled easily without requiring inert gas techniques. Polysilanes are often insoluble, not meltable, and intractable, which impedes their applicability in many practical processes. Chemical and physicochemical properties are strongly influenced by the molecular weight and the nature of the silicon-bonded substituents.

Polysilanes were first synthesized in the early 1920s by Kipping and Sands [20] by a Wurtz-Fittig type dehalogenation reaction of dichlorodiphenylsilane with molten potassium in xylene delivering poly(diphenylsilane).

\[
\begin{align*}
\text{Cl} & \quad \text{Si} & \quad \text{Cl} \\
\text{Ph} & \quad \quad & \quad \text{Ph} \\
- & \quad \text{KCl} & \quad \text{Ph} & \quad \text{Si} & \quad \text{Ph} & \quad n \\
\text{Ph} & \quad \text{C}_6\text{H}_5
\end{align*}
\]

In 1949, Burkhardt published the first clear description of the synthesis of a polysilane with a potential for technological application [23]. Dichlorodimethylsilane was reacted with sodium in a Wurtz-Fittig type reaction in benzene in analogy to the abovementioned procedure.

\[
\begin{align*}
\text{Cl} & \quad \text{Si} & \quad \text{Cl} \\
\text{CH}_3 & \quad \quad & \quad \text{CH}_3 \\
- & \quad \text{NaCl} & \quad \text{CH}_3 & \quad \text{Si} & \quad \text{CH}_3 & \quad n \\
\text{CH}_3 & \quad \text{CH}_3
\end{align*}
\]

Poly(dimethysilane) is a white powder, which is poorly soluble in organic solvents and which does not melt without decomposition. It will be highlighted below that it represents the first precursor for the production of refractory silicon carbide fibers.

In the meantime, a large number of polysilanes with different silicon-bonded substituents and molecular weight have been synthesized. Product yields are frequently unsatisfactory because of (i) the low solubility of the polymers in organic solvents and (ii) the expandable processing (purification) including hydrolysis to remove excess alkali metal. A further essential drawback of the Wurtz coupling is the limited functional group tolerance, which allows only for the synthesis of polysilanes with more or less “inert” side groups such as alkyl or aryl groups.

The mechanism of the Wurtz-type coupling of chlorosilanes seems to be very complex. Alkali metal surface reactions are probable, but dehalogenation reactions in solution have also to be considered. Moreover, the mechanisms involving silyl radicals are discussed as well as mechanisms based on silyl anions, which appear to be key species in the chain growth of polysilanes.

An alternative procedure for the synthesis of polysilanes, which delivers products in high purity and higher yields, is the dehydrocoupling of silanes in the presence of suitable catalysts such as \(\eta^2\)-alkynyl titanocene or -zirconocene as described by Chang and Corey in 1989 [24].

\[
\begin{align*}
\text{R} & \quad \text{H} & \quad \text{Si} & \quad \text{H} & \quad - & \quad \text{H}_2 & \quad \text{Si} & \quad \text{R} \\
\text{Catalyst} & \quad - & \quad \text{H}_2 & \quad \text{Si} & \quad \text{R} \\
\text{Catalysts:} & \quad & \eta^2\text{-Alkynyl-Titanocenes,} & \quad \eta^2\text{-Alkynyl-Zirconocenes}
\end{align*}
\]

In contrast to the Wurtz-Fittig reaction, no solid byproducts are formed. The only byproduct is hydrogen. As a consequence, purification steps such as hydrolysis and filtration are not required. The authors also suggested a conclusive mechanism of the silane polymerization by dehydrocoupling.

By elimination of the side-on coordinated alkynyl ligand, a highly reactive 14VE (VE = valence electron) metalloocene species forms, which inserts into a Si–H bond of the silane, thus forming a silymetalallocene. This species most likely eliminates hydrogen thereby forming a metallasilanediyl intermediate and subsequently adding a silane molecule by hydrosilylation of the Si–M bond. The disilylated metalallocene rearranges by a hydrogen shift to the transition metal center and reductive elimination of the oligomer/polymer (Scheme 1).

Woo et al. who investigated dehydrocoupling reactions of silanes using zirconocene and hafnocene hydride complexes as catalysts postulated alternative mechanisms [25]. In contrast to the above, they suggested a mechanism that is based on two \(\sigma\)-bond metathesis reactions that pass through four-center transition states (Scheme 2).

2.2.2. Polycarbosilanes

In contrast to polysilanes, polymer backbones in PCSs are not exclusively composed of silicon atoms. Here, aliphatic
units link the silicon atoms. Frequently, also unsaturated units as well as aromatic ring systems serve as building blocks for the formation of PCSs.

PCSs of the general structure \([\text{R}_2\text{Si} \text{CH}_2]_n\) have been synthesized by a variety of reaction pathways. The most important and intensely investigated access to PCSs certainly is the Yajima process\(^{[26]}\). This process gained much attractiveness as it thermally converts insoluble and not meltable poly(dimethylsilane) into soluble and processable poly(methylsilylene-methylene), \([\text{HSi(CH}_3]_n\text{CH}_2\text{]}_n\). This precursor can be either cast or spun. Silicon carbide fibers obtained from this precursor are commercially available under the trade names Nicalon\(^TM\) or High Nicalon\(^TM\), depending on the processing applied (cf. Section 5.5).

The key step in the Yajima process is a rearrangement of the polymer backbone involving pendent methyl groups, referred to as Kumada rearrangement or Kumada reaction, which occurs at around 400 °C. It is a methylene migration from one of the silicon-bonded methyl groups attached to the polymer backbone into the polymer chain. The process is radical induced.

The mechanism including the single key steps in the Kumada rearrangement of poly(dimethylsilane) is shown in Scheme 3. Initially, Si—Si bonds in the polymer are homolytically cleaved thus forming silyl radicals, followed by H-radical migration from a methyl group to the silyl radical (initiation). The methylene radical that is thereby generated inserts into an Si—Si bond whereby a new silyl radical is formed (rearrangement). Propagation proceeds by a repeated H-radical migration to the silicon radical with the formation of a methyl radical. Termination most probably takes place by radical recombination.

PCSs with Cl building blocks can also be obtained by Grignard reactions of chloromethyl trichlorosilane \(\text{Cl}_3\text{Si—CH}_2\text{Cl}\) in diethyl ether solution \(^{[27]}\). Subsequent treatment of the as-obtained poly(dichlorosilylene-methylene) with LiAlH_4 delivers branched PCS with an idealized structure \([\text{H}_2\text{Si—CH}_2]_n\). Due to its compositional analogy to polyethylene, it is also referred to as polysilaethylene. Thermalysis of polysilaethylene delivers near-stoichiometric SiC.
The disadvantages of this reaction pathway are long reaction times that are required for obtaining sufficient polymer yields. Changing the solvent toward a more polar one such as tetrahydrofuran strongly influences the selectivity of the reaction. If starting from chloromethyl dialkyl- or diarylchlorosilanes, according to reaction pathway (a), 1,3-disilacyclobutane derivatives are obtained.

The transformation into linear polymers can be achieved by ring-opening polymerization (ROP) reactions using Pt-group metal complexes, such as chloroplatinic acid as catalysts [28]. Wu and Interrante developed a modified multistep polysilaethylene synthesis [29] starting from methyltrichlorosilane, \( \text{Cl}_3\text{Si} - \text{CH}_3 \), which is a cheap, commercially available compound. Halogenation using chlorine in the presence of ultraviolet light delivers chloromethyl trichlorosilane, \( \text{Cl}_3\text{Si} - \text{CH}_2\text{Cl} \) (this reaction is not shown in Scheme 4). \( \text{Cl}_3\text{Si} - \text{CH}_2\text{Cl} \) could not directly be converted into 1,1,3,3-tetrachloro-1,3-disilacyclobutane. Rather, two of the three silicon-bonded chlorine atoms were initially substituted by ethoxy (\(-\text{OEt}\)) groups, which...
could be selectively realized in alcoholysis reactions before the reductive coupling. The ethoxy groups in 1,1,3,3-tetraethoxy-1,3-disilacyclobutane were subsequently replaced with chlorine atoms by a treatment with acetic chloride acid in the presence of FeCl₃. Pt-catalyzed ROP and subsequent reduction with LiAlH₄ then delivered linear polysilaphenylene in an approximately 60% yield.

Procopio and Berry published a procedure for the synthesis of oligosilaethylene by dehydrogenative coupling reactions [30] (Scheme 5).

Oligomeric carbosilanes were obtained from trimethylsilane under mild conditions in the presence of H₂Ru(SiMe₃)(PMe₃)₃, which was obtained in situ from H₂Ru(PMe₃)₄ and HSiMe₃. The proposed mechanism involves the formation of a disilyl ruthenium complex, which upon a β-hydride shift transforms into a η²-silylene species (in analogy to ethylene, this is the monomer unit of poly(silylene-methylene)). Insertion of the silylene into the remaining Ru–Si bond results in the formation of a singly silylated ruthenium complex and chain growth by one SiMe₂CH₂ unit.

PCSs with aliphatic C₂ building blocks in the polymer backbone are precursors to SiC, which have reached remarkable attractiveness. Most of the publications in this field trace back to work of Corriu et al. who have applied hydrosilylation of vinylsilanes, (R¹)(R₂)SiH(HC=CH₂) (R¹, R₂ = H, Cl, alkyl, aryl, NR₂), for the preparation of linear PCSs of the general type [(R¹)(R₂)Si–C₂H₄]ₙ [31,32]. Even though this type of hydrosilylation does not occur regioselectively, formation of the β-addition product (C₂H₄=CH₂CH₂) is strongly preferred.

Polymerization by hydrosilylation of vinylsilanes requires a catalyst such as H₂PtCl₆ dissolved in 2-propanol, also known as Speyers’ reagent [33]. In 2-propanol solution, PtCl₆²⁻ is reduced to PtCl₄²⁻, which is the catalytically active species.

The mechanism which involves 16VE and 18VE species is shown in Scheme 6. Initially, one of the chloride ligands in the 16VE fragment PtCl₄²⁻ is replaced with a vinylhydridosilane ligand, thus forming a η²-olefin PtCl₃(CO) complex. In a second step, a vinylhydridosilane molecule adds oxidatively (insertion of [Pt] into Si–H) whereby an octahedral 18 VE complex forms, followed by an insertion of the η²-coordinated olefin into a platinum–hydrogen σ-bond. The vacancy, which forms upon this migration, is occupied by an external chloride ligand that adds to the complex. Finally, the carbosilane is eliminated reductively with regeneration of the catalytic species PtCl₄²⁻.

Similarly, PtCl₄²⁻-catalyzed hydrosilylation of alkynylhydridosilanes of the general type (R¹)(R₂)SiH₂(C≡CR₃) is a synthetic pathway for obtaining linear PCSs in which alternating silicon atoms and olefinic C₂ units constitute the polymer backbone [34,35].

Formation of poly(alkenylsilanes) of this type is only observed if R¹, R₂ ≠ H. The use of dihydridoalkynylsilanes (R)SiH₂(C≡CH) as starting compounds, in contrast, delivers highly crosslinked, glass-like PCSs. The reason is that the olefin units, which are generated by the hydrosilylation of the alkynyl units, may react further and transform into aliphatic C₂-building blocks.
PCSs of the general type \([ (R^1)(R^2)Si\equiv C\equiv C ]_n \)
in which silicon atoms are linked via \(C\equiv C\) building blocks, are not accessible by hydrosilylation reactions. In contrast, metathesis reactions of dichlorosilanes with dialkalimetal acetylides or the respective Grignard reagents deliver this type of polymers in good yields [34].

The alkynyl unit represents a highly reactive site within the polymer backbone. This is an important issue with respect to obtaining highly crosslinked precursors for ceramics to prevent polymer degradation and thus volatilization of low molecular weight species during thermolysis. Moreover, the alkynyl group can be functionalized by, for example, hydroboration, hydrosilylation, or attachment of transition metal complex fragments by coordination, providing an access for the synthesis of multicomponent ceramics. In addition, poly(butadiynylsilanes) \( [(R^1)(R^2)Si\equiv C\equiv C\equiv C ]_n \) (also referred to as poly[(silylene)diacetylenes] or poly[(silylene)diethynylenes]) obtained from 1,4-dilithio-1,3-butadiyne proved to be precursors with interpenetrating networks to silicon carbide/metal carbide \((M = Ti, Nb, Ta)\) nanocomposites [36]. Preparation of branched polymers by regiospecific hydrosilylation of poly[(silylene)diethynylenes] was also reported [37].

2.3. Precursors to Silicon Nitride

For the preparation of silicon nitride-based PDCs oligo- or polysilazanes, \([(R^1)(R^2)Si\equiv N(R^3)]_n\) \((R = H, alkyl, aryl, etc.)\) or poly(silylcarbodiimide)s, \([(R^1)(R^2)Si\equiv N\equiv C\equiv N ]_n \) have been studied intensively during the last 30 years, resulting in an immense number of structurally different precursors that were published. In general, such precursors deliver silicon nitride/carbon or silicon nitride/silicon carbide/carbon composites if \(R^1\), \(R^2\), and/or \(R^3\) contain
carbon atoms. The formation of phase-pure silicon nitride can only be realized by thermolysis of perhydridopolysilazane (PHPS) in which \( R^1, R^2, \) and \( R^3 \) consist of hydrogen and/or nitrogen or by performing the polymer-to-ceramic conversion in an ammonia atmosphere. The latter, however, is limited to powders or parts with comparably small dimensions and/or cross sections.

### 2.3.1. Carbon-free Polysilazanes

Narsavage et al. published the polymer-to-ceramic conversion of tetrakis(ethylamino) silane in an inert gas atmosphere (Ar), from which they expected to obtain phase-pure silicon nitride [38]. The idea behind this concept was that silicon is exclusively bonded to nitrogen. It was expected that Si\(_3\)N\(_4\) formation by volatilization of the nitrogen-bonded ethyl groups takes place during thermolysis. However, carbon was not completely degassed, and the resultant material was an amorphous silicon carbonitride.

Consequently, the only remaining possibility for obtaining carbon-free Si\(_3\)N\(_4\) is thermolysis of polysilazanes, containing only Si, N, and H. The first attempts for the preparation of such polymers were reported already in 1885 by Schutzenberger and Colson [39], by ammonolysis of tetrachlorosilane according to Eqn (10). Polymer formation was accompanied by precipitation of ammonium chloride as a coproduct, which was removed by filtration. Glemser and Naumann performed more detailed investigations on this system [40].

\[
\text{Cl-Si-Cl + 3 NH}_3 \rightarrow [\text{Si(NH)}_2]_n + 4n \text{NH}_4\text{Cl}
\]

(Schematic representation)

Silicon diimide, Si(NH\(_2\))\(_2\), is a colorless solid, which is insoluble and infusible. With increasing temperature, it decomposes with the elimination of ammonia to form amorphous Si\(_3\)N\(_4\) at 1000 °C. Further heat treatment to 1400—1500 °C results in the crystallization of alpha-Si\(_3\)N\(_4\).

A processable Si\(_3\)N\(_4\) precursor, PHPS, can be obtained by the ammonolysis of dichlorosilane in polar solvents such as diethyl ether or dichloromethane as shown in Scheme 7 [41].

It is composed of linear and cyclic motifs but rapidly ages by loosing hydrogen and further crosslinking, thus increasing its viscosity gradually from becoming oily over waxy to glassy. Thermolysis to 1050 °C releases a ceramic material in 70% yield, which according to X-ray investigations consists of alpha-Si\(_3\)N\(_4\), beta-Si\(_3\)N\(_4\), and elemental silicon. Excess Si could be removed almost completely by performing thermolysis in an ammonia atmosphere. However, a critical issue with respect to transferring this process into a technical scale is the use of H\(_2\)SiCl\(_2\), which is a highly flammable gas, which can disproportionate with the formation of SiH\(_4\) and SiCl\(_4\). A modified process, allowing for a safer handling of the highly reactive starting compounds was developed at the Tonen Company. Before ammonolysis, H\(_2\)SiCl\(_2\) was modified by reaction with pyridine, which resulted in the formation of H\(_2\)SiCl\(_2\)+NC\(_3\)H\(_5\) [42].

Blanchard and Schwab modified the process developed by Seyferth et al. They avoided the formation of free silicon by synthesizing a PHPS, which delivers stoichiometric silicon nitride even after thermolysis in an inert gas atmosphere, using coammonolysis of dichlorosilane and trichlorosilane [43].

### 2.4. Precursors to Ternary Si—C—N Ceramics

#### 2.4.1. Carbon-containing Polysilazanes

While the synthesis of PHPS is limited to a small number of suitable starting compounds such as H\(_2\)SiCl\(_2\), HSiCl\(_3\), SiCl\(_4\), and NH\(_3\) or H\(_2\)N—NH\(_3\) as silicon and nitrogen sources, respectively, the chemistry of carbon-containing polysilazanes of the general type [(\( R^1 \))(\( R^2 \))Si—N(\( R^3 \))]\(_n\) (\( R^1 \), \( R^2 \), \( R^3 \) = H, alkyl, aryl etc.) is a very multifaceted field. The first publications in this regard were published by Rochow appeared already in 1966 [44].

In contrast to H\(_2\)SiCl\(_2\) and HSiCl\(_3\), functionalized chlorosilanes of the general type (\( R^1 \))(\( R^2 \))SiCl\(_2\) (\( R^1 \), \( R^2 \) = H, alkyl, aryl, halide etc.), which are in general used in the synthesis of Si—C—N polymers, are mostly cheap and commercially available. Therefore, much effort has been made within the last 30 years toward specifically designing polysilazanes for numerous applications. In contrast to PHPS that deliver amorphous Si—N ceramics, carbon-containing polysilazanes release amorphous Si—C—N ceramics after thermolysis to approximately 1100 °C. Interestingly, these materials possess improved high temperature mechanical properties compared to silicon nitride derived from PHPS. A particularly critical issue is “free,” that is, excess silicon. The reason is that silicon segregation is known to lower the crystallization temperature of silicon-nitride-based materials drastically [45]. In carbon-containing Si—C—N materials,
which usually contain “free” carbon, silicon segregation is suppressed because of the formation of silicon carbide.

There have been a remarkable number of synthetic pathways described in the literature, which deliver poly- silazanes. The desired application, the choice of the chemical composition and molecular structure of the monomers, reaction conditions and eventually further crosslinking finally determine the synthetic approach to the precursors. The procedure used most frequently for the synthesis of polysilazanes is ammonolysis or aminolysis of chlorosilanes.

Ammonolysis of chlorosilanes in general proceeds according to the following:

$$R_{4-x}SiCl_x + \frac{3}{2} \times NH_3 \rightarrow \frac{1}{n} [R_{4-x}Si(NH)_{x/2}]_n + x NH_4Cl$$

(11)

The single steps are (i) initial substitution of Cl atoms with NH$_2$ units followed by (ii) a condensation step upon which Si—NH—Si linkages are built. Depending on the number of chlorine atoms per silicon atom and the nature of the silicon-bonded groups R, monomeric, dimeric, cyclic, short chain linear, or highly crosslinked polysilazanes form.

Ammonolysis of monochlorosilanes R$_2$SiCl results in monomeric species. In the case of “small” silicon-bonded substituents R (R = H, CH$_3$), disilazanes R$_3$Si—NH—SiR$_3$ are formed. Certainly, if silicon-bonded substituents are sufficiently small and reaction temperatures are high, these species condense further, thereby transforming into tertiary amines bearing Si$_3$N motifs [46]. Against this, bulky substituents R, such as isopropyl, isobutyl, or mesityl inhibit the abovementioned condensation step due to steric effects, thus causing the formation of monomer silylamines R$_3$SiNH$_2$ [47,48].

Hexamethyldisilazane (HMDS), (H$_3$C)$_2$Si—NH—Si (CH$_3$)$_3$, synthesized by the ammonolysis of trimethylchlorosilane on a large scale, is industrially by far the most important monomeric silazane. Because it is a volatile liquid (boiling point: 125 °C), it cannot be directly used for the solid-state thermolysis process. However, it is an important single-source precursor for the preparation of Si—C—N coatings by the chemical vapor deposition (CVD) processes [49] and a very important source for the synthesis of polysilazanes by trans-amination (redistribution, metathesis) reactions [50].

Ammonolysis of dichlorosilanes R$_2$SiCl$_2$ delivers oligo- or polysilazanes of the general type [R$_2$Si—NH]$_n$, presuming sufficiently high reaction temperatures and sterically less demanding silicon-bonded substituents R such as hydrogen, methyl, ethyl, or phenyl. Reaction of ammonia with dichlorosilanes having sterically demanding substituents R such as i-propyl, t-butyl, or mesityl delivers silyldiamines, R$_2$Si(NH)$_2$ according to the above-described ammonolysis of monochlorosiloxanes R$_3$SiCl.

Usually, product mixtures are obtained, which consist of 6 and 8-membered rings (trimers and tetramers, respectively) besides low molecular (M < 2000) chain molecules. The low molecular weights of the tri- and tetrameric species result in low boiling temperatures and cause their volatilization during thermolysis, whereas the chain molecules frequently degrade during the heat treatment with volatilization of low molecular gas species [51]. As a consequence, ceramic yields of polysilazanes are frequently unsatisfactory.

Seyferth and Wiseman published a method for increasing the ceramic yields of cyclic silazanes bearing Si—H and N—H units considerably by crosslinking with catalytic amounts of KH [52,53]. In the case of cyclo-tetramethylsilazane, the ceramic yield could be increased from <30% to approximately 84%. The authors proposed a mechanism, which involves silylene—imine motifs as key species [53] (Scheme 8).

It is supposed that the Si—H units of a second Si$_4$N$_4$ ring add intermolecularly in a hydrosilylation type reaction to the highly reactive silylene—imine intermediate, which could not be isolated. The as-formed species can again be deprotonated and react further. However, there are also mechanisms desirable excluding silylene—imine formation, which were postulated for the dehydrocoupling of boron-modified silanes using n-butyl lithium as a catalyst [54]. Here, n-butyl lithium initially deprotonates the silazane N-atoms resulting in highly nucleophilic amides. Subsequent intermolecular substitution of a silicon-bonded hydride by an amide results in the formation of a new Si—N bond. The hydride released deprotonates an N—H unit, thereby closing the catalytic cycle.

A possibility for the synthesis of highly branched high ceramic yield polysilazanes is the attachment of reactive sites, for example, vinyl groups to the polymer backbone. These allow for further crosslinking reactions during thermolysis and thus very efficiently avoid polymer backbone degradation. Vinyl-substituted oligo- or polysilazanes can be obtained by ammonolysis of vinylchlorosilanes (H$_2$C—CHSi(R)Cl$_2$ (R = H, CH$_3$) according to the following:

$$\begin{align*}
\text{Cl-Si-Cl} + 3 \text{NH}_3 & \rightarrow \frac{1}{n} [\text{R-Si(NH)_{x/2}}]_n + 2 \text{NH}_4\text{Cl} \\
\text{R} & = \text{H, CH}_3
\end{align*}$$

(12)

Depending on the silicon-bonded unit R and reaction conditions applied, [(H$_2$C—CH)Si(R)—NH]$_n$ can crosslink
by three different reaction pathways (cf. Scheme 9): (i) olefin polymerization, (ii) hydrosilylation, and (iii) dehydrocoupling (the latter reactions only take place if $R = H$). It should be noted that both olefin polymerization and hydrosilylation are addition reactions, which take place without the formation of byproducts. This is an important issue that makes vinyl-substituted polysilazanes candidates for the production of polymer-derived ceramic matrix composites [55].

Nuclear magnetic resonance (NMR) spectroscopy and chemical analysis suggest that hydrosilylation is the main reaction in the thermally induced crosslinking of $[(H_2C=CH)SiH-NH]_n$ [56]. Addition of catalytic amounts of Speyers’ reagent can accelerate it. Thermally induced trans-amination (not shown in Scheme 9) in contrast, which would be accompanied with ammonia evaporation, occurs to a minor extent. By adding catalytic amounts of
potassium hydride, dehydrocoupling reactions become favored. Crosslinking in this case proceeds by Si–N bond formation with the elimination of hydrogen. Possible mechanisms for catalysis, hydrosilylation, and dehydrocoupling were already mentioned above and are not discussed here in detail. In contrast, potassium hydride promoted anionic polyaddition reaction of the vinyl groups (anionic olefin polymerization) was not observed.

Due to the absence of Si–H motifs in [(H₂C=CH)Si(CH₃)–NH]n, neither hydrosilylation nor dehydrocoupling can be considered for thermal crosslinking. However, based on the structural characterization by solid-state¹³C- and ²⁹Si-magic angle spinning Nuclear Magnetic Resonance (MAS NMR), Fourier Transform Infrared (FTIR), and elemental analysis, Bill et al. concluded that between 250 °C and 350 °C, which is approximately 120 °C higher than the temperature required for hydrosilylation reactions of [(H₂C=CH)SiH–NH]n, the vinyl groups are transformed into aliphatic hydrocarbons [57]. More detailed information on thermolysis reactions of polysilazanes are provided in Section 3.2.

Very important silazane-based precursors are copolysilazanes obtained by coammonolysis or α-aminolysis of chlorosilane mixtures. Polymerization of such mixtures allows for the synthesis of modified polymers with tunable properties such as solubility, rheology, and (latent) reactivity, which cannot be achieved using single-source precursors. Originally, copolysilazanes were developed to improve the processing of the precursors for silicon nitride fibers, that is, to adjust proper rheology and latent reactivity. Moreover, the importance of chemical constitution, that is, phase composition was recognized [58]. The main subject in this regard was to decrease the amount of free carbon in Si–C–N ceramics to shift decomposition temperatures toward higher values.

A commercially available precursor in this series is NCP200, a “polyhydridomethylsilazane,” which is obtained by the coammonolysis of dichlorodimethylsilane and dichloromethylsilane [59]. Because of commercial availability, low cost, good processability, and high ceramic yield of approximately 80% (this value, however, depends on the processing), this precursor could successfully be applied for the preparation of dense silicon carbonitride ceramics [60]. Structurally comparable is NN710 a “polyhydridomethylsilazane,” which can be synthesized by coammonolysis of dichlorodimethylsilane and dichloromethylsilane [61].

Recently, a structurally similar precursor, which is derived from this system and which delivers stoichiometric silicon carbide/silicon nitride, was published [62].

The polymer was obtained according to Scheme 10 by coammonolysis of H₂SiCl₂ and H₃CSiHCl₂ in a 3:1 molar ratio and subsequent base-catalyzed crosslinking using n-BuLi as a catalyst. While the as-obtained polymer is a viscous oil, which rapidly ages, the crosslinked precursor is a hard glass-like infusible solid, which is insoluble in all common organic solvents. In mass-spectra-coupled simultaneous thermogravimetric analysis (TGA-MS) investigations, it was shown that this type of polymer could be pyrolyzed without the formation of gaseous byproducts other than hydrogen, thus delivering amorphous SiC/Si₃N₄ with a ceramic yield of 94.5% (c.f. Figure 10).

According to the procedures described above, ammonolysis of trichlorosilanes RSiCl₃ usually results in the formation of [(R)Si(NH)₁.₅]ₙ. Such poly(silsesquiazane)s (these are polymeric silazanes in which three silicon-bonded substituents are nitrogen atoms and in which N:Si = 1.5) were first published in 1967 by Andrianov and Kotorelev [63] who performed ammonolysis of methyltrichlorosilane.

![Scheme 10](1035)

**Scheme 10** Synthesis of a precursor for stoichiometric SiC/Si₃N₄ [62]. A key step is subsequent crosslinking of the initially obtained precursor in order to avoid thermal degradation with volatilization of low-molecular weight species other than hydrogen.
In contrast to poly(silsesquioxane)s, \((\text{RSiO}_{1.5})_n\), which can be obtained by hydrolysis of trichloro- or trialkoxysilanes under water-starved conditions and which are structurally well investigated [64–66], there is not much known about the molecular structure of poly(silsesquiazane)s. However, similar to poly(silsesquioxane)s, they are obtained as a mixture of molecules with cage structures as well as oligomeric and highly branched polymeric products.

The structure of a hexameric silesquiazane \((\text{H}_2\text{C})_6\text{Si}_6\text{(NH)}_6\), which was obtained from sodium amide and \((\text{H}_2\text{C})\text{SiCl}_3\) in \(n\)-hexane solution at \(-78 \, ^\circ\text{C}\), was reported by Räke et al. Sodium amide was used instead of ammonia and thus formation of ammonium chloride precipitate avoided. Single crystal X-ray analysis displayed a cage-type structure of \((\text{H}_2\text{C})_6\text{Si}_6\text{(NH)}_6\) in which two six-membered ring systems \([\text{(H}_2\text{C})\text{Si(NH)}]_3\) with chair conformation are linked via the silicon centers with each one NH unit [67].

Polysilazanes are also accessible by transition metal-catalyzed dehydrocoupling of silanes and ammonia or amines. In contrast to the above-described ammonolysis and aminolysis of chlorosilanes, no solid byproducts are formed in the dehydrocoupling reactions; the only byproduct is hydrogen, which evaporates during the reaction. In this regard, it has to be mentioned that dehydrocoupling of silanes and ammonia allows for the preparation of highly crosslinked polysilazanes that due to their comparably low solubility are difficult to be obtained using the conventional salt elimination process.

Laine and Blum found that ruthenium catalysts such as \(\text{Ru}_3(\text{CO})_{12}\) are most suitable. They reported the first catalytic synthesis of polysilazanes in 1986 [68–71].

\[
\begin{align*}
\text{H-Si-H} + \text{NH}_3 & \xrightarrow{\text{[Ru}_3(\text{CO})_{12]} - \text{H}_2} \left[ \begin{array}{c}
\text{H-Si-NH} \\
\text{R}^1 \\
\text{R}^2 \\
\end{array} \right]_n \\
\text{R}^1, \text{R}^2 &= \text{alkyl}
\end{align*}
\]

An alternative method for the synthesis of polysilazanes, which also avoids the formation of solid byproducts during the polymerization step, relies on the redistribution reactions of chlorosilanes with HMDS [50].

\[
\begin{align*}
\text{Cl-Si-Cl} + \text{Me}_3\text{SiN}=\text{SiMe}_3 & \rightarrow \left[ \begin{array}{c}
\text{Si-NH} \\
\text{R}^1 \\
\text{R}^2 \\
\end{array} \right]_n \\
+ 2 \text{Me}_3\text{SiCl}
\end{align*}
\]

The reaction can be performed in solution or without solvent using excess HMDS. Both HMDS and the couple product trimethylchlorosilane are volatile liquids, which can easily be removed from the crude reaction mixture by a distillation step under reduced pressure.

As pointed out above, the molecular weight of polysilazanes can be increased by a crosslinking step, which occurs either by trans-amination or dehydrocoupling reactions. In both cases, polymers are obtained with retention of the silazane ring structures.

The conversion of low molecular weight cyclic silazanes, for example, cyclodi-, tri-, or tetrasilazanes, by ROP in contrast results in the loss of cyclic structural motifs. The idea behind this process is to obtain linear high molecular weight polysilazanes with proper rheological properties, as they are required for polymer fiber spinning.

Ring opening of cyclosilazanes can be induced by treatment with catalytic amounts of a strong base such as MeLi, \(n\)-BuLi, or \(t\)-BuLi [73], strongly acidic compounds such as triflic acid methyl ester, \(\text{F}_3\text{CSO}_2\text{CH}_3\) [74–77], or by transition metal catalysts such as \(\text{Ru}_3(\text{CO})_{12}\) [68]. For example, Suom and coworkers investigated in detail \(\text{F}_3\text{CSO}_2\text{CH}_3\)– and organo alkali salt-induced ROP of cyclodisilazanes covering small substituents in dichloromethane solution [78].

\[
\begin{align*}
\text{H}_3\text{C} & \xrightarrow{[\text{F}_3\text{CSO}_2\text{CH}_3]} \left[ \begin{array}{c}
\text{CH}_3 \\
\text{N} \\
\text{R} \\
\end{array} \right]_n \\
\text{R} &= \text{CH}_3, \text{C}_2\text{H}_5, \ldots
\end{align*}
\]

They observed that four-membered ring systems, which have higher tension, react faster and more readily than cyclotrisilazanes, independent of basic or acidic conditions that were applied. Remarkably, polysilazanes with molecular weights of up to 18,000, using acidic conditions and up to 100,000 using basic conditions, could be obtained by this method. Another remarkable issue using ROP is the possibility for synthesizing copolymers when starting from cyclodisilazanes with differently substituted silicon atoms.
2.4.2. Poly(silylcarbodiimide)s

Disilylated carbodiimides are known since the early 1960s and mainly Ebsworth and Mays, as well as Pump and Wannagat investigated the synthesis and reaction behavior of these monomeric compounds [79–84]. However, because of their volatility disilylated carbodiimides cannot be used directly as precursors for ceramics.

In 1964, Pump and Rochow reported for the first time on poly(silylcarbodiimide)s, which were obtained by the reaction of dichlorosilanes with disilver cyanamide [85].

However, for a transformation into a technical scale and the application of the polymers as a preceramic material, this approach is by far too expensive.

A patent of Klebe and Murray in 1968 described a modified procedure for the synthesis of poly(silylcarbodiimide)s [86]. In a trans-silylation reaction starting from bis(triorganosilyl)carbodiimides (usually bis (tri-methylsilyl)carbodiimide) and di- or trichlorosilanes, the precursors could be obtained in simple procedures and high yields.

Besides the simplified processing—no solid byproducts formed in this reaction—tremendously reduced costs for the precursor synthesis were also advantageous.

In the beginning of the 1990s (>20 years after the appearance of the patent of Klebe and Murray!), poly(silylcarbodiimide)s were rediscovered and their applicability as preceramics studied in much detail. The synthetic approach was improved significantly by developing a non-oxide sol–gel process, and the molecular structure of cyclic tetra(dimethylsilylcarbodiimide) using single crystal X-ray diffraction (XRD) determined [12,87–95].

Recently, air-stable linear carbon-rich polysilylcarbodiimides $[R^1R^2Si\equivCN]_n$ with $R^1 = \text{phenyl}$, $R^2 = \text{Ph, Me, H, and vinyl}$ were reported, which were synthesized via the liquid phase reaction of $R^1R^2SiCl_2$ with bis(trimethylsilyl)carbodiimide. The increased carbon content of these precursors induces an increase in the carbon content in the final ceramics, resulting in an improved thermal stability of the derived Si–C–N ceramics and higher crystallization temperatures. Remarkably, Si$_3$N$_4$ crystallization in the resulting ceramics upon annealing at high temperatures is inhibited [96].

Due to the similar electronegativity of the carbodiimide group and oxygen, polysilylcarbodiimides can be regarded as pseudo chalcogenides. Accordingly, polysilylcarbodiimide gels are non-oxide analogous to classical oxide gels, for example, to tetraethoxysilane- (TEOS) Si(OC$_2$H$_5$)$_4$ or tetramethoxysilane-based (TMOS) Si(OCH$_3$)$_4$ systems, with respect to structural, spectroscopic, and rheological properties [89,91,97] (Scheme 11).

---

**SCHEME 11** Nonoxide sol–gel process for the reaction of trichloromethylsilane with bis(trimethylsilyl)carbodiimide (left) and its analog oxide sol–gel type (right) [91]. For color version of this figure, the reader is referred to the online of this book.
Thus, highly crosslinked poly(silylcarbodiimide) can be obtained via a nonoxide Stöber process by reacting a diluted solution of tetrachlorosilane and bis(trimethylsilyl)carbodiimide in the presence of pyridine (pyridine was added to accelerate the trans-amination reaction) according to Eqn (20) [98]. By using this method, the formation of the first crystalline phases in the Si–C–N system upon thermal treatment was observed (Figure 2) [95].

\[
\begin{align*}
\text{Cl} & \text{Si} \text{Cl}^+ + 2 \text{Me}_3\text{SiN}=\text{C}=\text{N}=\text{Me}_3 \rightarrow \text{Tol}, [\text{Py}] \\
\frac{1}{n} \text{Si}(\text{N}=\text{C}=\text{N})_2 & \text{n}
\end{align*}
\]

Figure 2: Ternary Si–C–N phase diagram (T < 1440 °C, 1 atm N₂). A and B: compositions SiC₂N₄ and Si₂CN₄, respectively, the first crystalline ternary phases in the Si–C–N system obtained from poly(silylcarbodiimide) [Si(NCN)₂]ₙ [95].

β-SiC₂N₄ (space group Pn₃m) crystallizes >400 °C. It possesses an anticuprite structure, which can be described as two interpenetrating crystobalite lattices. Above 920 °C, it decomposes with the loss of cyanogen and nitrogen to form Si₂CN₄ (space group Aba₂), which is similar to the crystalline structure of sinoite [95].

2.5. Precursors to Quaternary Si–B–C–N Ceramics

Organometallic polymers containing the elements silicon, boron, carbon, nitrogen, and hydrogen are a comparably young class of preceramic polymers. However, the exceptional high temperature stability of their derived ceramics, which is superior to those noted in the previous sections, coupled with a high resistance toward oxidation, that is, self-healing capabilities under oxidative conditions, has prompted considerable interest in these new materials.

In 1986, Takamizawa et al. first described a method for the preparation of an inorganic fiber containing silicon, carbon, boron, and nitrogen. For this purpose, polydimethylsilane, [(H₃C)₂Si]ₙ, and B-trimethyl-N-triphenylborazine, [B(CH₃)–N(C₆H₅)]ₙ, were mixed and reacted at an elevated temperature. In a similar manner, the authors reacted block-copolymers carrying Si(CH₃)₂ and Si(C₆H₅)₂ units in the polymer backbone with the above-noted borazine derivative [99,100]. Such polymers were spun and rendered infusible by either subsequent curing in oxidative atmosphere or by irradiation with actinic rays. Investigations to determine the molecular structure of the precursors were not performed.

2.5.1. Borazine-based Si–B–C–N Precursors

The chemistry of borazine and its derivatives is an important field in inorganic chemistry. It is thus natural that scientists deal with potential precursors to Si–B–C–N ceramics long before this topic became relevant for materials science.

Such an example is work published by Noeth in the early 1960s [101]. He investigated the reaction of trichloroborane or alkylidichloroborane with (trialkylsilyl)dialkylamines, R₃SiNR₂, and HMDS. It turned out that chloroboranes readily react with R₃SiNR₂ under Si–N bond cleavage and B–N bond formation. For example, (trimethylsilyl)dimethylamine, Me₃Si–NMe₂ reacted with trichloroborane to form dimethylaminodichloroborane and trimethylchlorosilane.

\[
\begin{align*}
\text{Me}_3\text{SiNMe}_2 + \text{BCl}_3 & \rightarrow \text{Me}_2\text{N-BCl}_2 + \text{Me}_3\text{SiCl} \\
3 \text{Me}_3\text{SiNH-SiMe}_3 + 3 \text{RBCl}_2 & \rightarrow 3 \text{Me}_2\text{N-BCl}_2 + 6 \text{Me}_3\text{SiCl}
\end{align*}
\]

Using HMDS, (Me₃Si)₂NH, instead of Me₃Si–NMe₂ and alkylidichloroborane RBCl₂, resulted in the formation of monomeric B-alkylated borazine derivatives.

\[
\begin{align*}
3 \text{Me}_3\text{SiNH-SiMe}_3 + 3 \text{RBCl}_2 & \rightarrow 3 \text{Me}_2\text{N-BCl}_2 + 6 \text{Me}_3\text{SiCl}
\end{align*}
\]

Highly crosslinked borazine-based precursors to Si–B–C–N ceramics with a comparably low silicon content were synthesized by Narula et al. starting from B-chloroborazine derivatives and HMDS [102].

\[
\begin{align*}
\text{ClBi} & \rightarrow \text{Me}_3\text{SiNMe}_2 \rightarrow \text{Me}_3\text{SiNMe}_2 \rightarrow \text{Me}_3\text{SiNMe}_2 \rightarrow \text{Me}_3\text{SiNMe}_2 \\
\end{align*}
\]

Even though not evident from the molecular structure depicted above, silicon and carbon are present in the oligomer-capping groups.
Seyferth and Plenio published further borazine-based polymers in 1990. They reported on the synthesis of “borasilazanes” as polymeric precursors for borosilicon nitride [103,104].

Initially, borane undergoes addition to the cyclic silazane with the formation of a silazane–borane adduct (not shown) and elimination of dimethyl sulfide. Subsequent dehydrocoupling (intermediate I) and trimerization result in the formation of a borazane (proposed intermediate II in Scheme 12) in which silazane rings are maintained. A shift of a boron-bonded hydride to silicon, accompanied by an ROP reaction of the silazane motif finally releases the precursor.

Blum and Laine independently published a similar procedure starting from poly-N-methylsilazane, \([\text{SiH}_2-N(\text{CH}_3)]_n\), and ammonia or trimethylamine borane complexes \(\text{BH}_3+\text{NH}_3\) or \(\text{BH}_3+N(\text{CH}_3)_3\), respectively, in the presence of catalytic amounts of \(\text{Ru}_3(\text{CO})_{12}\) [105,106]. They obtained viscous products, which were not characterized by spectroscopic methods in more detail. Ceramic yields were in the range of 50%.

A series of \(\text{Si-B-C-N}\) precursors, which were obtained by dehydrogenative, and/or dehydrosilylative coupling reactions of monomeric or oligomeric silazanes with borazine as the boron source, were published by Sneddon et al. [107–113].

For example, hydridopolysilazane (HPZ) was reported to react smoothly with liquid borazine in different stoichiometry at temperatures between 50 ºC and 90 ºC to give borazine-substituted polysilazanes in high yields. Spectroscopic investigations indicated that the polysilazane used retained its backbone but was substituted with pendent borazine substituents by means of a borazine–boron to polymer–nitrogen linking, resulting from both hydrogen and trimethylsilane elimination [107,108].

Blum and Laine independently published a similar procedure starting from poly-N-methylsilazane, \([\text{SiH}_2-N(\text{CH}_3)]_n\), and ammonia or trimethylamine borane complexes \(\text{BH}_3+\text{NH}_3\) or \(\text{BH}_3+N(\text{CH}_3)_3\), respectively, in the presence of catalytic amounts of \(\text{Ru}_3(\text{CO})_{12}\) [105,106]. They obtained viscous products, which were not characterized by spectroscopic methods in more detail. Ceramic yields were in the range of 50%.

A series of \(\text{Si-B-C-N}\) precursors, which were obtained by dehydrogenative, and/or dehydrosilylative coupling reactions of monomeric or oligomeric silazanes with borazine as the boron source, were published by Sneddon et al. [107–113].

For example, hydridopolysilazane (HPZ) was reported to react smoothly with liquid borazine in different stoichiometry at temperatures between 50 ºC and 90 ºC to give borazine-substituted polysilazanes in high yields. Spectroscopic investigations indicated that the polysilazane used retained its backbone but was substituted with pendent borazine substituents by means of a borazine–boron to polymer–nitrogen linking, resulting from both hydrogen and trimethylsilane elimination [107,108].

However, molecular weight and distribution studies suggested an inhomogeneous distribution of borazine throughout the obtained polymers.

The evidence for dehydrosilylative coupling reactions was confirmed by reaction of tris(trimethylsilylamino)-silane with borazine [109,110].

\[
\begin{align*}
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\end{align*}
\]

\[
\begin{align*}
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\end{align*}
\]

\[
\begin{align*}
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\end{align*}
\]

\[
\begin{align*}
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\end{align*}
\]

\[
\begin{align*}
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\text{[Si]} & = \text{SiHMe} \\
\end{align*}
\]

SCHEME 12  Synthesis of borazine-based polyborosilazanes via dehydrogenative coupling of borane dimethyl sulfide and cyclic silazanes [103,104].
This finding was traced back to the observation that after vacuum fractionation of the volatile polymerization byproducts 0.25 mol of trimethylsilane per mole of borazine could be isolated. However, polymerization again occurred due to B—H—N—H dehydrocoupling reactions. Moreover, the formation of borazylene motifs was discussed, and the following idealized product structure was suggested [109].

Depending on the silazane to borazine stoichiometry, ceramic yields between 38 and 42% were obtained.

Treatment of borazine with cyclic silazanes such as hexamethyldicycloisilazane also seems to be complex. The main reaction however is a dehydrocoupling of borazine—boron atoms with the nitrogen atoms of the silazanes according to Eqn (27), whereas ring-opening reactions with Si—N bond cleavage, which may also appear is less likely. Thermolysis up to 1400 °C delivers ceramics with 55–62% yield, depending on the borazine to silazane ratio used.

A crucial issue in the synthesis of borazine-based polymers is the lack of selectivity in the polymerization reactions. This fact causes difficulties in controlling the structure and properties of the precursors and the ceramic materials. Especially if a precursor is designed for drawing fibers, latent reactivity must—though it is absolutely required for obtaining high ceramic yields and rendering green fibers infusible—be adjusted very carefully. The pendent borazine units in the above-noted precursors possess both B—H and N—H functions, which upon a thermal treatment may react further and increase molecular weight and viscosity of the polymeric precursors. Consequently, they are not suitable for processes that require stable melt viscosities, as, for example, in the generation of polymer fibers by melt spinning [113].

As a possible approach out of this dilemma, Sneddon et al. designed monofunctional and difunctional borazine derivatives such as B-diethylborazine (DEB) or B-monoethylborazine, which can be prepared by metal-catalyzed hydroboration of ethylene by borazine [111]. In contrast to “pure” borazine, these species have only one or respectively two reactive B—H units. The modification by dehydrocoupling of polysilazanes such as HPZ with DEB resulted in a precursor with pendant borazine units (Scheme 13).

As a side reaction Si—N bond cleaving of HPZ with the formation of DEB—NHSiMe3 was observed. However, the polymeric precursor, which had a sufficient glass transition temperature and an onset of weight loss >200 °C, could be melt spun to Si—B—C—N—H green fibers which were subsequently transformed into ceramic fibers by thermolysis up to 1400 °C.

Borazine was also used in the hydroboration of polyalkylsilanes. Therefore, \( x \) equivalents (referred to the monomer unit; \( x = 1, 2, 3 \)) of poly[1,2-ethynediyl-(methylsilylene)] (PEMS, \([\text{HSi(CH}_3]^\text{C}≡\text{C}]_n\)) were mixed with borazine. The mixture was stepwise heated in a Teflon-coated steel autoclave to 200 °C [114]. A catalytic hydroboration at an ambient temperature as reported by Evans and Fu [115] was not successful.

The polymers were obtained as large red—brown glassy grains. Their color brightened with increasing borazine content. Characterization of the insoluble products was
performed by IR and solid-state NMR spectroscopy as well as elemental analyses. The latter revealed a loss of borazine in the order of 10–20%, which presumably evaporated while manipulating the precursor mixture. The atomic ratios Si:C = 1:3 and B:N = 1:1 of the starting compounds were maintained. Thus, depolymerization or borazine cleavage did not occur.

Hydroboration is expected to appear on C=C bonds of the initial polymer and C=C double bonds resulting after the first hydroboration step. As a side reaction, the C=C unit may also be transformed into an olefinic C=C or aliphatic C–C unit by single or double hydrosilylation, respectively (not shown).

Dehydrocoupling of Si–H with N–H as shown in Eqn (28) is another desirable reaction pathway, which has to be considered.

IR spectra revealed the presence of Si–H ν(Si–H) and borazine or polyborazylene units (ν(N–H), ν(B–H), and ν(B–N)). The relative intensity of borazine absorption bands increased with increasing borazine content of the materials. Absorption bands of C–H vibrations of sp-hybridized carbon atoms (terminal C≡C–H groups) as present in the starting material PEMS were not detected. Likewise, there was no unambiguous evidence for the formation of C=C double bonds.

Solid-state NMR spectroscopy did not give a clear indication about the favored reaction pathway. 1H NMR spectra displayed broad signals without much structural information. 13C NMR spectra of polymers obtained from mixtures with \( \chi = 1, 2 \) exhibited resonances at 114 ppm pointing to a considerable amount of C=C units. Additional broad resonance between 10 and 40 ppm were assigned to carbon atoms with mixed chemical composition (CHSiBC, CH₂SiC, CH₂BC).

The \(^{29}\text{Si}\) NMR resonance signals lie within the range expected for the structural components evolving from the proposed reaction pathways (cf. Eqns 28 and 29).

SCHEME 13 Synthesis of borazine-based Si–B–C–N precursors by dehydrogenative coupling of monofunctional borazine derivatives DEB and PHPS [111].
$^{11}$B NMR spectra were identical for all of the three samples. They exhibited line shapes typical for boron atoms in borazine or polyborazylen with a quadrupolar coupling constant $C_Q = 2.9$ MHz and an asymmetry parameter $\eta = 0$ [116].

In summary, the spectroscopic results indicated that reaction between PEMS and borazine most probably proceeded via single or double addition of borazine to the C≡C bond with the formation of both C=C and/or C–C units. In contrast, there is no hint for depolymerization reactions of the starting compound or ring-opening reactions of borazine.

2.5.2. Si–B–C–N Precursors from Ammonolysis and Aminolysis Reactions of Boron-modified Chlorosilanes

Totally different approaches to Si–B–C–N ceramic precursors, which do not make use of the borazine derivatives and which—with one exception—do not produce borazine motifs during the polymerization, step were published by Baldus and Jansen et al. [117–121] and by Riedel et al. [122–126].

Baldus and Jansen reported on ammonolysis and aminolysis reactions of the single-source precursor trichlorosilyl-amino-dichloroborane (TADB) [127], which they obtained starting from HMDS in a two-step reaction with (i) tetrachlorosilane and (ii) trichloroborane.

![Equation](31)

IR and NMR data of the meltable and soluble polyborosilazane, which was obtained in 80% yield, as well as elemental analysis, suggested the above polymer structure in which $\text{Si}_3(\text{NCH}_3)_3$ six-membered ring systems are connected via HN–B– and N(CH$_3$)B-units. Because of the lack of resonance signals at approximately 30 ppm in the $^{11}$B NMR spectrum, the existence of borazine-like structures was excluded. The as-obtained polymer could be transformed into Si–B–C–N–H green fibers, which upon thermolysis delivered ceramic fibers with extraordinary high temperature properties (cf. Section 5.5).

Accordingly, derivatives Cl$_2$MeSi–NH–BCl$_2$ (methylchlorosilylaminodichloroborane, MADB) and ClMe$_2$Si–NH–BCl$_2$ (dimethylchlorosilylaminodichloroborane, DADB) as novel processable single-source precursors of amorphous Si–B–C–N ceramics have been published [128]. MeSiCl$_3$ and Me$_2$SiCl$_2$ were used as starting compounds instead of SiCl$_4$. Interestingly, DADB crystallizes with two molecules associated via intermolecular B–N coordination, thus forming planar B$_2$N$_2$ rings (Figure 3).

MADB and DADB were each polymerized by both ammonolysis and aminolysis using methylamine, releasing more or less highly crosslinked polyborosilazanes. The degree of polymerization was determined by NMR spectroscopy. The results suggested that methyl groups, either present in the molecular precursor (Si–Me) or the nitrogen component (N–Me) inhibit quantitative condensation of the initially formed Si–NHR (R = H, Me) units. As a consequence, ammonolysis of MADB released a more highly crosslinked precursor compared with that obtained by aminolysis of DADB with H$_2$NMe. The latter was a low-molecular-weight polyborosilazane with a substantial amount of Si–NHMe end groups.

Hydroboration of (H$_2$C═CH)(CH$_3$)SiCl$_2$ with borane Lewis base adducts such as borane dimethyl sulfide, H$_3$B•SM$_2$, results in the formation of tris(methylidichlorosilyl-ethyl)borane [129,130]. Subsequent ammonolysis as reported by Riedel et al. results in Si–B–C–N ceramic precursors, which are composed of silazane chains that are crosslinked via C–B–C bridges.
A crucial subject is the filtration process, which has to be applied for removing the solid coproduct ammonium chloride. This separation can be extremely time intensive, because the precipitate partly appears colloidally distributed within the product mixture. Moreover, ceramic yields are comparably low at around 50%. It will be shown below that structurally similar precursors can be obtained by a simple modification of the reaction coordinate.

Using $\text{H}_2\text{BCl}_2\text{SMe}_2$ or $\text{HBCl}_2\text{SMe}_2$ as hydroboration reagents instead of $\text{H}_3\text{B}\text{SMe}_2$, results in the formation of bis(methyl dichlorosilyl-ethyl)chloroborane and (methyl dichlorosilyl-ethyl)dichloroborane [130], which after ammonolysis, release more highly crosslinked precursors, in which the B:Si ratio increases from 1:2 to 1:1 compared to 1:3 in $\text{BH}_3$-derived precursors. Ceramic yields increase in the same order from 50% over 56% to 76%. Interestingly, polymers derived from the dichloroborane derivative $\text{H}_3\text{SiCl}_2\text{C}_2\text{H}_4\text{BCl}_2\text{SMe}_2$ possess molecular structures that are composed of six- and eight-membered silazane ring systems, Si–B–N heterocycles and borazine units [131, 132].

In contrast to the methyl-substituted derivative, these species provide latent reactivity, which enables further crosslinking during thermolysis. As a result, ceramic yields are >80%. However, a remarkable disadvantage is the decreased solubility of these precursors compared with the methyl derivative. Consequently, the processing is difficult and polymer yields are low. In the case of the highly crosslinked precursor with $R = (\text{NH})_{0.5}$, the polymer yield is even <5%, which is a result of its high crosslinking density along with its low solubility.

Both polymer yields and processing were significantly improved by switching the above reaction sequence, that is, by first performing ammonolysis followed by hydroboration [133–137].
Ammonolysis of vinyl-substituted chlorosilanes \((\text{H}_2\text{C}═\text{CH})\text{Si}(\text{R})\text{Cl}_2\ (\text{R} = \text{H, Cl, CH}_3)\) delivers the respective oligo- or poly(vinylsilazane)s (PVSs), which are colorless liquids. Subsequent hydroboration then releases the polymeric precursors in quantitative yields. Whereas the methyl- and the hydrogen-substituted polymers possess sufficient solubility to remain in solution, the more highly crosslinked precursor \([\text{B}[\text{C}_2\text{H}_4\text{Si}(\text{NH})_{1.5}]_3\] precipitated as a fine-grained solid from the reaction during borane addition. Remarkably, the workup of Si–B–C–N precursors obtained in this way is rather simple and only requires the removal of all volatile components by vacuum evaporation. According to this procedure, boron-modified polysilazanes with different boron contents were synthesized and the high temperature properties of their derived materials evaluated (cf. Section 4.4, Figure 26) [138,139].

2.5.3. Si–B–C–N Precursors from Boron-modified Hydridosilanes

An alternative method for the formation of Si–B–C–N polymers as precursors to ultrahigh temperature ceramics is a dehydrocoupling of ammonia or alkyl amines with tris(hydridosilylethyl)boranes, \(\text{B}[\text{C}_2\text{H}_4\text{Si(\text{CH}_3)}_n\text{H}_n]_3\) \((n = 1–3; \text{C}_2\text{H}_4 = \text{CH}_2\text{CH}_2, \text{CHCH}_3)\) [54,140,141] (Scheme 14).

It was observed that the ammonolysis of \(\text{B}[\text{C}_2\text{H}_4\text{Si(\text{CH}_3)}_3\text{H}_n]_3\) (route A) does not occur directly and requires basic catalysts such as \(n\)-butyl lithium similar to a method described by Seyferth and Wiseman in which potassium hydride was used for the crosslinking of cyclic silazanes [52].

A possible mechanism for this base-catalyzed dehydrocoupling was given [54] (Scheme 14).

---

**Scheme 14** Dehydrocoupling reactions of tris(hydridosilylethyl)boranes with ammonia and methylamine [54,140,141].

**Scheme 15** Proposed mechanism for the formation of boron-modified polysilazanes by base-catalyzed dehydrogenative coupling of tris(hydridosilylethyl)boranes with ammonia [54].
It was suspected that \( n \)-butyl lithium initially deprotonates ammonia with the formation of lithium amide and evaporation of \( n \)-butane (1). The more nucleophilic amide then replaces a silicon-bonded hydride which subsequently deprotonates ammonia with the evolution of molecular hydrogen (2/1). The proposed silyldiamine or -triamine was not stable under the reaction conditions applied, and by fast elimination of ammonia (3), the polymeric precursors were formed.

A precursor system that was designed especially for the production of fiber-reinforced high temperature Si\( _3 \)Be\( _3 \)Ce\( _3 \)N ceramic matrix composites was also published [142–144]. Glass-like polymeric precursors were obtained by a thermally induced hydrosilylation reaction of oligo(vinylsilazane) (OVS) using the above-noted tris(hydridosilylethyl)boranes.

\[
\begin{align*}
3 & \quad \text{Si - N} \quad \text{H} \quad \text{H} \\
& + n \quad \text{H} \quad \text{H} \quad \text{H} \\
& \quad \text{B} \quad \text{N} \quad \text{R} \\
\end{align*}
\]

\[
\begin{align*}
\text{R'} &= \text{C}_2\text{H}_4\text{SiH}_2\text{R} \\
\text{R} &= \text{H, CH}_3
\end{align*}
\]

\[
\text{120-200ºC}
\]

(35)

Remarkably, neither solvents nor catalysts were required for this process nor were byproducts formed. For the crosslinking step, the starting compounds were mixed at room temperature, and the resulting colorless reaction mixture was carefully degassed in a moderate vacuum. During a stepwise heating of the mixtures to 120 °C within 2 h and then to 180–200 °C within 4–5 h, a remarkable viscosity increase was observed, which resulted in the formation of the transparent, colorless, and glass-like solid.

Preliminary experiments have proven the applicability of the silane/silazane mixture for the preparation of fiber-matrix composites in the resin transfer molding (RTM, cf. Chapter 5.6) process [145]. Prerequisites for this process, which the precursor mixture fulfills, are, for example, low viscosity (sufficient fluidity) of the starting compounds to allow for sufficient penetration, adequate vapor pressure, appropriate chemistry (no solvents or byproducts), controllable reactivity during crosslinking, and high thermolysis conversion yields/low shrinkage during thermolysis. For further details, see Section 5.6.

2.5.4. Si–B–C–N Precursors Based on Poly(silylcarbodiimides)

According to the procedure described for the synthesis of poly(silylcarbodiimides) as precursors to ternary Si–C–N ceramics described in Section 2.4.2, their boron-modified counterparts are precursors to Si–B–C–N materials. They are accessible by different methods: first, by the hydrosilylation of vinyl-substituted poly(silylcarbodiimides) with borane dimethyl sulfide. The polymeric starting compounds can be obtained either from the respective vinyl-substituted chlorosilanes with cyanamide, H\( _2 \)N–CN, in the presence of pyridine or from vinyl-substituted chlorosilanes and bis(trimethylsilyl)carbodiimide, Me\( _3 \)Si–N=C=N–SiMe\( _3 \) [146–148].

\[
\begin{align*}
\text{NR} &= \text{C}_2\text{H}_4\text{Si(R)}\text{NCN} \\
\text{R} &= \text{H, CH}_3, (\text{NCN})_{0.5}
\end{align*}
\]

(36)

IR and NMR results pointed out that boron is attached exclusively to the silicon-bonded vinyl units. An addition of boron to the carbodiimide unit, as was reported for alkyl aluminination reactions of monomeric silylcarbodiimides, was not observed [149]. Alternatively, boron-modified poly(silylcarbodiimides) were obtained in a nonoxide sol–gel process involving tris(chlorosilylethyl)boranes, which were reacted with excess bis(trimethylsilyl)carbodiimide [146–148].

\[
\begin{align*}
\text{R'} &= \text{C}_2\text{H}_4\text{Si(R)}\text{Cl}_2 \\
\text{R} &= \text{H, Cl, CH}_3 \\
\end{align*}
\]

\[
\text{exc. BTSC} \quad \text{-(H}_3\text{C})_3\text{SiCl}
\]

\[
\begin{align*}
\text{R'} &= \text{C}_2\text{H}_4\text{Si(R)}\text{NCN} \quad \text{n} \\
\text{R} &= \text{H, (NCN)}_{0.5}, \text{CH}_3 \\
\end{align*}
\]

\[
\text{BTSC} = (\text{H}_3\text{C})_3\text{Si-N=C=N-Si(CH}_3)_3
\]

(37)

Depending on the silicon-bonded substituents R, the viscosity of the solution (sol) increased considerably within 30–60 min (R = Cl, H) or within one day (R = CH\(_3\)). After approximately three days, the as-obtained reaction mixtures,
which were then transformed into a gel, began to separate from excess bis(trimethylsilyl) carbodiimide and the byproduct chlorotrimethylsilane. Purification was preformed by removing the volatile components in a high vacuum.

Remarkably, this synthetic pathway is rather cheap and easy to perform. Excess starting compound as well as the byproduct chlorotrimethylsilane can easily be recycled. The precursors are obtained in quantitative yields and without the use of solvents. Moreover, time-intensive processing steps are not essentially necessary.

Unfortunately, the high temperature properties of ceramics derived from such boron-modified poly(silylcarbodiimides) were not superior to those obtained from boron-free Si—C—N precursors. This was traced back to the chemical composition of the materials, that is, their nitrogen content, which was supposed to be too high. Note that, it was not possible to reduce the nitrogen content in these materials by this process to improve their high temperature performance.

For this reason, a synthetic pathway was developed that allows for the synthesis of boron-modified poly(silylcarbodiimides) with adjustable nitrogen contents while the boron and silicon concentrations were kept almost constant[150,151].

Dehydrocoupling was performed according to Scheme 16 in a one-pot reaction by dissolving the silanes in tetrahydrofuran and subsequent mixing with different amounts of cyanamide. A catalyst as required for the dehydrocoupling of tris(hydridosilylethyl) boranes and ammonia or amines that is described above was not used. To ensure quantitative turnovers, the reaction mixtures were refluxed for 12 h whereby strong hydrogen evolution and precipitation of the precursors occurred.

The above examples show that preparative organometallic chemistry allows for the production of a wide variety of molecular precursors for PDCs. Nevertheless, there is still a need for further development. Future work will focus on precursors that release phase-pure ceramics as well as composites with tunable properties.

2.5.5. Si—B—C—N Precursors: attempts Toward Industrial Synthesis

For any kind of large-scale testing and application, the availability of preceramic polymers in 10–100 kg amounts is a crucial prerequisite. In a first attempt, Bayer AG (Germany) scaled up the laboratory synthesis of Cl3Si(NH)BCl2, TADB, by a factor of 100, demonstrating the principle feasibility of a large-scale production of this precursor. Unfortunately, major problems were faced, in particular contamination by oxygen, degrading the materials' high temperature performance. The contaminations were due to...
the extreme sensitivity of the molecular and polymeric intermediates against moisture and oxygen in combination with the difficulty of generating sufficiently inert production lines.

At Fraunhofer Institute for Silicate Research (Würzburg, Germany), attempts were made, to process CH$_3$Cl$_2$Si(NH)$_2$BCl$_2$, MADB, and the respective polymer batchwise, rather analogous to the laboratory synthesis and the Bayer process [152]. However, similar problems appeared, and the fussy purification and inertization of the apparatus prevented a breakthrough in the industrial synthesis.

Synthetic routes that allow for continuous processing are clearly advantageous. They allow the use of miniaturized plants, have facilitated maintenance, and usually exhibit high conversion rates. In this regard, Weinmann et al. published the first approach toward a non-batchwise synthesis of a single-source precursor to Si–B–C–N ceramics. Starting from SiCl$_4$, MeNH$_2$, and BCl$_3$, Cl$_3$SiNMeBCl$_2$ dichloroborylmethyltrichlorosilylamine (DMTA) was obtained in the gas phase in a two-step reaction involving salt elimination reactions via Cl$_3$SiNMeH, trichlorosilylmethyamine (TSMA) as an intermediate [153]. Remarkably, conventional batchwise synthesis is not suitable to produce substantial amounts of DMTA.

A scheme and the simplified flow diagram of the continuous process are shown in Figures 4 and 5, respectively. A photograph of the laboratory setup is provided in Figure 6.

A constant flow of MeNH$_2$ gas diluted with molecular nitrogen and excess SiCl$_4$ vapor are injected simultaneously

![FIGURE 4](image1.png)

**FIGURE 4** Scheme of the continuous two-step synthesis of Cl$_3$SiNMeBCl$_2$, DMTA. For color version of this figure, the reader is referred to the online version of this book.

![FIGURE 5](image2.png)

**FIGURE 5** Flow diagram of the reactor for the two-step synthesis of Cl$_3$SiNMeBCl$_2$, DMTA. Straight lines indicate mass flow during the production cycle, whereas dotted lines indicate mass flow during the regeneration process. Scheme taken from Ref. [153].
into the first reactor that operates at approximately 100 °C. Immediate formation of TSMA (gas) and MeNH3Cl (solid) occurs. The hydrochloride precipitates and deposits on the walls of the reactor. N2, TSMA, and excess SiCl4 are completely separated from MeNH3Cl by filtration through a silicon carbide filter and subsequently exposed to a stream of BCl3/N2. Spectroscopically pure DMTA and excess starting compounds SiCl4 and BCl3 are separated from the solid couple product TSMA hydrochloride by filtration. Excess SiCl4 and BCl3 are subsequently removed from DMTA by evaporation under reduced pressure.

The ongoing deposition of hydrochlorides in the course of the production decreases the performance of the reactor and finally clogs the filters. Therefore, precursor synthesis based on a cyclic mode, consisting of sequentially occurring production and regeneration steps, is required. Production is performed until the partial clogging of the filters causes a pressure increase within the system, which is monitored online. The supply of reactants is then interrupted and the system subsequently regenerated by vaporizing the amine hydrochlorides in a high vacuum. Fast and quantitative sublimation is enabled by inductively heating the ceramic filters.

The maximum production time of the initial setup was 25–30 min. Quantitative regeneration was achieved within another 60 min. Accordingly, a full cycle took up to 2 h. The maximum flow rate for the actual laboratory setup (diameter of the reactors was ~3.0 cm) was 4.9 mmol (0.15 g) of MeNH2 per minute, corresponding to a daily synthesis of approximately 440 mmol (108 g) of DMTA.

An even more straightforward process is the continuous salt-free two-step synthesis of TADB (cf. Eqn (30)) by silazane cleavage, starting from HMDS, SiCl4, and BCl3 [154].

The reaction is performed in an array of two subsequent high surface area columns filled with Raschig rings. A scheme of the first step, that is, synthesis of N,N,N-trichloro-N’,N’,N’-trimethylidisilazane (TTDS) (Cl3SiNH-SiMe3) is given in Figure 7. Pressure and temperature are chosen in a way that HMDS and TTDS are liquid and SiCl4 as well as the couple product MeSiCl3 are in the gas phase. Typical reaction conditions were 45 °C/160 mbar. A stream of SiCl4 vapor is passed through the system from the
bottom to the top. HMDS is simultaneously introduced on the top of the column. Gravity causes the liquid to flow downward thereby wetting the surface of the Raschig rings. MeSiCl₃ elimination and conversion of HMDS to TTDS occur on the surface of the glass rings, causing a concentration gradient of the two liquids within the reactor. The upper part mostly contains HMDS, whereas—on choosing proper reaction conditions—the lower part only contains TTDS. The latter is collected in a vessel attached to the bottom of the system. Excess SiCl₄ vapor and the vapor of the couple product MeSiCl₃ are collected in a cooling trap attached to the upper part of the reactor.

The second reaction step is performed accordingly. BCl₃ gas is introduced into a column filled with Teflon rings and liquid TTDS given into the column in the reverse flow. The product Cl₃SiNHBCl₂, TADB, is finally purified by distillation under reduced pressure. However, the determination of proper reaction conditions is of major importance, since otherwise, formation of trichloroborazine as a side product occurs, which precipitates and clogs the column.

According to US 20110028302 A1 [154], TADB is mixed with HMDS, and the mixture is polymerized in a salt-free reaction by a controlled stepwise heat treatment to finally 200 °C. Evaporation of MeSiCl₃ and completion of the condensation are supported by simultaneously reducing the pressure to finally 10 mbar.

2.6. Precursors to Ternary Si—O—C and Quaternary Si—M—O—C Ceramics (M = B, Al, Ti, Zr, Hf, etc.)

Poly(organosiloxanes) are preceramic polymers that can be used for the synthesis of silicon oxy carbide-based ceramics on thermal decomposition in inert gas atmosphere. They are generally denoted as silicones, are usually inexpensive and exhibit unique chemical, physical, and electrical properties [155]. The general synthesis method for the preparation of polysiloxanes relies on the reaction of chloro(organo)silanes with water (Scheme 17a).

(a)

(b)
In recent years, novel silicon-rich poly(organosiloxanes), named polysilaethers [156], have been synthesized by the polycondensation of \( z,m \)-functionalized linear silanes [156] or through the ROP of cyclic siloxanes [157] (Scheme 17b).

Furthermore, highly branched poly(organosilsesquioxanes) \( [\text{RSiO}_{1.5}]_n \) have been used as preceramic polymers for the synthesis of \( \text{Si}–\text{O}–\text{C} \)-based ceramics, since their ceramic yield upon pyrolysis is very high due to their increased crosslinking degree.

Crosslinked polysiloxanes or silicone resins can also be prepared by the sol–gel process via hydrolysis and condensation reactions of hybrid silicon alkoxides. This class of precursors was used for the first time to synthesize silicon oxycarbide glasses [158–160]. They are modified silicon alkoxides of the general formula: \( \text{R}_x \text{Si}(\text{OR})_{4-x} \), \( \text{R} = \text{alkyl, allyl, aryl; R'} = \text{methyl, ethyl} \), which on gelation convert into silicone resins of the composition \( \text{R}_x \text{SiO}_{(4-x)/2} \). The sol–gel process allows for precisely controlling the composition of the silicone resin, because different hybrid silicon alkoxides can be used for cohydrolysis and subsequent polycondensation. Thus, precursors for stoichiometric silicon oxycarbide, as well as for \( \text{Si}–\text{O}–\text{C} \) materials, showing excess of carbon or silicon have been prepared [161]. Moreover, this preparative technique allows for introducing additional elements within the preceramic network, for example, Al, Ti, B, by using their corresponding metal alkoxides [162].

Recently, the modification of polysilsesquioxanes with transition metal alkoxides (such as zirconium(IV)-propoxide or hafnium(IV)butoxide) via sol–gel-like condensation reactions has been reported (Scheme 18), which leads to preceramic precursors for quaternary \( \text{Si}–\text{M}–\text{O}–\text{C} \) ceramics (M being e.g. Zr, Hf) [163,164]. Thus, a hydroxy- and ethoxy-substituted polysilsesquioxane has been reacted with the alkoxides and subsequently crosslinked and pyrolyzed to furnish \( \text{SiZrOC} \) and \( \text{SiHfOC} \) ceramics. The modification of the polysilsesquioxane with the metal alkoxides has been shown to strongly increase the crosslinking degree of the preceramic precursor, which consequently leads to a strong decrease of the open porosity of the resulting ceramics [164].

3. POLYMER-TO-CERAMIC TRANSFORMATION

The thermally induced degradation of precursors is a rather complex process, which involves numerous chemical reactions in the solid state as well as in the vapor phase. Due to limitations of monitoring such reactions, there is still a lack of understanding of thermolysis mechanisms in detail. However, because thermolysis significantly influences both the composition and microstructure of PDCs, an understanding of the single steps that occur during the heat treatment is of considerable interest. Accordingly, thermolysis of different types of precursors has been studied [51,55,165–182] and reviewed [12,58,183] extensively.

As noted in previous sections, preceramic polymers are composed of a polymer backbone to which either single atoms or molecular units are attached. These substituents influence physical–chemical properties, such as solubility, viscosity, and softening or melting, thereby directly determining the processability of the precursor. With respect to thermolysis, it is important that pendant groups may also provide functionality and as a consequence latent reactivity. The latter is known to be a key issue for obtaining high polymer-to-ceramic conversion yields (high “ceramic yields”), because it provides a means for further crosslinking, thus inhibiting polymer backbone degradation and volatilization of low-weight molecular backbone components. In contrast, pendant, nonfunctional groups are eliminated as gaseous byproducts, which do not only induce a mass loss upon the thermal treatment, but are also responsible for retention of impurities or creation of gas-generated pores and cracks. As a logical consequence, molecular precursors for ceramics should be designed in such a way that exclusively molecular hydrogen is split off during the heat treatment, which would provide the benefit of an easy outgassing because of the high mobility of hydrogen in any ceramic material. Unfortunately, high ceramic yields and good processability are often contradictory. Therefore, compromises are necessary in the precursor design, usually mandating processing with “low-ceramic-yield” precursors in order to obtain higher quality ceramic products [183].

A relevant aspect in precursor thermolysis, which until now limits its applicability to the formation of ceramics in

![SCHEME 18 Chemical modification of a polysiloxane with hafniumtetra(\(n\)-butoxide).](image-url)
small dimensions (thin films, fibers), is the change in density and volume with increasing temperature. Even if ceramic yields are high, dramatic volume shrinkage occurs during thermolysis, which is due to the increase in density. Whereas the density of polymeric precursors is typically in the range of <1 g/cm³ the respective values for fully thermolyzed amorphous ceramics can be at around 2.3—2.8 g/cm³. This means that the increasing density of a specimen during thermolysis causes volume shrinkage to 45% of the initial value or lower. Considering the loss of gaseous byproducts during thermolysis, the volume shrinkage is even higher. For a precursor with a density of 1 g/cm³, which delivers ceramics with a density of 2.3 g/cm³ in 60% yield, the shrinkage is almost 75%. Considering an increase in density up to >3.0 g/cm³ due to the crystallization of the amorphous material at elevated temperature (cf. Section 5.2), the whole situation worsens further. Near net shape manufacturing of PDCs is thus difficult to perform and mostly requires either passive or reactive filler materials to inhibit shape distortion [184,185].

The complexity of the thermolysis process requires the combination of contrasting investigation tools such as thermal and elemental analysis as well as spectroscopic and diffraction methods. Common techniques applied are, for example, simultaneous thermogravimetric analysis (TGA) in which mass changes during the heat treatment are monitored combined with differential thermal analysis (DTA) to receive valuable information on energetic changes during the polymer-to-ceramic conversion. Elemental analysis of the thermolysis intermediates is essential, because it is the only method, which allows for a highly accurate examination of the compositional changes of the precursors that occur during the polymer-to-ceramic conversion. Structural characterization of thermolysis intermediates provides distinct information on chemical reactions that occur during ceramization. Solid-state MAS NMR (in the case of silicon containing precursors especially ²⁹Si-NMR [186]) and FTIR spectroscopy are state of the art. Both are nondestructive methods. NMR spectroscopy represents a tool for probing the local environment of single NMR-active nuclei in amorphous insoluble materials. IR spectroscopy additionally allows for a rapid determination of characteristic vibrational frequencies of structural units, that is, stretching vibrations of O—H, N—H, C—H, Si—H, C≡N, C=C, N≡C≡N, etc. or vibrations caused by deformation of three-atom arrays in the precursor or the thermolysis intermediates. A helpful, supplementary method in this regard is Raman spectroscopy.

### 3.1. Thermolysis of Precursors to Silicon Carbide

Among numerous precursors, which are described in the literature for the formation of silicon carbide ceramics, there is a small number of polymers to silicon carbide that have reached commercial interest. In this regard, the most important systems are poly(dimethylsilane), which is historically seen as the first commercial precursor to SiC/C composites and poly(methylsilane), which is a precursor of near-stoichiometric silicon carbide.

As noted in Section 2.2.2 poly(dimethylsilane), [Si(Me)₂]ₙ, can be thermally rearranged to yield poly(methyl-silylenemethylene), [(H)Si(Me)—CH₂]ₙ. In contrast to [Si(Me)₂]ₙ, the latter is a soluble and meltable polymer. [(H)Si(Me)—CH₂]ₙ was first applied in 1975 to produce SiC-containing fibers in the so-called Yajima process [26,187]. The procedure is schematically presented in Figure 50 in Section 5.5.

Poly(methyl-silylenemethylene) has a molecular weight MW = 1200. It can be melt spun and delivers green fibers with poor tensile strengths of around 5—10 MPa. Even though the precursor is obtained in a rather simple procedure and in high yields, it suffers from the lack of latent reactivity, that is, the ability to thermally crosslink after the spinning process. As a consequence, fiber integrity is lost during thermolysis. To maintain fiber integrity, it is thus necessary to subsequently cure the green fibers. In this step, the fiber surface is chemically modified and the fiber rendered infusible. Curing can be performed by different procedures: according to Scheme 19, heating of the fibers in air results in a partial oxidation of surface Si—H (rather than C—H) groups with formation of Si—OH units. A subsequent condensation step produces a thin Si—C—O film on the surface of the fibers [183].

This silicon oxyxide film efficiently inhibits melting of the fiber during thermolysis. Oxygen, which is introduced in this step, cannot be removed during the heat treatment and remains in the fiber, limiting the maximum application temperature to approximately 1150 °C. At higher temperature decomposition with the formation of volatile species such as CO and SiO as well as crystallization occurs, leaving pores and large crystallites behind, which substantially decrease the mechanical properties. Alternatively, electron beam irradiation can serve for fiber curing. Compared to green fiber curing in air, it is much more expensive and more difficult to perform, but no oxygen is introduced into the green fiber. Consequently, thermolysis delivers oxygen-free SiC/C fibers with enhanced mechanical properties and higher maximum application temperature.

Thermolysis of the green fibers has to be performed in an inert atmosphere. Yajima et al. reported on thermolysis of Nicalon™ green fibers in a vacuum at finally 1500 °C to obtain 7-μm-thick SiC fibers, which exhibit a density of approximately 2.5 g/cm³, tensile strengths of 1.5—3 GPa, and a Youngs modulus of 170—200 GPa, depending on the processing applied [188—190]. In contrast, Hi-Nicalon™ fibers exhibit a density of 2.74—3.1 g/cm³, tensile strengths...
of 2.8 GPa, and elastic moduli of 270–420 GPa, depending on their (poly-) crystallinity. Moreover, their maximum application temperature is >1400 °C compared to 1150 °C for that of Nicalon™ fibers. For further details, see Section 5.5.

TGA of the thermolysis step, which released ceramics in 60% yield, indicated that the thermal decomposition of the polymer proceeds gradually >300 °C and is completed at 800 °C. The pyrolytic degradation of the polymer does not take place “stoichiometrically” by elimination of methane and hydrogen according to

$$\frac{1}{n} \left[ \text{CH}_3 \text{Si}-\text{CH}_2 \right]_n \xrightarrow{\Delta T} \text{SiC} + \text{H}_2 + \text{CH}_4 \quad (39)$$

since almost 50% of carbon, which is present in the precursor as methyl groups, is retained in the ceramic fibers. Consequently, SiC/C composite fibers without defined microstructures rather than phase-pure SiC fibers are obtained.

To yield phase-pure SiC, alternative improved precursors were developed. The idea behind these concepts was to design polymers in which the final elemental composition of the ceramic, that is, the 1:1 stoichiometry of the constituting elements silicon and carbon is already pre-formed. Moreover, the synthesis of precursors was aspired, which does not produce gaseous byproducts upon thermolysis other than hydrogen. Polymers, which (at least theoretically) fulfill these requirements, are polymethylsilane, [SiMeH]_n [24], and polysilaethylene, [H_2SiCH_2]_n [27,29], which can be obtained according to procedures described in Section 2.1. Although the precursors exhibit a 1:1 Si:C ratio, it is difficult to obtain stoichiometric or near-stoichiometric SiC. The reasons are (i) that the Kumada rearrangement is a radical reaction, which takes place at elevated temperature with low selectivity and (ii) that depolymerization of the precursors during thermolysis occurs with the volatilization of carbon and/or silicon species, which cause changes in the chemical composition of the precursors. IR and NMR investigations of the thermolysis intermediates provide information on chemical reactions that occur during the heat treatment. In the case of [Si(Me)(H)CH_2]_n, at least four different steps can be distinguished. Below 400 °C, low weight oligomers evaporate. Up to 550 °C, molecular weight increases due to the formation of new SiC bonds that can be traced back to a dehydrocoupling of Si–H and C–H units [166,191]. In contrast, there is no evidence for the formation of Si–Si linking. A further heating to 800 °C results in the degradation of residual Si–H, CH_3, and Si–CH_2–Si units with evaporation of low molecular weight organic species, that is, molecular hydrogen and methane. This stage corresponds to the organic-to-ceramic transition and releases an inorganic covalent SiC network. Additionally, graphitic carbon forms, which is evident from 13C MAS NMR investigations [191]. Above 800 °C, only very small weight loss is observed, which is due to the evaporation of residual hydrogen. Heating to temperatures exceeding 1100 °C finally results in the crystallization of the predominantly amorphous ceramic. Details on this topic are presented in Section 4.
3.2. Thermolysis of Precursors to Silicon Carbide/Nitride Ceramics

Due to the structural diversity of Si–C–N precursors compared to SiC polymers, thermolysis seems to be much more complex. Although for SiC usually polysilanes or PCSs serve as preceramics, polysilazanes or poly(silylcarbodiimide)s are mostly used as precursors for ternary Si–C–N ceramics.

The ternary phase diagram in Figure 8 distinguishes different types of compositions in the Si–C–N system. Poly(silesquiazanes) [192] or poly(silylcarbodiimides) [13,89,91,92], which are both comparably rich in nitrogen, deliver ceramics with compositions located on the tie-line C–Si,N₄ (A). In contrast, ceramics derived from polysilazanes with the general structure [(R¹)(R²)Si–NH–]ₙ (R¹, R² = single-bonded organic unit) [41,51,55,56,167,170,193–197] have compositions located in the three-phase field SiC/Si₃N₄/C (B). Laine et al. published an exception from this general rule. They obtained Si₃N₄/C (type A) compositions by thermolysis of [H₂Si–N(CH₃)–]ₙ rather than silicon nitride and silicon carbide-based composites [177]. Compositions located in the three-phase field SiC/Si₃N₄/Si (C) were obtained by co-pyrolysis of PCS and PHPS [198–200].

Even though a comparison of the elemental composition of the precursors and their derived ceramics gives first hints for possible reactions that occur during the heat treatment, it is inevitable to structurally characterize thermolysis intermediates accurately. In general, methods such as NMR and IR spectroscopies, elemental analysis, and (mass-spectra-coupled) TGA are considered for this purpose.

3.2.1. Thermolysis of Polysilazanes

Most of the work in this field concentrated on thermolysis of polysilazanes. It turned out that the molecular architecture of the polymer backbone, the pendent functional groups, and the crosslinking of the precursors significantly influence the polymer-to-ceramic conversion. As mentioned in Section 2.3, ammonolysis of dichlorosilanes usually delivers six- or eight-membered cyclosilazanes, [SiR¹R²–NH]ₙ (n = 3, 4). Because of their low molecular weight, as-obtained polymers are mostly not qualified as precursors for ceramics. In the worst case, thermolysis of such volatile molecules results in a 100% mass loss. Moreover, cyclosilazanes are generally liquids, which have poor rheological properties, disqualifying them as precursors for ceramic fibers or bulk materials.

A procedure for increasing ceramic yields of low-weight polysilazanes is a catalyst-induced crosslinking by dehydrocoupling of Si–H and N–H units. For example, Seyferth and Wiseman described crosslinking reactions of cyclo-tetrasilazane, yielding a highly branched polymer, which is soluble in tetrahydrofuran [52,53]. Details on the procedure including a proposed mechanism are presented in Section 2.3. The increased crosslinking was directly mirrored in increased ceramic yields from <30% to approximately 84%.

The incorporation of reactive sites in oligomeric silazanes, for example, vinyl groups, also results in increased ceramic yields. This is directly reflected by a comparison of ceramic yields of [SiMeH–NH]ₙ (20%) [195] with that of [SiViH–NH]ₙ (Vi = HC≡CH₂, 83%) [56] or [SiMe₂–NH]ₙ (50%) and [SiViMe–NH]ₙ (76%) [55], as determined by TGA.

Detailed investigations on the thermolysis of vinyl-substituted oligosilazanes of the general type [SiViR–NH]ₙ and the influence of the vinyl group in crosslinking reactions were published by Bill et al. (R = Me, (NH)₀.₅) [57,201] and by Mocaer et al. (R = H) [176].

Based on solid-state ‘H, ‘C, and ²⁹Si NMR and FTIR investigations, Bill et al. suggested conclusive thermolysis mechanisms.

For example, crosslinking reactions involving Si–C coupling accompanied by the elimination of hydrogen (Eqn (40)) or methane (Eqn (41)) were considered.

\[
\text{Si} - \text{CH}_3 + \text{H} \rightarrow \text{Si} - \text{CH}_2 - \text{Si} + \text{H}_2
\]  

(40)

\[
\text{Si} - \text{CH}_3 + \text{H}_3\text{C} \rightarrow \text{Si} - \text{CH}_2 - \text{Si} + \text{CH}_4
\]  

(41)

They occur at around 500 °C and most likely follow radical mechanisms. Further increasing the temperature, finally results in the formation of CSi₄ tetrahedra.
Vinyl groups that are attached to the silazane skeleton enable crosslinking by olefin polymerization at low temperatures, usually around 300 °C. It will be discussed below (Figure 9) that $^{13}$C NMR spectroscopy with cross-polarization and magic angle spinning is an ideal tool for monitoring the progress in such reactions. The sp$^2$-carbon atoms of the vinyl groups generate NMR signals at approximately 130–140 ppm. During the crosslinking, they are transformed into sp$^3$-carbon atoms, which generate resonance signals at around 10–30 ppm.

Raising the temperature to higher values results in Si–C bond cleavage and β-hydride shift with the formation of Si–H bonds and hydrogenated sp$^2$-carbon atoms. Under the given conditions, the latter eliminate hydrogen and transform into sp$^2$-carbon.

Si–H units play an important role in that they have reactive sites for hydrosilylation and/or dehydrocoupling. The former additionally requires C≡C or C≡C units in the polymer and mostly takes place below <200–250 °C. Si–N bond formation by dehydrocoupling of Si–H and

![Figure 9](image-url)
N–H units is important, because in combination with the above-discussed C=C polymerization and subsequent Si–C bond cleavage, it sufficiently explains the transformation of Si–C units, present in the precursor, into Si–N (finally SiN₄) units.

\[
\text{Si} + \text{H}_2 \text{N} \rightarrow \text{Si} \text{N} + \text{H}_2 \tag{45}
\]

In the pyrolytic degradation of oligo- or polysilazanes of the type [SiR\(^1\)R\(^2\)–NH]ₙ, N–H bonds are frequently involved in trans-amination reactions. These take place at temperatures around 200–500 °C and result in the formation of NSi₃ sites among gaseous ammonia, which is eliminated.

\[
3\text{Si} \text{N} \rightarrow 2\text{Si} \text{N} + \text{NH}_3 \tag{46}
\]

The loss of ammonia leads to the reduction of the nitrogen content in the ceramic materials. The tendency for trans-amination reactions is especially pronounced for poly(silsesquiazane)s of the general type [SiR(NH)\(^1.5\)]ₙ, because these polymers are rich in nitrogen. Their composition is located in the three-phase field C–Si₃N₄–N (cf. Figure 8), whereas that of the derived ceramic lies on the tie-line C–Si₃N₄, as will be discussed below.

Finally, Si–N coupling with the elimination of methane is often considered. This reaction requires a cleavage of thermodynamically relatively stable Si–C bonds. It thus takes place exclusively at elevated temperature.

\[
\text{Si} \text{CH}_3 + \text{H}_2 \text{N} \rightarrow \text{Si} \text{N} + \text{CH}_4 \tag{47}
\]

In the following, the pyrolytic degradation of two structurally different oligosilazanes, which yield ceramics with different chemical composition, is discussed in detail. For this purpose, poly(methylvinylsilazane) (PMVS), [SiViMe–NH]ₙ, and nitrogen-rich poly(vinylsilsesquiazane) [SiVi(NH)\(^1.5\)]ₙ (PNVS) are considered. Both polymers do not only differ in their chemical composition but also in their crosslinking densities. The first is a linear oligomer (or cyclomer), whereas the second constitutes a highly branched network.

Reliable information on the pyrolysis mechanism is received especially from NMR spectra, which for this intention, have to be analyzed in detail. In the following, such a spectral breakdown is represented as an example for PMVS [202].

The \(^{13}\text{C}\) CP MAS NMR spectrum of as-synthesized PMVS precursor shows NMR signals at around 130 and 140 ppm, which are generated by the olefin carbon atoms and one at 0 ppm, which is assigned to the silicon-bonded methyl group. The appearance of the spectrum does not change after heating the polymer to 200 °C, indicating that in this temperature range, no chemical modifications of the precursor involving carbon atoms take place. On heating the sample to 300 °C, drastic changes are observed. The signals at 130 and 140 ppm significantly decrease in intensity, whereas a broad resonance appears at approximately 25 ppm. This observation points to an olefin polymerization of the pendent vinyl groups in the precursor according to Eqn (43). In Scheme 20, this reaction indicates the onset of the thermal degradation of the polymer. In the \(^{13}\text{C}\) NMR spectrum of the sample heated to 400 °C, the latter signal decreases in intensity, whereas a new broadened signal at around 130 ppm is found. This points to an Si–C bond cleavage and the formation of sp²-carbon atoms

![Scheme 20 Proposed thermolysis intermediates and structural ceramic units that evolve during the polymer-to-ceramic conversion of PMVS, [SiViMe–NH]ₙ [202].](image)
to 500 °C, causes only little changes in the appearance of the spectrum. However, signals appear broader, and the resonance of the sp²-carbon atoms is somewhat more intensive. Remarkably, the signal of the methyl groups still remains unchanged. At 600 °C, again drastic changes are observed. The intensity of the signal at 130 ppm strongly increases, whereas that at 0 ppm significantly broadens. This temperature marks the decomposition of residual organic functions, that is, the silicon-bonded methyl groups. It is accompanied by the loss of hydrogen and/or methane and the formation of Si—C bonds (cf. Eqns (40—42)). Moreover, methane that is split off partly decomposes to graphitic carbon and molecular hydrogen.

Further information on the thermal degradation of PMVS is received from ²⁹Si NMR spectroscopy. The spectrum of the precursor displayed in Figure 9 shows two sharp lines at −14 and −17 ppm, which point to a mixture of cyclic trimers and tetramers. Heating to 200 °C causes no changes in the signal appearance, whereas raising the temperature to 300 °C results in a low field shift of the signals to around −3 ppm. This is in accordance with the transformation of the sp²-hybridized vinyl carbon atoms to sp³ hydrocarbons. At 400 °C, a new signal at −21 ppm evolves. This value is typical for C(sp³)SiN₂H sites and clearly points to a rearrangement with Si—C bond cleavage and Si—H bond formation as is suggested in Eqn (44). However, it is also possible that in this stage, C(sp³)SiN₃ sites form by a dehydrocoupling reaction of Si—H and N—H units. Their resonance appears in a similar chemical shift range as that of C(sp³)SiN₂H sites. At 500 °C, a new signal at −45 ppm emerges, which can unequivocally be assigned to SiN₄ sites, most probably due to the reactions according to Eqns (44) and (45). Above 500 °C, the NMR signals broaden and overlap substantially. However, it can be concluded without doubt that SiC₂N₂, SiCN₃, and SiN₄ sites generate the broad signal observed at and >600 °C.

According to the observations in the ¹³C and ²⁹Si NMR spectra of the pyrolysis intermediates of PMVS and the proposed pyrolysis mechanism in Scheme 20, a ceramic evolves, which is composed of sp²-carbon and an amorphous Si—C—N phase, in which silicon is bonded in SiCₓNₓ (x = 2—4) environments. This is in accordance with the finding of the elemental analysis, which confirms that the elemental composition of PMVS-derived ceramic is located in the three-phase field C—SiC—Si₃N₄ (cf. Figure 8).

The lack of aliphatic silicon-bonded hydrocarbons and the higher crosslinking in PNVS, in contrast, result in a ceramic material with different composition. Chemical reactions that take place during the heat treatment are depicted in Scheme 21. They are basically comparable with those of PMVS. This is reflected by a similar onset of the decomposition temperature, marked by a polymerization of the silicon-bonded vinyl groups at around 300 °C.

During the further heat treatment, the aliphatic hydrocarbon chains are split off with the formation of SiHN₃ and SiN₄ sites among sp²-carbon that segregates. Subsequent dehydrocoupling of Si—H and N—H units as well as transamination reactions result in the formation of a ceramic material, which is composed of sp²-carbon and silicon nitride. This is nicely reflected in both ¹³C and ²⁹Si NMR spectra of the final thermolysis products (not shown here), which give no evidence for any Si—C units [201,202]. Consequently, the compositions of ceramics obtained from PNVS are located on the tie-line C—Si₃N₄, which is verified by elemental analysis.

Recently, amorphous ceramics with compositions along the tie-line SiC—Si₃N₄ could be obtained [62]. For the development of such materials, the above-discussed findings were considered carefully. The synthesis of the precursor is discussed in detail in Section 2.3 (cf. Scheme 10).

In this regard, the main objective was to design a precursor in which the final elemental composition of the ceramic, that is, the Si₃N₄ stoichiometry of the constituting elements is already preformed. It was presumed that the polymer-to-ceramic conversion does not suffer from the elimination of gaseous thermolysis byproducts other than hydrogen.
Therefore, the following idealized thermolysis reactions were considered, taking only hydrogen evolution on heat treatment into account.

- thermolysis of PHPS \([\text{SiH}_2\text{NH}]_n\) (sum formula \(\text{Si}_n\text{N}_4\text{H}_{12}\) for \(n = 4\)) results in a composite \(\text{Si}_3\text{N}_4 + \text{Si}\);
- thermolysis of poly(methylsilazane) \([\text{H}_3\text{CSiH} = \text{NH}]_n\) (sum formula \(\text{Si}_n\text{N}_4\text{C}_n\text{H}_{20}\) for \(n = 4\)) delivers a material with a chemical composition \(\text{Si}_3\text{N}_4 + \text{SiC} + 3\text{C}\).

Accordingly, copolymers composed of \((\text{SiH}_2\text{NH})\) and \((\text{H}_3\text{CSiH} = \text{NH})\) building blocks in a 3:1 ratio such as \([\text{SiH}_2\text{NH}]_3(\text{H}_3\text{CSiH} = \text{NH})]_n\) (sum formula \(\text{Si}_n\text{N}_4\text{C}_n\text{H}_{20}\)) should thus deliver \(\text{Si}_3\text{N}_4 + \text{SiC}\) ceramics without “free” carbon or silicon. Consequently, precursor synthesis was performed by coammonolysis of \(\text{H}_2\text{SiCl}_2\) and \(\text{H}_3\text{CSiHCl}_2\) in a 3:1 ratio at \(-10 \, ^\circ\text{C}\) using tetrahydrofuran as solvent according to Scheme 10.

To avoid depolymerization and volatilization of low molecular species during the polymer-to-ceramic conversion, the as-obtained precursor was further crosslinked by base-catalyzed dehydrocoupling according to a method described earlier by Seyferth et al. using 0.1 mol-% of n-butyl lithium as a catalyst. The theoretical ceramic yield of the crosslinked precursor, presuming only hydrogen evaporation, is around 95%.

The polymer-to-ceramic conversion was monitored by means of TGA—MS as shown in Figure 10.

Thermolysis up to \(1200 \, ^\circ\text{C}\) takes place in a two-step decomposition in the 200 \(\, ^\circ\text{C} \sim 900 \, ^\circ\text{C}\) range. The ceramic yield of 94.5% is the highest value observed for \(\text{Si} - \text{C} - \text{N}\) precursors so far. Thermolysis of the non-cross-linked precursor, in contrast, delivers ceramics in only 78% yield [203]. Accordingly, the high ceramic yield of the crosslinked precursor is mainly a consequence of the high crosslinking density of the polymer combined with sufficient latent reactivity due to the presence of both \(\text{Si-H}\) and \(\text{N-H}\) units. These enable thermally induced dehydrocoupling reactions according to Eqn (46) and efficiently avoid depolymerization and volatilization of low molecular species during the heat treatment. The observed mass loss of 5.5% corresponds to the hydrogen content of the precursor determined by elemental analysis and points to the fact that exclusively hydrogen evaporates upon thermolysis. This assumption was confirmed by the results of the mass spectroscopic investigations. Hydrogen elimination starts at approximately 200 \(\, ^\circ\text{C}\) and continues up to 1000 \(\, ^\circ\text{C}\). Above 1000 \(\, ^\circ\text{C}\), the concentration of hydrogen increases again, parallel with that of a species with \(m/z = 18\). The latter corresponds to water, which in this temperature range is eliminated from the alumina crucible in which the measurement was performed. The amount of gaseous species \(<1000 \, ^\circ\text{C}\) with an \(m/z\) ratio \(>2\), however, is negligible, suggesting that the \(\text{Si}:\text{C}:\text{N}\) ratio is retained during the polymer-to-ceramic transformation. However, a very small amount of ammonia and methane and their corresponding fragments can be detected unequivocally, though in low intensity, at approximately 380 \(\, ^\circ\text{C}\) and approximately 580 \(\, ^\circ\text{C}\), respectively. A precise comparison of the relative intensities of methane and ammonia with that of hydrogen is not yet possible, because the large amount of hydrogen exceeded the detector’s capacity.

### 3.2.2. Thermolysis of Polysilylcarbodiimides

In contrast to the \(\text{Si-N}\) skeleton in polysilazanes, which is retained during the thermal degradation of the polymers, the \(\text{Si-N}=\text{C}=\text{N}\) backbone in poly(silylcarbodiimide) is decomposed quantitatively. Neither findings in NMR nor IR investigations gave a hint for the presence of such structural units in the amorphous ceramics. The only

---

**FIGURE 10** The TGA—MS of crosslinked \([\text{SiH}_2\text{NH}]_3(\text{H}_3\text{CSiH} = \text{NH})]_n\). Fragments ranging from \(m/z = 2\) to \(m/z = 84\) are considered. Heating rate: 5 \(\, ^\circ\text{C/min}\); atmosphere: flowing Ar [62].
exception are two metastable crystalline phases, SiC$_2$N$_4$ and Si$_2$CN$_4$ (cf. Figure 2), which were published by Riedel et al. [95]. As already pointed out in Section 2.4.2, these phases were obtained by a reaction between tetrachlorosilane and bis(trimethylsilyl)carbodiimide and subsequent heating to 400 and 920 °C, respectively.

Basically, there are two different degradation mechanisms that additionally have to be considered in the thermolysis of poly(silylcarbodiimide)s. The first mechanism involves the elimination of acetonitrile, H$_3$C\text{CN}, or HCN from poly(silylcarbodiimide)s exhibiting (NCN)Si\text{CH}_3 or (NCN)Si\text{H} units, respectively.

\[
\text{Si N C N Si} + 3 \text{(CN)}_2 \rightarrow 4 \text{Si N C N Si} + \text{N}_2
\]  

According to the formerly mentioned degradation, Si–N–Si linkages form. In addition, this type of reaction is responsible for the thermal degradation of SiC$_2$N$_4$ to Si$_2$CN$_4$, as already mentioned above (Figure 2 [95]).

Even though structurally different, polysilazanes and poly(silylcarbodiimide)s follow comparable degradation...
routes. As an example, the progress in the thermal decomposition of poly(methylvinylsilylcarbodiimide), [SiViMe−NCN]ₙ (PMVSC), which was investigated by solid state $^{13}$C and $^{29}$Si MAS NMR is discussed (Figure 11, Scheme 22) and compared with that of PMVS [SiViMe−NH]ₙ (see Scheme 20) [182].

Similar to the $^{13}$C CP MAS NMR spectrum of PMVS, PMVSC shows NMR signals at around 135 ppm, which are generated by the vinyl carbon atoms, as well as a line at 0 ppm, which is assigned to the silicon-bonded methyl group. An additional signal at 125 ppm can be attributed to the carbodiimide carbon atom, which vanishes after heating to 200 °C. According to the thermal degradation of PMVS, further heating to 300 °C causes polymerization of the pendent vinyl groups and results in the appearance of a broad resonance at around 20–25 ppm, which can be attributed to aliphatic carbon atoms. The latter disappears after heating to 400 °C. Moreover, the line width of the N=C=N resonance at 125 ppm increases considerably, most probably because of the formation of “free” carbon, which also shows resonance in this chemical shift range. At 500 °C, signals broaden further and the low field signal increases in intensity, indicating further formation of graphite. The absence of C−H resonance signals after heating to 600 °C points to the fact that the thermal degradation of the precursor is almost completed. Similar insights are obtained from $^{29}$Si MAS NMR spectroscopy. Olefin polymerization at temperatures up to 300 °C causes a low field shift of the $^{29}$Si signal from −35 to −20 ppm, which is a consequence of the transformation of SiC(sp³)C(sp³)(NCN)₂ sites into SiC(sp³)₂(NCN)₂ motifs. Further heating causes multiple reactions, resulting in the formation of at least 4 differentiable $^{29}$Si sites. According to Scheme 22, the formation of SiNC₂C(sp³)(NCN) (δ = −32 ppm) and SiN₂C(sp³) (δ = −16 ppm) sites is preferred. Above 500 °C, all low field signals vanish, whereas a broad signal forms that can unequivocally be assigned to SiN₄ sites. In conclusion, a comparison of the reactive intermediates in the thermolysis of [SiViMe−NH]ₙ (PMVS) and [SiViMe−NCN]ₙ (PMVSC) clearly shows that the polymer structure and chemical composition significantly influence the polymer-to-ceramic conversion and therefore the chemical composition of the PDCs. In contrast to PMVS which delivers a ceramic with a composition located in the three-phase field C−SiC−Si₃N₄ (Figure 8), thermolyis of PMVSC results in a material with a composition located on the tie-line C−Si₃N₄.

![Scheme 22](image.png)

**Scheme 22** Proposed thermolysis intermediates and structural ceramic units that evolve during the polymer-to-ceramic conversion of PMVSC [SiViMe−NCN]ₙ [182,202].

### 3.3. Thermolysis of Precursors to Si−B−C−N Ceramics

Until now, not much is published on the thermolysis behavior of precursors for Si−B−C−N ceramics. So far, only precursors of the general structure [B(C₂H₄SiR−X−)₃]ₙ (R = single bonded organic ligand; X = NH, NCN) have been investigated. It was Prof. Dr. Klaus Müller, together with his students, who for the first time in detail examined the conversion of the multinary precursors to quaternary Si−B−C−N ceramics by means of solid-state NMR spectroscopy [54,201,202,204].

The extension of the ternary Si−C−N system with boron results in additional reactions during the thermal degradation of the polymeric precursors. In this regard, rearrangements that involve B−C bond cleavages and B−N bond formations have to be considered especially.

In this regard, two different reactions must be taken into account: B−C bond cleavage and B−N bond formation with (Scheme 23, reaction pathway a) or without (Scheme 23, reaction pathway b) Si−N bond cleavage. In the first case (a), Si−CH(CH₃)−Si motifs form, whereas simultaneous B−C and N−H bond cleavages according to reaction pathway (b) result in the formation of pendant Si−CH₂─CH₃ units.

Such reactions are best monitored by $^{11}$B, $^{15}$N, or $^{29}$Si MAS NMR spectroscopy, whereas $^{13}$C MAS NMR does not allow for a distinct differentiation of individual resonance signals, because of a significant signal overlap of all...
aliphatic carbon atoms. Unfortunately, the natural abundance of $^{15}$N is too low for obtaining NMR spectra with a sufficient signal-to-noise ratio. For performing $^{15}$N NMR spectroscopy, $^{15}$N-enriched samples must therefore be synthesized.

As an example, the single steps in the thermal degradation of boron-containing $[\text{B}(\text{C}_2\text{H}_4\text{SiH}_e\text{NH})_3]_n$, which can be obtained from the above mentioned Poly(hydridovinylsilazane) (PHVS) $[\text{SiViH}--\text{NH}]_n$ in a hydroboration reaction using borane dimethyl sulfide $[136]$, will be discussed on the basis of $^{13}$C, $^{29}$Si (Figure 12) and $^{11}$B MAS NMR spectroscopy (left spectrum in Figure 13). Additionally, the results of $^{15}$N MAS NMR investigations (Figure 13, right spectrum) will be reviewed.

The $^{13}$C NMR spectrum of the polymeric precursor $[\text{B}(\text{C}_2\text{H}_4\text{SiH}--\text{NH})_3]_n$ (Figure 12, bottom left) exhibits two strong signals at 12 and 28 ppm, which refer to B—C$_2$H$_4$—Si groups along with two very weak resonance signals at 124 and 139 ppm. The latter corresponds to remaining C=C units that due to steric effects did not react during the hydroboration of PHVS. After heating to 200 °C, the weak olefin signals observed in the precursor vanished completely. This points to the fact that increasing the temperature enables quantitative hydroboration of the C=C units. Upon heating to 400 °C, the signal at 28 ppm

![SCHEME 23](image)

**SCHEME 23** Possible polymer skeleton rearrangements during the thermolysis of boron-modified polysilazanes $[\text{B}(\text{C}_2\text{H}_4\text{SiH}--\text{NH})_3]_n$ due to B—C bond cleavage: (a) including Si—N bond cleavage and Si—C bond formation; (b) without Si—N bond cleavage.

**FIGURE 12** Experimental $^{13}$C (left) and $^{29}$Si (right) NMR spectra of boron-containing $[\text{B}(\text{C}_2\text{H}_4\text{SiH}--\text{NH})_3]_n$ $[133]$, which was obtained by hydroboration of PHVS $[\text{SiViH}--\text{NH}]_n$ $[56]$, at various stages of the thermolysis process. All spectra were obtained under MAS conditions using the CP technique $[202,204]$.

**FIGURE 13** Experimental $^{11}$B (left) and $^{15}$N (right) NMR spectra of $[\text{B}(\text{C}_2\text{H}_4\text{SiH}--\text{NH})_3]_n$ $[133]$, which were obtained by hydroboration of PHVS $[\text{SiViH}--\text{NH}]_n$ $[56]$, at various stages of the thermolysis process. All spectra were obtained under MAS conditions $[202,204]$.
formation of a mixture of dehydrogenative Si signal centered at 30 ppm, which is caused by additional raising the temperature results in a broad resonance 2124 cm\(^{-1}\) are observed in the IR spectra of the polymeric precursors. The resonance signals in the \(^{11}\)B NMR spectra of the thermolysis intermediates confirm the above-proposed rearrangement (Scheme 23). Further raising the temperature results in a broad resonance signal centered at 30 ppm, which is caused by additional dehydrogenative Si–N coupling, finally resulting in the formation of a mixture of SiC(sp\(^3\))N\(_2\) and SiC(sp\(^3\))N\(_3\) sites. FTIR spectra of the thermolysis intermediates confirm the above assumption. Due to the quantitative dehydrogenative Si–N coupling, N–H and Si–H stretching vibrations, which are observed in the IR spectra of the polymeric precursors and which appear as very strong signals at around 3380 and 2124 cm\(^{-1}\), respectively, completely disappear after annealing the samples to \(\geq 600\) °C.

Accordingly, \(^{11}\)B and \(^{15}\)N MAS NMR spectra show pronounced changes in the 400–600 °C range. The experimental \(^{11}\)B NMR spectra shown in Figure 13 refer to the excitation of the central transition (\(m_1 = -\frac{1}{2}\) to \(m_1 = +\frac{1}{2}\)).

Owing to the quadrupolar moment of the boron nucleus, the signals of all \(^{11}\)B NMR spectra are significantly broadened (a second-order broadening, which cannot be eliminated by fast rotation at the magic angle).

The resonance signals in the \(^{11}\)B NMR spectra of the polymeric precursor and the thermolysis intermediates \(< 500\) °C are centered at around 0 ppm. They appear broadened, featureless, and without any fine structure. Their line shapes point to a significant heterogeneity in the local environment of the boron nuclei, presumably due to the presence of a mixture of trigonal and tetrahedral coordination spheres. However, exceeding 500 °C results in a low field shift to 15 ppm and the formation of a signal with distinct special features (quadrupolar coupling constant \(= 2.8–2.9\) MHz), which are typical for trigonal-coordinated boron nuclei. These spectra are almost identical to the spectrum of h-BN. This observation confirms the above-proposed rearrangement (Scheme 23) of the precursor at 500 °C, which is accompanied by a B–C bond cleavage and B–N bond formation.

\(^{15}\)N NMR spectra of \(^{15}\)N-enriched samples are given in the right column of Figure 13. The NMR spectrum of the precursor exhibits a line at around \(-360\) ppm arising from NSi\(_2\)H units, whereas the 400 °C sample exhibits an additional broad signal at \(-245\) ppm. The latter can be attributed to NH\(_2\) and NHSi units, which are generated by the above-discussed rearrangement. After heating the sample to 600 °C, a very broad signal is seen in the \(^{15}\)N NMR spectrum, which can be traced back to a superposition of NH\(_2\), NHSi, NSi\(_3\), and NSi\(_2\)H sites. Finally, after heating to 1050 °C, only one signal remains at around \(-320\) ppm, which clearly reflects the formation of an amorphous ceramic network including NB\(_3\) and NSi\(_3\) structural units.

Remarkably, the ceramic yield of [B(C\(_2\)H\(_4\)SiH\(_3\))\(_n\)]\(_{[133]}\), is \(> 86\%\), which is an exceptionally high value. Figure 14 shows the result of thermogravimetric investigations of the precursor.

Thermal degradation of the precursor occurs in one step and starts at around 280–300 °C (the insignificant mass loss \(< 200\) °C is due to the evaporation of residual solvent, which was used in the synthesis of the polymer). This is exactly the temperature range in which the onset of the B–C bond cleavage followed by the abovementioned rearrangement (cf. Scheme 23) was observed by a solid-state MAS NMR spectroscopy. The mass loss of 15% is due to the volatilization of both methane, which is eliminated from Si–(CH\(_3\))–Si motifs and hydrogen. The latter evaporates due to the elimination from hydrocarbons and dehydrogenative Si–N coupling.

![Figure 14: TGA of [B(C\(_2\)H\(_4\)SiH\(_3\))\(_n\)]\(_{[133]}\), which was obtained by hydroboration of PHVS [SiV\(_2\)H–NH\(_2\)]\(_n\) [56]; heating rate: 2 °C/min; atmosphere: flowing argon [133].]
3.4. Thermolysis of Precursors to Si–O–C and Si–M–O–C Ceramics

In the case of polysiloxanes, crosslinking is achieved via condensation, transition metal-catalyzed addition reactions (e.g., hydrosilylation) as well as free radical initiation techniques. Polymers containing methyl or vinyl groups can be crosslinked thermally by using peroxides [205]. Furthermore, silanol groups can be reacted with a moisture-sensitive silane crosslinker. Silicon hydride functionalities in combination with vinyl groups allow for hydrosilylation reactions and the synthesis of infusible materials, which are resistant toward hydrolysis and thermal degradation (Scheme 24) [206,207]. The crosslinking process of polysiloxanes comprising functional groups, for example, hydroxy or alkoxy groups, can be achieved by the condensation of silanol groups with in situ water release and subsequent hydrolysis reactions of the alkoxy substituents and Si–O–Si bond formation (Scheme 24). Using appropriate catalysts, such as a tetrakis(pentafluorophenyl) borate in the case of a polysiloxanol or [bis(2-ethylhexanoate)tin] in the case of poly(methoxymethylsiloxane) or poly(methylsiloxane), these reactions can occur already at room temperature [208,209].

The pyrolysis of crosslinked polysiloxane resins lead to the formation of amorphous silicon oxycarbide ceramics [210–212]. Mineralization occurs primarily through the evolution of hydrocarbons (essentially CH$_4$) and hydrogen. In parallel, a variety of redistribution reactions between the various bonds Si–O, Si–C, Si–H may occur (Scheme 25) [213], leading to the evolution of silanes (usually between 400 and 600 °C), thus decreasing the final ceramic yield.

At higher temperatures, that is, 600–1000 °C, processes involving extensive C–H, Si–C, and Si–O cleavage leads to ceramic materials consisting of an amorphous silicon oxycarbide phase and residual free carbon.

The structural investigation of the polymer-to-ceramic transformation has been essentially characterized by IR spectroscopy [214], solid-state NMR experiments ($^{29}$Si, $^{13}$C) [215–217], and TGA/MS [216].

The onset of thermal degradation strongly depends on the type of organic groups present in the polysiloxanes. The decomposition of their organic substituents on pyrolysis leads to the formation of a silicon oxycarbide phase consisting of SiC$_x$O$_{4-x}$ ($0 \leq x \leq 4$) sites. Their relative amount

\[
\text{SCHEME 24 Crosslinking processes in polysiloxanes: (a) hydrosilylation reaction (as for vinyl-substituted polyhydridosiloxanes); (b) condensation and hydrolysis reactions.}
\]

\[
\text{SCHEME 25 Structural units in polysiloxanes and possible redistribution reactions thereof during thermal treatment.}
\]
can be extracted from the integration of the various resonance signals, which allows one to exactly determine the composition of the oxycarbide phase. A comparison with the elemental analysis allows for the estimation of the amount of segregated C phase present within the ceramic (so-called free carbon).

\[ ^13 \text{C} \text{MAS-NMR spectroscopy can also be used to follow the polymer-to-ceramic transformation: thus, Si–CH}_3 \text{ groups were found to be stable up to 600 °C, whereas at temperatures around 800 °C, the formation of aromatic C sites can be observed. These findings are in agreement with those obtained by means of TGA coupled with mass spectrometry: the polymer-to-ceramic transformation is accompanied by the evolution of H}_2 \text{ and CH}_4 \text{, which form due to the cleavage of Si–C and C–H bonds.} \]

A large number of polysiloxane compositions have so far been studied [161,218]. It appears that the composition of the final Si–O–C phase (SiO}_2 – C(\text{CH}_3)_2) can be predicted because the O/Si molar ratio remains almost constant during the pyrolysis process. Consequently, the elemental composition of the polysiloxane can be adjusted (C/Si ≈ 1/2) to finally obtain SiCO glasses with minimized contents of free carbon. This approach was applied by Soraru et al., who introduced a proper amount of Si–H groups in the polysiloxane network. The authors modulated the content of free carbon and thus obtained a transparent Si–O–C glasses [161]. Indeed, the concurrent presence of Si–H and Si–CH}_3 \text{ units promotes the formation of Si–CH}_2–Si bonds, and thus the insertion of C into the SiCO framework [218].

The chemical modification of polysiloxanes with the main group and transition metal alkoxides has been shown to be a convenient access to Si–M–O–C ceramics (M = B [219,220], Al [221], V [222], Nb [223], Ta [224], Ti [225], Zr [163,226], Hf [164]).

In the case of a chemical modification of polymethylsilsesquioxane (PMS) with hafnium tetra(\text{butyl-}) oxide, the polymer-to-ceramic transformation occurs differently compared to that of the nonmodified polysilsesquioxane. TGA/MS studies revealed the release of small amounts of water as well as of ethanol and butanol. The release of ethanol was attributed to the condensation reaction of Si–OH and Si–OEt groups (which are present as functional groups in PMS), while butanol is released via the condensation reactions of Si–OH and Hf–O\text{Bu} \text{ groups, thus revealing the chemical modification of PMS via hafnium alkoxide (Scheme 26) [164]. The same behavior was observed also for a zirconium alkoxide-modified polysilsesquioxane [226].}

In both cases, FTIR spectra of the as-prepared alkoxide-modified polysilsesquioxane-based materials revealed the presence of Si–O–M units (M = Hf, Zr; absorption bands at 959 cm}^{-1} \text{ for Hf and 950 cm}^{-1} \text{ for Zr, Figure 15).}

The ceramic yield of the hafnium-alkoxide PMS (ca. 81%) was found to be lower than that of pure PMS (ca. 86%). This can be explained by the high mass loss induced by the condensation reactions and decomposition of Hf(O\text{Bu})_4. Considering its theoretical ceramic yield of 44.7% in combination with a ceramic yield of PMS of 86%, and weight fractions of 10.1% and 89.9%, respectively, a ceramic yield of 81.9% is expected, which is in agreement with the recorded value.

The polymer-to-ceramic transformation of hafnium alkoxide-modified PMS was also investigated by means of MAS NMR techniques. The \text{Si}^{29} \text{NMR spectrum of hafnium alkoxide-modified PMS (Figure 16) was found to be rather similar to that of the pure polymer. The \text{Si}^{29} \text{Si signals expected for the silicon units =Si(\text{OEt})(\text{CH}_3) \text{ and =Si(OH)(CH}_3 \text{) as well as for the terminal =Si–CH}_3 \text{ groups appear at } \sim65 \text{ (denoted by label e), } \sim57 \text{ (d) and } \sim18 \text{ (a) ppm, respectively. The observed experimental chemical shifts are in agreement with the data reported recently for PMS [227].}

The NMR spectra of the Si–Hf–O–C ceramic nano-composites are significantly broadened. At 800 °C, the sample consists of an amorphous SiC}_3O}_7 \text{ phase and the \text{Si}^{29} \text{Si MAS NMR spectrum shows a distribution of silicon sites, namely, SiC}_3O \text{ (6 ppm, (a)), SiC}_2O}_2 \text{ (25 ppm, (b)), SiCO}_3 \text{ (65 ppm, (c)), and SiO}_4 \text{ (106 ppm, (d)). The formation of SiC}_2O}_2 \text{ and SiO}_4 \text{ units was seen as a result of rearrangement processes occurring during pyrolysis [228].}

The \text{Si}^{29} \text{Si NMR signal of tetrahedral SiO}_4 \text{ units (106 ppm) of the 800 °C sample was found to be shifted by approximately 4 ppm to lower field compared with that of pure silicon oxycarbide prepared at the same temperature (110 ppm) [215,229]. This low-field shift was attributed to the presence of Si–OH–Hf bonds within the ceramic backbone. It is generally accepted that modification of silicon oxides by the formation of Si–O–M mixed bonds causes a low-field shift of the respective \text{Si}^{29} \text{Si resonance; furthermore, the low-field shift increases with the number of M atoms in the second coordination sphere of silicon [230]. Downfield shifts up to 10 ppm have been reported, for example, for aluminosilicates [231], zeolites [232], mixed oxides SiO}_2–ZrO}_2 [233,234], and Si–O–C–ZrO}_2 \text{ ceramics [226] prepared by sol–gel techniques.}
An increasing pyrolysis temperature induces a phase separation within the material and the formation of silica- and SiC-rich domains. The SiO$_4$ resonance increases in intensity and dominates the $^{29}$Si NMR spectra of the samples pyrolyzed at 1000 and 1200 °C. In contrast, the SiC$_n$O$_{4-n}$ ($n = 1, 2, \text{or } 3$) resonances decrease considerably in intensity SiCO$_3$ ($-68$ ppm, (a’’)), or disappear completely (SiC$_2$O$_2$ and SiC$_3$O). The spectrum of the 1300 °C sample exhibits a SiO$_4$ resonance with a chemical shift of $-109$ ppm (c’’’). The amount of mixed SiCO$_3$ sites (b’’’) decreases further, while a new signal at $-12$ ppm (a’’’) evolves, corresponding to SiC$_4$ units. This finding indicates the initiation of the phase separation process within the Si–O–C matrix [235,236].

The information obtained by thermal analysis, FTIR and MAS NMR spectroscopy revealed that Hf/PMS up to 800 °C consists of an amorphous SiHfOC phase (Figure 17). Especially the chemical shift of SiO$_4$ centers

**FIGURE 15**  (a) FTIR spectra of PMS, and as-prepared PMS modified with 10 and 30 vol% hafnium alkoxide (arrows indicate the Si–O–Hf absorption band); (b) FTIR spectra of PMS and zirconium alkoxide-modified PMS.
and the presence of a Si–O–Hf band in the FTIR spectrum are strong indications for the presence of a single amorphous phase in the materials. At higher temperatures hafnia begins to precipitate as revealed by a slight high field shift of the SiO₄ signal in the ²⁹Si MAS NMR spectrum. The segregation of hafnia within the SiHfOC amorphous matrix was confirmed by transmission electron microscopy (TEM) studies. At temperatures >1300 °C, additional phase separation of the silicon oxycarbide phase with the formation of amorphous silica and amorphous silicon carbide regions occurs, as indicated by the presence of SiC₄ sites in the ²⁹Si MAS NMR spectra.

4. HIGH TEMPERATURE PROPERTIES

4.1. General Comments

The choice of a material basis for high temperature structure materials suffers from a principal dilemma. Oxide materials show creep affinity at high temperatures, limiting their application in energy facilities and aerospace industry [237]. Nonoxide materials, which possess increased creep stability at high temperatures, are not thermodynamically stable in air and react with oxygen and/or moisture [238–240]. Those that form protective oxide layers either contain oxide type grain boundary phases promoting creep and increased internal oxygen diffusion at higher temperatures. Promising candidates with oxidation resistance and mechanical stability at high temperatures are covalent ceramics, which are prepared without sintering additives. According to the above sections, such materials can be obtained by the thermolysis of suitable organometallic precursors in an inert gas atmosphere [241–243]. Due to low self-diffusion coefficients in covalent ceramics, conventional sintering of powder compacts without

![FIGURE 16](image_url) ²⁹Si MAS NMR spectra of PMS modified with 10 vol% Hf(O₂Bu)₄. The spectra at room temperature (r.t.) and 800 °C were recorded under CP conditions, while the spectra at 1000, 1200, and 1300 °C were recorded by means of SP experiments. Peaks labeled with asterisks are spinning sidebands. The insert is a magnification of the r.t. spectrum.

![FIGURE 17](image_url) Structural evolution of hafnium-alkoxide-modified PMS during polymer-to-ceramic transformation.
additives is not suitable for the production of composites based on covalent ceramics. Multinary nonoxide ceramics are of special interest because these materials combine the properties of the constituting binary phases. In this context, amorphous Si–B–C–N ceramics obtained from organosilicon precursors became of fundamental interest, because these materials recently demonstrated a short-time thermal stability up to 2000 °C without mass loss and only minor crystallization as will be discussed at the end of this section. These attributes, coupled with a high resistance toward oxidation even at elevated temperatures, give rise to extensive studies.

4.2. Thermodynamic Aspects

A major aspect in understanding the high temperature behavior of precursor-derived amorphous ceramic composites is thermodynamics, that is, the consideration of phase reactions of the constituting thermodynamically stable unary, binary, or ternary phases. Solid-state reactions of such phases may in the worst-case result in the decomposition of the ceramic composites.

For example, the stable and metastable solid phases of the quaternary system Si–B–C–N are shown in Figure 18.

Aside from the pure components silicon, boron, carbon, and nitrogen, a number of well-characterized binary phases such as SiC, Si₃N₄, BN, B₄⁺C as well as intermetallic silicon boron phases are known. Besides that, several metastable binary and ternary phases have been revealed in the subsystem. For example, in the binary C–N system, β-C₃N₄ has been predicted by Liu and Cohen [246]. In the system Si–C–N β-SiC₂N₄, which decomposes >900 °C and Si₂CN₄ (see also Sections 2.3 and 3.2), which is stable up to around 1000 °C, have been reported [95]. Crystalline, phase-pure ternary B–N–C phases are to our knowledge not known so far. However, because of the almost similar lattice parameters of h-BN and graphite (both are layered materials with sheets of trigonally coordinated atoms) they build solid solutions. Stoichiometric (but not phase-pure) compositions of BCN shown in Figure 18 have usually been obtained by the reactions of trichloroborane with N-containing hydrocarbons such as acetonitrile, polyacrylonitrile, CVD processes using boranes, ammonia, and hydrocarbons or by thermalysis of borane–pyridine complexes [247–255].

Moreover, several amorphous compositions have been disclosed in as-thermolized PDCs, for example, amorphous carbon, amorphous silicon nitride, and amorphous silicon carbonitride [57,256–260]. A set of consistent thermodynamic data of the stable phases of the quaternary Si–B–C–N system has been developed [244,245,261], which enables a detailed description of phase equilibria.

With respect to possible applications of precursor-derived nonoxide ceramics, Si–C–N is the most important subsystem in the quaternary system. It was stated above that such ternary compositions can be obtained by the thermolysis of polysilazanes or poly(silylcarbodiimide)s at 1000–1400 °C. As-obtained Si–C–N materials are usually amorphous. In other words, they are in a nonequilibrium state, where phase configuration cannot be reproduced by equilibrium calculations. However, spectroscopic methods such as solid-state MAS NMR, and IR spectroscopy (IR) in combination with wide-angle scattering reveal a partial segregation of the amorphous ceramics into structural units that refer to thermodynamically stable phases. At higher temperature, segregation proceeds and the amorphous state is transformed into crystalline composites, consisting of Si₃N₄, SiC, and/or C, depending on the molecular structure and composition of the precursor and the conditions applied during thermalysis.

An important tool for predicting thermally induced decomposition reactions are calculations of phase equilibria by the CalPhaD (Calculation of Phase Diagrams) approach using the set of thermodynamic data mentioned above.

A number of thermodynamic calculations in the Si–C–N system have already been published [244,245,262–265]. It turned out that the decomposition of such ceramics is to a significant extent controlled by the decomposition reactions of silicon nitride either by a carbothermal reduction according to the following.

$$\text{Si}_3\text{N}_4 + 3\text{C} \xrightarrow{1484 ^\circ \text{C}} 3\text{SiC} + 2\text{N}_2 \quad (1013 \text{ mbar N}_2) \quad (50)$$

or by decomposition into the elements:

$$\text{Si}_3\text{N}_4 \xrightarrow{1841 ^\circ \text{C}} 3\text{Si} + 2\text{N}_2 \quad (1013 \text{ mbar N}_2) \quad (51)$$

**FIGURE 18** Si–B–C–N concentration tetrahedron including stable and metastable binary and ternary solid phases in the quaternary Si–B–C–N system [244,245]. Note that the compositions BC₃N are stoichiometric but not crystalline; C₅N₄ was calculated in detail [246] but so far not isolated in crystalline form.
It can be concluded from Figure 19 that the onset of the above decomposition reactions strongly depends on the nitrogen partial pressure. In 1 bar N$_2$ atmosphere, the above reactions take place at 1484 $^\circ$C (1757 K) and 1841 $^\circ$C (2114 K), respectively.

Increasing the nitrogen partial pressure to 10 bar results in a shift of the decomposition temperature due to carbothermal reduction to as much as 1700 $^\circ$C (1973 K), whereas at 1 mbar, one would—consider the results of the CalPhaD computations—expect decomposition already at around 1050 $^\circ$C.

Thermochemistry studies have also been performed on polymer-derived Si–O–C ceramics to assess their thermodynamic stability relative to crystalline phases. Calorimetric measurements of heats of dissolution in a molten oxide solvent indicated that these ceramics possess a negative enthalpy of formation (−15 to −72 kJ/mol) relative to their crystalline constituents (Table 1) [266,267]. Therefore, their crystallization resistance does not only rely on kinetic reasons but also on the thermodynamic characteristics.

In contrast, calorimetric measurements performed on poly(silylcarbodiimide)-based ceramics provided evidence that amorphous carbon-rich Si–C–N PDCs have slightly positive or close to zero heats of formation relative to graphite and the binary crystalline components Si$_3$N$_4$ and SiC (Table 1). Therefore, Si–C–N PDCs are less stable than Si–O–C ceramics. However, amorphous ceramics have a higher entropy than the corresponding crystalline phases, which may stabilize the amorphous PDCs at high temperatures considering the relatively small (positive) enthalpies of formation in Table 1. Therefore, the experimentally observed thermal stability of the Si–C–N ceramics may still reflect at least in part thermodynamic stability.

---

**TABLE 1** Enthalpy of formation from elements as well as from carbon and crystalline binary systems (SiO$_2$ and SiC for Si–O–C; Si$_3$N$_4$ and SiC for Si–C–N) for one Si–C–N– and two Si–O–C-based ceramics pyrolyzed at 1100 $^\circ$C in inert atmosphere [268–271]

<table>
<thead>
<tr>
<th>Sample</th>
<th>Composition</th>
<th>$\Delta H_f$ from elements at 25 $^\circ$C [kJ/mol]</th>
<th>$\Delta H$ From binaries and carbon at 25 $^\circ$C [kJ/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si$_1$O$<em>1$C$</em>{0.73}$</td>
<td>SiO$_2$ [mol%] SiC [mol%] C [mol%]</td>
<td>$-242 \pm 3.6$</td>
<td>$-53.9 \pm 3.6$</td>
</tr>
<tr>
<td></td>
<td>43.10</td>
<td>34.98 21.92</td>
<td></td>
</tr>
<tr>
<td>Si$<em>1$O$</em>{0.72}$C$_{2.10}$</td>
<td>SiO$_2$ [mol%] SiC [mol%] C [mol%]</td>
<td>$-149.9 \pm 4.4$</td>
<td>$-51.9 \pm 4.4$</td>
</tr>
<tr>
<td></td>
<td>14.73</td>
<td>25.97 59.30</td>
<td></td>
</tr>
<tr>
<td>Si$<em>1$C$</em>{2.80}$N$_{1.20}$</td>
<td>Si$_3$N$_4$ [mol%] SiC [mol%] C [mol%]</td>
<td>$-50.6 \pm 6.4$</td>
<td>$-2.8 \pm 6.4$</td>
</tr>
<tr>
<td></td>
<td>14.30</td>
<td>4.80 80.90</td>
<td></td>
</tr>
</tbody>
</table>
In conclusion, the unique thermal stability of the PDCs and their exceptional crystallization resistance relies on (i) kinetic stabilization by the “free” carbon, which suppresses diffusion processes and (ii) thermodynamics that is, negative enthalpy formation of (amorphous) PDCs compared to their crystalline binary counterparts.

4.3. High Temperature Properties of Si–C–N Ceramics

As an example to illustrate the influence of chemical composition on high temperature properties, the decomposition behavior of two different Si–C–N ceramics obtained by the thermolysis of commercially available PVS VT50 [57] and polymer-hydridosilazane NCP200 [59], which possess different chemical compositions, is briefly discussed.

In Figure 20, calculated isothermal sections in the ternary Si–C–N system are shown. Figure 20a shows phase equilibria between 1414 and 1484 °C. In this temperature range, three three-phase fields exist: C/Si3N4/Gas, C/Si3N4/SiC, and Si3N4/SiC/Liquid. Below 1414 °C (melting point of silicon), the latter is composed of Si3N4/SiC/Si. The composition of VT50 ceramic (symbolized by a triangle) is located on the tie-line Si3N4–C. Presuming quantitative segregation into the pure (crystalline) phases, the sample is composed of Si3N4 and graphite. In contrast, the composition of NCP200 ceramic is located in the three phase fields C/Si3N4/SiC. Accordingly, a fully crystalline sample would be composed of SiC, Si3N4 and graphite.

CalPhaD predicts that raising the temperatures to >1484 °C (Figure 20b) initiates the carbothermal reduction of Si3N4 according to Eqn (50). The expected quantitative loss of nitrogen in the case of VT50-derived ceramics should result in the formation of a SiC/C composite. In contrast, thermally induced degradation of ceramics obtained from NCP200 (symbolized by a rhombus) should deliver a composition, which is located on the tie-line Si3N4/SiC. The calculations are confirmed by XRD of annealed ceramic samples. A comparison of XRD patterns of VT50 and NCP200-derived ceramics after annealing at 1800 °C in a nitrogen atmosphere (5 h, 1 bar) is shown in Figure 21.

Whereas in the XRD pattern of annealed VT50 ceramic only β-SiC reflections appear, the pattern of annealed NCP200-derived ceramic exhibits both α-Si3N4 and β-SiC reflections. As expected, an increase of the temperature to 2000 °C (not shown in the figure) results in the dissociation of α-Si3N4 into the elements. After cooling to room temperature, XRD patterns of the respective sample now clearly show α-Si reflections besides those of β-SiC.

The contrasting crystallization of VT50 and NCP200 ceramics is also reflected in their different mass-loss...
behavior. A suitable method for investigating mass loss-accompanied degradation is high temperature TGA.

Figure 22 shows a comparison of the TGA of VT50 (top) and NCP200 ceramics (bottom). It is evident from the TGA curves that VT50 decomposes in a one-step reaction with the loss of 30% of its original sample weight. The determined value exactly fits with the nitrogen content of the ceramic material. In contrast, NCP200-derived ceramics decompose in a two clearly separated steps. The onset of the first endothermic peak, centered at around 1665 °C, is in accordance with the thermodynamic calculations. The corresponding DTA curve displays a strong endothermic peak, centered at around 1665 °C.

In contrast, NCP200-derived ceramics decompose in a two clearly separated steps. The onset of the first decomposition step caused by a carbothermal reduction of Si₃N₄ is at around 1550 °C and thus shifted by 60 °C to higher temperature compared with the result of the calculations. The second step is caused by Si₃N₄ dissociation into the elements and starts at around 1800 °C. The mass loss of the individual decomposition steps is 12 and 15 mass-%, respectively. The total of 27 mass-% exactly matches the nitrogen content of the ceramic material. In addition, the values of the mass-loss of the single steps are purported by the amount of “free carbon.”

It is interesting to note that indeed all as-thermolyzed ceramics with an overall composition within the three-phase field C/Si₃N₄/SiC decompose like VT50 or NCP200-derived ceramics in one- or two-step reactions, only depending on whether their C:Si ratio is >1 or <1, respectively.

In this context, phase fraction diagrams are very suitable to quantitatively predict phase reactions. As an example, Figure 23 shows such a diagram of NCP200-derived ceramic. The total pressure was considered to be 1 bar.

As-pyrolyzed NCP200-derived ceramic is composed of 8 mass-% of graphite, 19 mass-% of SiC, and 73 mass-% of Si₃N₄. Due to the carbothermal reduction of Si₃N₄ at 1484 °C, graphite is quantitatively consumed and bonded as SiC. Correspondingly, it is expected that the amount of Si₃N₄ is reduced to 38 mass-% with the formation of 12 mass-% of gas (N₂). The second decomposition step at 1841 °C consumes residual Si₃N₄. As a consequence, additional 15 mass-% of gas form along with 22% of a liquid (Si). As expected, the amount of SiC is not affected by this reaction.

### 4.4. High Temperature Properties of Si–B–C–N Ceramics

Organometallic polymers containing the elements silicon, boron, carbon, nitrogen, and hydrogen are precursors for silicon nitride, silicon carbide, and boron nitride-based Si–B–C–N ceramics, which possess significantly improved thermal stability, and higher oxidation resistance compared with ternary Si–C–N ceramics [120,124,133].

Results of detailed investigations of the high temperature properties of such materials by high temperature TGA, XRD, and TEM of thermally annealed samples will be shown below. A comparison of the thermal stability by HT-TGA in an Ar atmosphere of an Si–C–N ceramic derived from poly(vinylsilazane) (PVS) and an Si–B–C–N material obtained from a boron-modified polysilazane is shown in Figure 24.

It is clearly seen that PVS-derived ceramics start to decompose below 1500 °C. As has been mentioned, this breakdown is primarily a consequence of the decomposition reaction of Si–N units, which are present in the amorphous material together with “free” carbon causing the formation of silicon carbide and the elimination of nitrogen. Si–B–C–N ceramics obtained from [B(C₂H₄SiH(NH)₃)]ₙ [133], in contrast, resist thermal degradation up to 1950 °C. This example nicely shows that by modification with around 5 mass-% of boron the onset of decomposition in Si–C–N ceramics can be shifted from around 1500 °C by roughly 450 °C to higher temperature.

From the thermodynamic point of view, the unusual high temperature stability of the Si–B–C–N ceramic cannot be understood. According to the calculated phase fraction diagram shown in Figure 25 (composition Si: 24.0, B: 8.0, C: 44.0, N: 24.0 atom%), the Si–B–C–N ceramic should at temperatures <1484 °C be composed of BN (16 mol-%), Si₃N₄ (28 mol-%), SiC (24 mol-%), and 32 atom-% of graphite [133].

In analogy to the abovementioned Si–C–N ceramic, it is expected that in the case of the quaternary material, Si₃N₄ also reacts above 1484 °C with graphite according to Eqn (50). Consequently, the original sample weight should decrease due to the loss of gaseous nitrogen. The amount of free carbon in this case decreases from 32.0 to 20.0 atom-%,
whereas the amount of silicon carbide increases from 24.0 to 48.0 mol-%. However, from the high temperature TGA investigations (cf. Figure 24), it can be assumed that in contrast to the ternary ceramics, the thermodynamically expected decomposition reaction does not take place in the case of the Si–B–C–N ceramic. In addition, boron, which is present as boron nitride, does not thermodynamically directly participate in any decomposition reactions <2000 °C.

Systematic investigations on the connection of chemical composition and thermal stability of precursor-derived Si–B–C–N ceramics reveal that both the boron and nitrogen content significantly influence thermal properties. By stepwise increasing the amount of boron in Si–B–C–N precursor-derived materials, Müller et al. showed that at least 4 weight-% of boron are needed to efficiently avoid thermally induced decomposition [138,139]. For example, a series of Si–B–C–N ceramics obtained from PVS, which was hydroborated with different amounts of borane dimethyl sulfide, is discussed in more detail in the following section.

Table 2 gives the chemical composition of the ceramics derived from the above precursors in weight-%; values in

**FIGURE 22** High temperature TGA/DTA investigations of VT50 (a) and NCP200-derived ceramics (b) in 0.1 MPa N₂. Heating rate: 5 °C/min [269].
Parentheses are atom-%. Phase compositions, which were obtained using the CalPhaD calculations on the basis of the results of the chemical compositions obtained by elemental analysis, are given in Table 3. Oxygen values are <1.5% and neglected.

From the chemical composition of the ceramics, which was determined by elemental analysis, it is obvious that the increasing boron concentration in the precursors is directly reflected in an increasing amount of boron in the ceramic materials. It is the highest in 5c, which was derived from quantitatively hydroborated PVS$_5$ (cf. Scheme 27) and decreases in the row 4c > 3c > 2c. Accordingly, the concentration of Si, C and N increases in that manner.

The results of the thermodynamic calculations in Table 3 point out that with increasing boron content the phase fraction of boron nitride increases from 7.3 mol-% in 2c to 18.2% in boron-rich 5c. Consequently, the amount of nitrogen available for the formation of silicon nitride is

\[
\begin{align*}
\text{Si}_3\text{N}_4 + 3\text{BH}_3 &\rightarrow 3\text{SiC} + 6\text{H}_2, \\
\text{Si}_3\text{N}_4 + 6\text{BH}_3 &\rightarrow 3\text{SiC} + 12\text{H}_2.
\end{align*}
\]

**FIGURE 23** Calculated Si–C–N phase fraction diagram ($p_{tot}$ = 1bar) for NCP200-derived ceramics [265]. A carbothermal reduction of Si$_3$N$_4$ at 1484 °C (1757 K) results in the formation of SiC and N$_2$ (Gas). Decomposition of remaining Si$_3$N$_4$ into the elements takes place at 1841 °C.

**FIGURE 24** Comparison of the thermal behavior of ternary Si–C–N and quaternary Si–B–C–N ceramics [133] by high temperature TGA. Heating rate: 5 °C, 1 bar argon.

**FIGURE 25** Calculated Si–B–C–N phase fraction diagram ($p_{tot}$ = 1bar) for ceramics derived from [B(C$_2$H$_4$SiH–NH)$_3$]$_n$. Chemical composition: Si: 24.0, B: 8.0, C: 44.0, N: 24.0; atom-%. It is expected that due to the carbothermal reduction of Si$_3$N$_4$, decomposition of the ceramic composite occurs at 1484 °C (1757 K), resulting in the formation of SiC and N$_2$ (Gas) [133].

Increasing Boron Content

**SCHEME 27** Synthesis of Si–B–C–N polymers with adjustable boron concentration [138].
reduced. This has two consequences: the silicon nitride phase fraction decreases from 41.6 mol-% in 2c to 31.3 mol-% in 5c, whereas the silicon carbide phase fraction increases from 17.0 to 23.3%. The relative increase of silicon carbide again results in a decrease of the amount of graphite. In conclusion, increasing boron concentration results in increased boron nitride and silicon carbide but decreased silicon nitride and graphite phase fraction and vice versa.

According to Scheme 27 and Tables 2 and 3, the total mass loss due to the carbothermal reduction of silicon nitride increases with increasing phase amounts of silicon nitride in the row 1c > 2c > 3c > 4c > 5c. This conclusion is confirmed by high temperature TGA as shown in Figure 26.
As expected, boron-free 1c starts to decompose at around 1550 °C and loses 24% (Δm₀) of its original sample weight. This is in accordance with the thermodynamically calculated (predicted) value of 25% (Δmₚ) given in Table 3. The incorporation of even small amounts of boron such as in 2c results in a shift of the onset of decomposition by 80 °C to 1630 °C. However, the total mass loss of 19% corresponds to the expected value (Δmₚ = 20%). Further increased boron contents in 3c and 4c result in an additional shift of the decomposition temperature to around 1700 °C. Moreover, a significant deviation of found (3c: Δm₀ = 13%, 4c: Δm₀ = 8%) and calculated (3c: Δmₚ = 18%, 4c: Δmₚ = 17%) mass losses is observed. Compound 5c, which possesses the highest boron content, shows no signs of decomposition in the expected temperature range. Thermal degradation is significantly retarded and starts at around 2000 °C. At this temperature, carbothermal reduction of boron nitride is expected:

$$4BN + C \rightarrow B₄C + 2N₂ \quad (52)$$

In analogy to the carbothermal reduction of silicon nitride (cf. Eqn (50)), the onset of BN decomposition also depends on the nitrogen partial pressure (Figure 27).

The different behavior in the high temperature TGA shown in Figure 26 is mirrored in the crystallization/phase composition of the materials after an additional heat treatment at elevated temperature. For example, after heating to 1800 °C (1 bar argon) for 5 h, XRD patterns of 2c and 3c correspond to the diffraction pattern of silicon carbide (Figure 28). This is in agreement with the high temperature TGA investigation that indicated quantitative (2c) or almost quantitative thermal degradation (3c) of silicon nitride at this temperature.

In the XRD patterns of the boron-richer materials 4c and 5c, β-Si₃N₄ reflections are distinctly detected besides SiC peaks. Their line widths suggest a nanocrystalline arrangement. At 2000 °C, the thermodynamically expected cleavage of Si–N bonds in 4c proceeds, resulting in the absence of Si₃N₄ reflections in the XRD diagram. In contrast, the diffraction pattern of 5c, which was shown by high temperature TGA to be the only thermally stable composite at this temperature, still represents β-Si₃N₄ reflections. In addition, all samples heat treated to 2000 °C exhibit an additional broad reflection at 26.6°, which is a characteristic of graphitic carbon.

Investigations on the influence of the nitrogen content on the thermal stability on Si–B–C–N ceramics were also published [150]. Therefore, boron-modified poly(silylcarbodiimide)s with adjustable nitrogen content were
synthesized by a dehydrocoupling reaction of tris (hydridosilylethyl)boranes, B(C2H4SiRH2)3 (R = H, CH3), and cyanamide (cf. Scheme 16) [151]. Investigations of the polymer-to-ceramic conversion of the precursors in an Ar atmosphere revealed that they deliver Si–B–C–N ceramics in 65–85% yield, depending on the elemental composition and the molecular structure, that is, the crosslinking of the polymers. For example, TGA of a series of polymers obtained by the reaction of B(C2H4SiH3)3 [140] with different amounts of H2N–(hydridosilyl)boranes, B(C2H4SiRH2)3 (R = H, CH3) was performed. TGA revealed that neither silicon-, boron-, nor nitrogen-containing species evaporated during thermolysis. It can thus be concluded that the Si:B:N ratio which was adjusted in the precursors was maintained in the ceramics.

High temperature properties of ceramics derived from H–N1 – H–N6 can clearly be correlated with their nitrogen content.

Ceramics obtained from H–N5 and H–N6, which possess the highest nitrogen content among the materials investigated, decompose earliest at around 1670 °C and finally lose 25% and 30% of their original sample weight. The decomposition temperature of H–N4 is shifted by 50 °C to 1720 °C, the final mass loss of 10% is significantly lower than that of H–N5 and H–N6. Compound H–N2, H–N3, and H–N3.5 behave like 5c ceramics (cf. Figure 26). Neither of these materials decomposes below 1950 °C.

The difference in the thermal mass loss behavior is again reflected in the different crystallization behavior of the ceramic materials. For example, the phase evolution observed by XRD of H–N2 and H–N5, which were annealed under similar conditions, is shown in Figure 31.

As-obtained ceramics (1400 °C) are predominantly amorphous. However, in the patterns of H–N2 ceramics, broadened reflections at 2θ = 36° and 61° point to the presence of nanocrystalline silicon carbide. The appearance

### TABLE 4

<table>
<thead>
<tr>
<th>Silane: cyanamide</th>
<th>Sum formula (monomer unit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H–N1</td>
<td>1:0.5</td>
</tr>
<tr>
<td>H–N2</td>
<td>1:1</td>
</tr>
<tr>
<td>H–N3</td>
<td>1:1.5</td>
</tr>
<tr>
<td>H–N4</td>
<td>1:2</td>
</tr>
<tr>
<td>H–N5</td>
<td>1:2.5</td>
</tr>
<tr>
<td>H–N6</td>
<td>1:3</td>
</tr>
</tbody>
</table>

### FIGURE 29


### TABLE 5

<table>
<thead>
<tr>
<th>Composition of ceramic materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>H–N2 – H–N6</td>
</tr>
<tr>
<td>[mass-%]</td>
</tr>
<tr>
<td>H–N2</td>
</tr>
<tr>
<td>H–N3</td>
</tr>
<tr>
<td>H–N3.5</td>
</tr>
<tr>
<td>H–N4</td>
</tr>
<tr>
<td>H–N5</td>
</tr>
<tr>
<td>H–N6</td>
</tr>
<tr>
<td>[mol-%]</td>
</tr>
<tr>
<td>H–N2</td>
</tr>
<tr>
<td>H–N3</td>
</tr>
<tr>
<td>H–N3.5</td>
</tr>
<tr>
<td>H–N4</td>
</tr>
<tr>
<td>H–N5</td>
</tr>
<tr>
<td>H–N6</td>
</tr>
</tbody>
</table>
of this XRD pattern remains unchanged, even after heating the sample to 1800 °C indicating that there is no further crystallization up to this temperature. After annealing to 1900 °C, reflections of both \( \alpha \)-SiC and \( \beta \)-Si3N4 appear. Remarkably, the latter do not decrease in intensity even after heating the samples to 2000 °C. The crystallization behavior of H–N5 ceramic is significantly different and reflects the decomposition behavior observed by TGA shown in Figure 30. No crystallization of phase formation is observed <1600 °C. Above this temperature, only \( \beta \)-SiC forms, whereas there is no evidence for the crystallization of silicon nitride. This finding again indicates the decomposition of Si–N units due to a carbothermal reduction either before or during crystallization.

The above examples clearly state that, under certain conditions, Si–B–C–N ceramics are thermally much more stable, than thermodynamically predicted. To understand in detail the thermal stability of such ceramics and to find a universal relationship between the thermal stability and the compositions of the materials, different types of Si–B–C–N ceramics were investigated in detail by means of TGA and XRD. Additionally, spectroscopic and microscopic methods were incorporated to receive information on the atomic structure of the amorphous and nano-crystalline assemblies.

Thermodynamic calculations reveal that almost all compositions of Si–B–C–N ceramics investigated so far are located in the four-phase equilibrium field Si3N4 + SiC + C + BN (cf. Figure 18). Thermally stable compositions in this equilibrium field are located close to the three-phase equilibrium Si3N4 + C + BN (Si3N4-enriched area). This fact is illustrated in Figure 32 schematically, showing a calculated isothermal section in the Si–B–C–N system at a constant boron content of 10 atom-%.

Structural analysis of as-obtained amorphous ceramics by means of small and wide angle scattering as well as by FTIR and solid-state NMR spectroscopy reveal a homogeneously ordered atomic array originating from the polymer structure. Remarkably, the structural units of the thermodynamically stable phases are already preformed within the amorphous states on an atomic and a medium range scale. Crystallization of these amorphous ceramics initially yields metastable amorphous phases \( \text{Si}_3\text{C}_y\text{N}_{4-y} \) (\( Y = 0 \rightarrow 4 \)) and BN/C. The former is related to the composition line Si3N4–SiC (Figure 18). As already shown in the X-ray studies of annealed H–N2 and H–N5 in Figure 31,

![FIGURE 30](image-url) Comparison of the thermal behavior of quaternary Si–B–C–N ceramics derived from H–N1 – H–N6 with different nitrogen concentrations by high temperature TGA. Heating rate: \( T < 1200 10 \degree \text{C/min, } T > 1200 2 \degree \text{C/min, 1 bar argon atmosphere. For details concerning the molecular structure of the polymeric precursors and the elemental and phase composition of the derived ceramics, consider Scheme 16 and Table 5, respectively [150].}

![FIGURE 31](image-url) XRD patterns of the ceramics obtained from H–N2 and H–N5 (cf. Scheme 16) annealed at 1400–2000 °C (100 °C steps, 1 bar N2) each for 3 h [150].
Detailed studies on the phase evolution of amorphous Si-B-C-N ceramics by solid-state NMR using MAS were published by Schuhmacher et al. [203,204]. They compared the crystallization behavior of high temperature stable Si-B-C-N ceramic, which was obtained from [B(C2H4SiH3-NH)3]n (MW33 [133]) with that of a less stable material, derived from nitrogen-rich boron-modified poly(silsesquiazane) [B(C2H4Si(NH)1.5)3]n (MW60 [133]).

The representative 29Si MAS NMR spectra given in Figure 33 cover the temperature range from 1050 to 2000 °C. After thermolysis to 1050 °C, 29Si NMR spectra of both ceramics appear almost similarly (the NMR spectrum of the 1050 °C MW60 sample is not shown). They are each characterized by a broad resonance centered at around −40 ppm. Apart from a line narrowing in the spectrum of MW60-derived ceramic, further heat treatment to 1400 °C causes only minor changes in the signal appearance. The chemical shift range in combination with the broadened resonance signals at and below this temperature point to the presence of a mixture of SiCₓN₄₋ₓ sites (x = 1–4), suggesting a fully amorphous structure. After heating to 1600 °C, significant progress in the phase evolution of both materials occurs. In the case of MW33-derived material, two new resonance signals centered at −19 and −49 ppm...
evolve, which indicate the formation of SiC₄ and SiN₄ sites, respectively. However, a broad signal, which appears in between, still points to a remarkable amount of SiC₄N₄−x sites. In contrast, annealing of MW60 ceramic to 1600 °C causes the disappearance of high field shifted signals. The only resonance signal in the ²⁹Si NMR spectrum of MW60 ceramic is found at −19 ppm and suggests quantitative degradation of Si−N motifs due to a carbothermal reduction with the formation of silicon carbide.

Heating MW33 ceramics further results in additional demixing. As a consequence, only two, well-separated signals remain after heating to 1800 °C, which exhibit considerably smaller line widths than those obtained from ceramics annealed at lower temperatures. It was claimed that this is a consequence of the presence of SiC and Si₃N₄ crystallites. Obviously, all structural components with silicon in mixed coordination, such as SiC₄N₄−x sites, are disintegrated at this temperature, as also found by X-ray studies on the same samples.

Even though obtained from structurally different precursors, the crystallization behavior of ceramics obtained from MW33 and MW60 shown in Figure 34 is comparable with that obtained from boron-modified polysilylcarbodiimides depicted in Figure 31. Ceramics with “high” nitrogen contents decompose around 1550–1600 °C with crystallization of silicon carbide, whereas in the case of ceramics with “low” nitrogen content, neither thermal degradation nor crystallization is observed <1750 °C. At and above this temperature, such materials transform into nanocrystalline silicon nitride/silicon carbide composites.

Investigations of the structural evolution of Si−B−C−N ceramics by high-resolution transmission electron microscopy (HR TEM) were also the subject of several publications. For example, Jalowiecki et al. investigated by HR TEM the microstructure of a boron-doped silicon carbide/nitride composite, which was derived from a boron-doped polyhydridomethylsilazane [259]. As-obtained amorphous material was subsequently annealed at 1800 °C for 50 h in an argon atmosphere. HR TEM investigations revealed that the annealed ceramic possessed a quite unusual microstructure. It was segregated into a turbostratic BNC phase that occurred along the grain boundaries of nanosized silicon carbide and silicon nitride crystals as shown in Figure 35.

It was supposed that the BNCx phase inhibits diffusion processes and consequently retards the crystal growth. Moreover, it binds free carbon, thus decreasing the carbon activity. The connection of carbon activity and decomposition temperature of silicon nitride due to the reaction with free carbon, which was calculated thermodynamically [264], is given in Figure 36.

Obviously, a decreasing carbon activity shifts the onset of the carbothermal reduction of silicon nitride according to Eqn (50) to higher temperatures (at 1 bar N₂, this reaction takes place at 1484 °C). Presuming a carbon activity of 0.5 results in a shift by 90 °C, whereas an activity of 0.1 would result in a shift of the decomposition temperature of around 330 °C. Apparently, a further decrease of the carbon activity does not result in a further shift of the onset of Si₃N₄ decomposition to higher temperature since the decomposition of silicon nitride into the elements (at 1848 °C), which is indicated by the horizontal line, becomes the dominant reaction.

A quantitative encapsulation of the silicon nitride grains by the BNCx grain boundary phase presumably prevents the
By combining the single topics discussed above, the unusual high temperature stability of Si–B–C–N ceramics can be understood unequivocally. The BNC grain boundary phase decreases the carbon activity and, due to an encapsulation effect, increases the internal nitrogen partial pressure considerably. Both effects result in a shift of the onset of the carbothermal reduction of silicon nitride to higher values, that is, to 2030 °C, at 10 bar nitrogen pressure and $a_C < 0.2$ (cf. Figure 36, upper line). However, as mentioned above, the stabilization of silicon nitride is only possible, if the amount of boron (i.e. the boron nitride phase fraction) exceeds a certain value, which was empirically found to be at around 8 atom%. This amount approximately corresponds to 16 mol% BN in the final ceramic, if derived from boron-modified polysilazanes [B(C2H4SiR–NH)3]n or boron-modified poly(silylcarbodiimide) [B(C2H4SiR– N≡C≡N–)3]n. If the boron content in the materials is lower, there is obviously no sufficient dissolution of free carbon within the turbostratic BN phase. The second critical point is the relative phase amount of Si3N4. Even though there is a lack of more detailed studies, it turned out that the relative phase amount of silicon nitride should not exceed 30 mol%. If the respective value is higher, an efficient encapsulation of silicon nitride crystals that evolve during a heat treatment is not possible. Several publications report in detail on the high temperature mass stability of Si–B–C–N by considering the relative ratios of (SiC + BN + C)/Si3N4, Si3N4/(Si3N4 + SiC), and BN/(BN + C) [54,138,139,150,151]. However, this issue needs by far more detailed investigations and the development of new synthetic approaches to novel precursors for Si–B–C–N ceramics, which allow for the preparation of ceramics with adjustable chemical compositions.

4.5. High-Temperature Properties of Si–O–C and Si–M–O–C Ceramics

Silicon oxy carbide undergoes a phase separation at temperatures beyond 1100 °C to form amorphous nano-composites consisting of silica, silicon carbide and excess carbon [235,236]. Two processes occur in the Si–O–C system: (i) crystallization of phase separated silicon carbide; (ii) carbothermal reaction of phase separated silica with excess carbon, accompanied by the formation of crystalline SiC and release of gaseous CO. Above 1500 °C, severe decomposition of the Si–O–C ceramic due to the reaction of silica with SiC to form silicon monoxide and CO is thermodynamically expected [270].

The incorporation of additional elements into the Si–O–C network has a substantial effect on the decomposition and crystallization behavior. In Si–B–O–C ceramics, silicon carbide crystallizes at lower temperatures than in boron-free Si–O–C [271,272].
Recently, the high-temperature behavior of Zr- and Hf-modified Si–O–C has also been investigated. The materials were prepared by the chemical modification of a poly-silsesquioxane with metal alkoxide, followed by crosslinking and pyrolysis in argon atmosphere at 1100 °C [163,164,273]. Crystallization and microstructural evolution occur in four temperature regimes: (i) at around 800 °C the materials consist of a single amorphous Si–M–O–C phase (M = Zr, Hf); (ii) between 800 and 1100 °C the metal oxide phase precipitates throughout the matrix; the metal oxide nanoparticles (particle sizes of 2–5 nm) are mainly amorphous (Figure 37a); (iii) >1100 °C the Si–O–C matrix starts to phase separate to silica, silicon carbide, and excess carbon, whereas the metal oxide nanoparticles crystallize to t-ZrO$_2$ and t-HfO$_2$; (iv) >1400 °C a solid-state reaction between silica and the metal oxide phase occurs leading to the formation of a crystalline MSiO$_4$ (M = Zr, Hf) phase in the form of well-dispersed nanoparticles (20–30 nm) throughout the amorphous PDC matrix (Figure 37b) [163,273].

Systematic studies on the thermal stability of Si–O–C/ ZrO$_2$ and Si–O–C/HfO$_2$ ceramic nanocomposites upon annealing at $T >> 1100$ °C have shown that the incorporation of the metal oxide phase within Si–O–C strongly suppresses decomposition processes (e.g. carbothermal decomposition) even at temperatures as high as 1600 °C. In contrast to the Si–O–C sample, which exhibits a mass loss of 48.8% if annealed for 5 h in Argon atmosphere at 1600 °C, significantly lower mass losses are recorded for the Hf-modified samples (Table 6). The modification of the polysiloxane with 30 vol% of zirconium or hafnium alkoxide leads to Si–O–C/ZrO$_2$ and Si–O–C/HfO$_2$ ceramic nanocomposites with mass losses of 4.7 wt% and 8.6 wt%, respectively, if annealed under similar conditions [163,273].

The strong improvement of the thermal stability of the ceramic nanocomposites by the incorporation of zirconia or hafnia into the Si–O–C matrix relies on the suppression of the carbothermal reaction of silica with carbon. Instead, a reaction between silica and metal oxide occurs and leads to the formation of MSiO$_4$ (see Scheme 28, as for Si–O–C/HfO$_2$) [273].

### Table 6 Mass loss of Si–O–C, Si–O–C/HfO$_2$ (two different contents of hafnia, as 10 vol% and 30 vol% hafnium alkoxide has been used for the chemical modification PMS) and Si–O–C/ZrO$_2$ (two different contents of zirconia, similar to Si–O–C/HfO$_2$) upon annealing at 1300, 1400, and 1600 °C for 5 h in an argon atmosphere

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mass Loss [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si–O–C</td>
<td>0.60 1.00 48.80</td>
</tr>
<tr>
<td>Si–O–C/HfO$_2$ (10 vol%)</td>
<td>1.19 1.88 18.28</td>
</tr>
<tr>
<td>Si–O–C/HfO$_2$ (30 vol%)</td>
<td>0.22 1.33 8.56</td>
</tr>
<tr>
<td>Si–O–C/ZrO$_2$ (10 vol%)</td>
<td></td>
</tr>
<tr>
<td>Si–O–C/ZrO$_2$ (30 vol%)</td>
<td></td>
</tr>
</tbody>
</table>

**Scheme 28** Possible processes occurring in Si–O–C/HfO$_2$ ceramic nanocomposites upon annealing at 1600 °C.
The carbon contents measured in the samples obtained upon annealing at temperatures \(>1300^\circ\text{C}\) (as for Si–O–C/HfO\(_2\), see Table 7) support the fact that the carbothermal reaction of silica is suppressed [273]. Thus, the carbon content was shown to remain practically constant upon annealing, indicating that it is not consumed on reaction with silica.

5. APPLICATIONS

5.1. General Comments

The wide variety of available organometallic polymers, suitable as preceramic materials, offers exceptional opportunities for the development of PDCs. According to Figure 38, bulk parts, fibers, coatings, infiltrated media, fiber matrix composites, and near net shape manufactured ceramic components are available [3,243,274–276].

The preforms—also referred to as green parts—can be produced using techniques well known from polymer process engineering. The different approaches and techniques illustrated in Figure 38 will be discussed in the following sections topologically.

5.2. Monolithic Ceramics

Bulk ceramics can be obtained by polymer powder compaction and subsequent ceramization according to the flow scheme shown in Figure 39. In contrast to conventional sintering processes, no sinter additives are required for the densification. Powder compaction is usually achieved by cold, warm, or hot isostatic or uniaxial pressing.
depending on the physical/chemical properties of the precursor. A key step is the initial transformation of volatile or liquid precursors into highly crosslinked preceramic infusible networks.

PDCs suffer from a high volumetric shrinkage (>50 vol percent) and residual porosity that evolves during pyrolysis of the preceramic polymers. This is a consequence of the loss of organic substituents and an increasing density from 1 g/cm³ (precursor) to approximately 3 g/cm³ (ceramic). Thus, the conversion into ceramic is accompanied by the generation of mechanical stresses that lead to defects and cracking of the ceramic parts [277].

Many studies in the last two decades were directed to reduce shrinkage and porosity by optimizing precursor compositions and finding suitable processing conditions, for example, with respect to crosslinking and ceramization. Fillers can be used to minimize the overall shrinkage (Figure 40). There are two types of fillers for this purpose. Passive fillers serve as a space holders, they are not involved in reactions during pyrolysis (their size and composition remain constant). In contrast, active fillers react during pyrolysis either with the polymer matrix, the decomposition volatiles, or the pyrolysis atmosphere, thus changing/tailoring the shrinkage and the composition of the final ceramic [184,185].

Inert fillers are, for example, metal oxides, carbides, or nitrides (e.g. Al₂O₃, SiO₂, Y₂O₃, SiC, B₄C, Si₃N₄, BN). The volumetric shrinkage of a system consisting of a preceramic polymer and an inert filler (ωᵥ(IF)) can be described as ωᵥ(IF) = (1 - V_F/V_F*)ωᵥ(P), where V_F is the volume fraction of the inert filler, V_F* is the critical volume fraction of the inert filler and ωᵥ(P) represents the volumetric shrinkage of the preceramic polymer. The critical volume fraction of the inert filler is the volume fraction at which the filler particles start to form a rigid network, that is, further shrinkage of the system is not possible. Hence, at volume fractions ≥V_F*, the total system shrinkage is zero.

Active fillers are mainly used to produce near net shape parts because they nearly compensate the shrinkage of the precursors during pyrolysis. Usually, reactive powders are used, such as pure metals, intermetallics, metal hydrides or metal carbonyl complexes. The active filler usually reacts with the decomposition products or with the reactive pyrolysis atmosphere forming new phases that expand in volume. Therefore, the volumetric shrinkage of a system consisting of a preceramic polymer and an active filler, ωᵥ(AF), can be described as: ωᵥ(AF) = [(1 - V_F/V_F*)ωᵥ(P)] + V_Fωᵥ(F), where ωᵥ(F) is the shrinkage/expansion value resulting from the filler transformation reaction. Thus, the critical volume fraction of an active filler, V_AF*, to achieve net zero shrinkage in the bulk composite part is defined as V_AF* = ωᵥ(P)/(ωᵥ(P)/V_F* - ωᵥ(AF)) [184,185].

The near net shape processing of PDC parts using active fillers has been denoted by Greil as Active Filler Controlled Pyrolysis [185].

Composite systems with near zero shrinkage can be designed with both passive and active fillers. However, in the case of active fillers, lower filler volume fractions are necessary to achieve zero shrinkage. The use of filler particles not only reduces shrinkage during pyrolysis but may also endow tailored properties to the composite. In this way, special properties such as mechanical strength, thermal and electrical conductivity, magnetism, or surface features can be adjusted.
The polymer powder compaction process for the preparation of PDC bulk parts by cold isostatic pressing was developed by Riedel et al. [60,279,280] in the early 1990s. The final density of ceramic materials obtained by this procedure was in the range of 90% maximum. For obtaining dense monolithic bodies, however, warm pressing of suitable (nonvolatile) precursors is the state of the art. Warm uniaxial pressing as described by Seitz and Bill of PVS VT50 [281] delivered glassy green bodies with 7.5% of open porosity. Subsequent thermolysis released ceramic bulk parts with relative densities of around 97%.

Optimized conditions for this process were determined by means of thermomechanical analysis (TMA). Figure 41 shows such a diagram obtained from a cold-isostatically pressed PVS body (precursor: VT 50, Hoechst AG) [281].

TMA investigations give information on the shrinkage (or expansion) of materials upon heat treatment. If it is performed in an oscillating modus, it additionally allows for the determination of the softening temperature of the polymeric precursor (cf. inset in Figure 41). This value is of major importance because an increasing temperature results in a decrease in viscosity of the polymeric precursor, which enables a facilitated powder compaction. On the other hand, the maximum temperature applied for the powder compaction must be considered carefully. Exceeding this value (which correlates to the maximum of the TMA curve) results in the shrinkage of the green part due to a beginning ceramization during compression, causing the formation of shrinkage cracks which usually appear parallel to the pressing direction. As a consequence, shape integrity is lost, and in the worst case, the monoliths break into pieces.

Haug et al. investigated in detail plastic forming of boron-modified poly(silylcarbodiimide)s for the production of Si–B–C–N ceramic monoliths and correlated the conditions applied for the powder compaction with the open porosity of the derived bulk parts [282]. It turned out that the open porosity of bulk ceramics obtained from green bodies prepared at 100 °C/20 MPa could be decreased from about 20% to 12% by increasing the applied pressure to 43 MPa. If the same polymer was plastically formed at 120°C, the porosity decreased from 6.6% at 20 MPa to 5% if 43 MPa were applied for the densification of the polymeric powder (Figure 42).

It is obvious that the temperature applied during plastic forming of the precursors has more influence on the open porosity than the pressure applied. However, in the above case, further raising the temperature did not result in an additional reduction of the open porosity, because at and above 150 °C, shape integrity was totally lost as is evident from Figure 43 [282].

The absence of low-melting grain boundary phases in precursor-derived materials leads to extraordinarily high temperature mechanical properties that can be very different from those of conventionally sintered ceramics.
Recent studies of the high temperature mechanical properties of as-thermolyzed Si–C–N and Si–B–C–N materials have shown that the amorphous state reveals an outstanding mechanical stability, even at rather high temperatures [283–285]. However, the deformation mechanisms are not yet well understood.

Christ et al. characterized amorphous Si–B–C–N ceramics using isothermal compression creep testing in the temperature range of 1200–1500 °C [283,284]. It was found that the deformation rate contains a stress-independent section as well as a stress-dependent component that is proportional to the applied stress, indicating that this portion of the deformation mechanism is based on viscous flow.

As shown in Figure 44 for Si–B–C–N ceramics derived from [B(C₂H₂SiH–NH)₃]ₕ (MW33) and [B(C₂H₂SiMe–NH)₃]ₕ (T2-1), a continuous decrease in the deformation rate from approximately 10⁻⁶ to 10⁻⁸ s⁻¹ with time was found, which was mainly explained with a reduction of free volume in the amorphous material. Stationary creep was not observed, even after 450 h, indicating that the materials improve their creep stability during the creep tests.

It is obvious and interesting to note that the curves of both materials can be divided into two ranges. The first one referred to as range I up to approximately 2*10⁻⁵ is similar to primary creep in conventional materials, which can often be described using the Norton power law [286]. Range II is characterized by an accelerated decrease of the deformation rate as compared to range I.

Christ et al. [283,284] as well as Thurn et al. [287,288] also observed that the temperature dependence of the creep behavior of both Si–C–N and Si–B–C–N ceramics is comparably low and qualitatively the same for all temperatures. For example, the influence of the temperature on the deformation rate of T2-1-derived Si–B–C–N ceramic is shown in Figure 45.

Among the investigated samples, deformation rate was the lowest at 1350 °C and the highest at 1500 °C. The range I to range II transition occurred earlier at elevated temperatures. This shift was explained by a stress-independent reduction of free volume, which proceeds diffusion-controlled and thus accelerates with increasing temperature.

All creep tests mentioned above were performed in air and thus represent long time oxidation experiments. Once tested, the specimen usually possessed porous oxide surface layers consisting of silicon and oxygen as determined by energy-dispersive X-ray spectroscopy (EDX) and wavelength dispersive X-ray spectroscopy (WDX). Although the thickness of such oxide layers increased with testing time, a saturation value was reached. Although after short time exposures, even up to 1700 °C, the thickness of the surface layers was <2 µm [133], the thickness of such layers was around 50 µm after annealing in air for 300 h at 1400 °C and did not increase even after annealing for 500 h.

5.3. Ceramic Coatings

Preceramic polymers solutions and slurries can be applied to surfaces to produce ceramic coatings. The coatings are prepared using various application techniques such as spin coating, spraying, or dip coating followed by crosslinking and pyrolysis. The choice of the coating procedure mainly depends on (i) slurry properties (composition, rheology, etc.), (ii) the substrate (size, shape complexity, roughness),
and (iii) the desired coating thickness [289]. Thin layers can be deposited on planar, smooth surfaces by spin coating; the thickness can easily be adjusted by varying the spin rate [290,291]. Dip coating provides thicker layers and complex-shaped substrates can be coated via this technique [292,293]. Spray coating or screen printing can be used for layers with thickness up to 50 μm [294]. Independent of the procedure applied, one should note that there is a critical thickness (hCCT) above which delamination and cracking of the coating occur during pyrolysis [295,296]. With respect to delamination processes, hCCT depends on the adhesive strength between coating and substrate (\(\vartheta_a\)), on the coating shear modulus (\(G\)) and on the shrinkage ratio \(x\):

\[
h_{CCT} = \frac{\vartheta_a}{G\alpha(1 + \alpha^6 - 2\alpha^3)}
\]

For thick coatings of up to several hundred micrometers, multiple layer deposition (with or without intermediate layer treatment) or deposition of polymer/filler particle systems may be applied to reduce shrinkage and cracking (Figure 46).

Figure 46 SEM micrographs of Si–C–N-based coatings with a thickness of 20–30 μm, prepared via pyrolysis at 1100 °C in inert atmosphere on alumina substrates: (a) without use of fillers; (b) with 4.2 vol% alumina; (c) with 15 vol% alumina. One can observe that the addition of fillers suppresses crack formation upon pyrolysis [298].

Ceramization in an inert (nitrogen, argon) or reactive (ammonia, air) atmosphere may be used to furnish the PDC coatings. In the case of low melting substrates, advanced pyrolysis techniques such as laser pyrolysis or ion-beam-induced crosslinking and ceramization may be used. For instance, laser pyrolysis has been used to obtain ceramic coatings on plastic or aluminum substrates [297]. Furthermore, using focused laser beams, well-defined ceramic patterns on different substrates can be prepared [298].

PDC-based coatings can protect metal or carbon-based substrates from thermal loading and oxidative degradation [299,300]. Coatings consisting of Si–O–C filled with TiSi₂ particles can protect stainless steel surfaces toward thermal oxidation up to 800 °C. PDC-based coatings on silicon wafers [301,302], titanium, graphite, quartz [303], glass rods [304], alumina [305], and ceramic porous parts have a beneficial influence on the high-temperature behavior of the substrates used. Recently, silver-filled Si–C–N-based PDC coatings have been shown to exhibit antibacterial functionality [306].

A multifaceted field is the protection of carbon-fiber-reinforced ceramic composites from oxidation, hydrolysis, and erosion by PDC coatings. Because of their high maximum application temperature and low density, carbon-fiber-reinforced composites are very promising materials for realizing mechanically chargeable, light weight structures with a high potential for aerospace applications [307]. However, under the harsh re-entry conditions, carbon fibers in such composites are oxidized and erode. As a consequence, shape integrity of exposed components is lost, which in the worst case results in the total damage of the composite.

Figure 47 shows such a carbon fiber-reinforced silicon carbide substrate coated with a polymer-derived Si–B–C–N ceramic after one coating–thermolysis cycle. The precursor was obtained by ammonolysis of B(C₂H₄SiMeCl₂)₃ according to Ref. [124]. Additionally, 42 vol.-% of Si powder was added to the polymer solution to avoid crack formation during the polymer-to-ceramic conversion [308,309].

It is evident that the ceramic coating homogeneously covered the surface of the substrate. The right part in the image in which the fiber structures are visible represents noncoated substrate, which was fixed to the dip-coating equipment. Figure 48 displays a comparison by TGA of the
oxidation behavior of the pure and the coated composite at 1400 °C.

Obviously, the ceramic coating efficiently prevented the substrate from degradation. Whereas an exposure of the uncoated substrate to air after 500 min resulted in the loss of approximately 55% of the original sample weight, there was nearly no mass change observed in the case of the protected substrate. The observed 55% mass loss roughly corresponded to the amount of carbon fibers in the substrate and was due to a total “fiber-burnout.”

5.4. Ceramic Membranes

Microporous ceramic membranes are excellent candidates for gas separation applications. For steam reforming of natural gas, high-temperature reactors that operate at approximately 800 °C are generally used. High-pressure membrane reactors operating at significantly lower temperatures (ca. 500 °C) can potentially achieve conversion efficiency similar to that obtained in conventional reactors [310–313]. Furthermore, the membrane reactors have the advantage of synthesizing and purifying hydrogen gas simultaneously, thus leading to a highly efficient hydrogen production.

Two types of membranes are used for high temperature membrane reactors: (i) dense palladium-based membranes and (ii) microporous ceramic membranes. Palladium-based membranes exhibit excellent hydrogen permeability. However, they are very expensive, susceptible to poisoning by sulfur and sensitive to cracking or pinhole formation due to hydrogen embrittlement. Microporous PDC membranes exhibit relatively high gas permeability, but it is their exceptional high temperature stability that makes them superior to palladium-based membranes [314]. Remarkably, the choice of the precursor and the technology applied for its deposition allow to control the microstructure development of the membrane and to adjust the desired microporous/mesoporous structural features [315].

Microporous amorphous-silica-based membranes can be prepared as thin films on permeable alumina porous supports having graded and layered porous structure, with pore sizes ranging from several hundred nanometers to a few nanometers [316,317]. The gas transport properties of amorphous-silica-based membranes deposited on mesoporous anodic alumina capillary were investigated. The gas permeability for small molecules (such as He or H2) was much higher than those for larger gas molecules. This behavior emphasizes the potential of these membranes for use in gas separation properties [318]. Oyama et al. synthesized amorphous-silica-based membranes by means of CVD techniques. Whereas He and H2 permeate through the membrane, the permeation of molecules with a kinetic diameter > 0.3 nm such as CO, CO2, or CH4 is suppressed. This emphasizes the great potential of these membranes in applications related to the hydrogen purification. An amorphous-silica-based membrane was also synthesized by pyrolysis in air of a polysilazane deposited on a silicon nitride porous support. It exhibited a hydrogen permeability of 1.3 × 10⁻⁷ mol/m² sPa at 300 °C and a H2/N2 selectivity of 141, which is comparable with the selectivity of other amorphous-silica- or silicon oxycarbide-based membranes [319].

Amorphous ceramic membranes were also prepared in nonoxide systems, such as Si–C, Si–N, Si–C–N, and Si–B–C–N. The possibility of a molecular sieve amorphous SiC ceramic membrane was first demonstrated for a polysilastystrene derived composite membrane on a porous Vycor glass [320]. Amorphous silicon carbide ceramic membranes were prepared by thermal [317], e-beam [321], or chemical [322] curing of PCSs followed by pyrolysis in inert atmosphere. Hydrogen-selective Si–O–C–based membranes were prepared via air curing of PCSs and subsequent pyrolysis in argon [323–325].

**FIGURE 47** Carbon-fiber-reinforced silicon carbide composite covered with a ceramic Si–B–C–N coating. The coating was obtained by a dip-coating process using a slurry of 42 vol-% of silicon in a toluene solution of [B(C2H4SiMeNH)3]n [124] and subsequent thermolysis [308,309].

**FIGURE 48** Comparison of the oxidation behavior at 1400 °C in air of C/SiC composites with (squares) and without (circles) Si–B–C–N coating. The mass loss of around 55% in the case of the uncoated CMC corresponds to the amount of carbon fibers in the material [308,309]. For color version of this figure, the reader is referred to the online version of this book.
Amorphous silicon-nitride-based ceramic membranes were prepared by the pyrolysis of a polysilazane in ammonia atmosphere at 650 °C. The as-synthesized membrane showed a hydrogen permeability of $1.3 \times 10^{-8}$ mol/m² s Pa at 200 °C and a H$_2$/N$_2$ selectivity of 165, whereas after hydrothermal treatment at 300 °C, the permeability was $>1.0 \times 10^{-7}$ mol/m² s Pa at 300 °C with H$_2$/N$_2$ selectivity beyond 100.

Recently, novel polymer precursors have been synthesized for the preparation of PDC membranes in the systems Si–Ce–N and Si–B–C–N. The Si–C–N-based ceramic membrane was prepared via pyrolysis of a novel polysilylcarbodiimide precursor, which was synthesized via a nonoxide sol–gel process based on reactions of bis(trimethylsilyl)carbodiimide with chlorosilanes. Nitrogen sorption isotherm analysis of a 500 nm Si–C–N layer deposited on a porous support surface indicated the existence of pores with sizes in the range 2–5 nm [326]. Amorphous Si–B–C–N-based ceramic membranes with a thickness of approximately 1.75 μm were prepared by dip coating of a polyborosilazane on a macroporous alumina support, followed by pyrolysis in inert atmosphere (Figure 49) [327]. Pore analysis revealed a trimodal distribution of the pore sizes with maxima at 0.6, 2.7, and 6 nm.

Materials based on Si–C–N or Si–B–C–N are highly promising amorphous nonoxide ceramics that can withstand extremely high temperatures and are, therefore, of great interest for hot gas separation or filtration applications.

### 5.5. Ceramic Fibers

The fabrication of high-modulus refractory ceramic fibers requires controllable polymer rheology and adjustable polymer reactivity. Precursor fibers are obtained by either melt or dry spinning. Different principles for fiber processes such as extrusion, downdrawing from preforms, or updrawing from precursor melt or solutions are known [328].

Once spun, the precursor fiber is rendered infusible by subsequent curing, which represents an additional chemical surface reaction. Without this step, fiber integrity is lost because of melting and/or creep of the green fiber before full transformation into the ceramic state.

The first nonoxide silicon-based ceramic fibers were obtained from poly(dimethylsilane) by the Yajima process [26].

Synthesis of polydimethylsilane through alkali-metal-promoted dehalocoupling of dichlorodimethylsilane and its thermal (Kumada) rearrangement into processable poly(methylsilylene-methylene), [HSi(CH$_3$)$_2$–CH$_2$]$_n$, also referred to as polysilapropene, were already discussed in detail in Section 3.1. Polysilapropene fibers were obtained by melt spinning in an inert gas atmosphere such as nitrogen or argon at around 300 °C.

According to Figure 50, curing of the green fibers was performed by controlled surface oxidation in air at approximately 150 °C, which resulted in the oxidation of Si–H units accompanied with the formation of Si–O–Si linkages on the fiber surface. Heating of such partially oxidized fibers in an inert gas atmosphere to 1200 °C delivered Si–C–O fibers, which are commercially available under the trademark Nicalon™ [329]. As evidenced from HR TEM and X-ray photoelectron spectra, such fibers are composed of free carbon, nanocrystalline β-silicon carbide and an amorphous nonstoichiometric Si–C–O matrix. Heating of the fibers to temperatures exceeding the final thermolysis temperature results in their decomposition caused by the elimination of SiO and CO. A further crucial subject is the hydrogen content of the ceramic fibers which is progressively reduced when heating the fibers to 1200–1300 °C or higher. All these effects are accompanied by a coarsening of the fiber morphology and result in
a dramatic deterioration of the high temperature mechanical properties.

Alternatively, electron beam (2 MeV) or γ-ray (60Co) irradiation served for the chemical modification of the fiber surfaces, by which contamination of the fibers with oxygen could almost be avoided. The mechanism behind these processes is based on radical reactions. Noticeably, free radicals, which remain in the green fibers, have to be removed by additional annealing steps under inert atmosphere to reduce the reactivity of the extremely sensitive radical species in order to avoid oxidation and/or hydrolysis when exposing the fibers to air. Commercially available silicon carbide fibers, which are obtained by melt spinning and subsequent electron beam curing, are, for example, Hi-Nicalon™ fibers [329]. Even though oxygen-free, these fibers also crystallize at comparably low temperatures (<1300–1400 °C). In contrast to Si–C–O fibers, the formation of nanocrystalline β-SiC and random segregation of carbon occur, whereas there is no evidence for an amorphous Si–C–O phase.

Thermodynamic calculations predict that in an inert atmosphere oxygen-free SiC/C, fibers are thermally stable up to approximately 2500 °C. Unfortunately, such calculations only consider phase transformation or decomposition reactions that involve evaporation of gaseous species, whereas crystal growth cannot be taken into account. The latter is driven by a reduction of the surface energy and results in a grain coarsening with increasing temperature. Both silicon carbide grain coarsening and the increase of the size of the carbon domains negatively influence the mechanical properties of the fibers.

A critical aspect is the oxidation behavior of SiC fibers [330]. At increased temperature, either passive or active oxidation occurs, depending on the oxygen partial pressure. Passive oxidation is accompanied with the elimination of CO or CO2 and appears at lower temperatures and high oxygen pressure. According to Eqn (53) or Eqn (54), it results in the conversion of (surface-near) silicon carbide into silica, which forms a protective layer on the ceramic fiber.

\[ \text{SiC} + 2\text{O}_2 \rightarrow \text{SiO}_2 + \text{CO}_2 \]  
(53)

\[ \text{SiC} + 1.5\text{O}_2 \rightarrow \text{SiO}_2 + \text{CO} \]  
(54)

Even though carbon-containing species are evaporated in these reactions, the ceramic fibers gain in weight.

In contrast, oxidation at low oxygen (partial) pressure according to Eqn (55)

\[ \text{SiC} + \text{O}_2 \rightarrow \text{SiO} + \text{CO} \]  
(55)

results in the volatilization of carbon and silicon-containing species. It is consequently accompanied by a weight loss and in the worst-case results in the total damage of the ceramic fiber.

Recent publications also report on the synthesis and high temperature properties of silicon nitride-based ceramic fibers obtained from polysilazanes (see Section 3.2) or polyborosilazanes (Section 3.3). In previous sections, it was discussed that the derived ternary or quaternary ceramics resist crystallization up to temperatures of 1500 °C and approximately 1800 °C, respectively. This is mainly a consequence of the random amorphous ceramic network in which silicon atoms are present in mixed coordination spheres.

Even though numerous Si–C–N precursors have been published since the mid-1970s, there are only a very limited number of procedures described in the literature, which so far deal with the conversion of such polymers into ceramic fibers [197,331–337]. For example, Si–C–N green fibers were obtained by the melt spinning of a polymeric silazane HPZ, which was synthesized by reacting trichlorosilane with HMDS [333]. The green fibers were chemically cured by exposing to a stream of argon/trichlorosilane at moderate temperature and subsequently thermolyzed at 1200 °C. Alternatively, curing using trichloroborane served for rendering polymeric Si–C–N fibers rendering insusible [336]. In contrast to as-obtained Nicalon™ fibers, which are composed of free carbon, nanocrystalline β-silicon carbide and an amorphous nonstoichiometric Si–C–O matrix, Si–C–N fibers derived from HPZ consist of an amorphous nonstoichiometric SiN, C1−(3x/4) phase and free carbon. Remarkably, grain growth in Si–C–N fibers is retarded compared to Nicalon™ fibers, but at temperatures exceeding 1500 °C in an Ar atmosphere, degradation by carbothermal reduction of the Si–N units leads to a total breakdown of the fiber.

Baldus and Jansen obtained nonoxide silicon-nitride-based ceramic fibers, which do not decompose <1800 °C and which were claimed to additionally possess superior oxidation resistance [117–120]. According to Eqn (30) (cf. Section 2.5.2), TADB was reacted with methylamine to deliver a polyborosilazane, which after crosslinking appeared glass-like but which was still soluble and fusible. The as-obtained precursor could be melt spun at 100–120 °C using a multifilament spinneret, which delivers 200 single fibers. Subsequent curing and continuous heating to 1500 °C released almost oxygen-free (<1 mass-%) ceramic fibers with an approximate elemental composition close to SiBN 3C. Remarkably, oxidation experiments in air at 1500 °C for 50 h resulted in the formation of an amorphous 1.7 μm Si–C–O layer, which protected the fiber from further oxidation. Interestingly, tensile strengths of such oxidized fibers were identical to those of the as-obtained oxygen-free fibers.

Bernard et al. reported in detail the preparation of high-temperature Si–B–C–N fibers from polymeric precursors of the general type \([B(C_2H_4SiR(NR'))_3]_{10}\) (R = H, CH3, (NR')0.5; R' = H, CH3) [338–342]. The authors
investigated the influence of the different substituents on viscoelastic properties, crosslinking chemistry, and structure as well as performance of the ceramic fibers.

A means to determine the principal applicability of a precursor for extrusion processes, that is, spinning from the melt, is differential scanning calorimetry (DSC) in combination with TMA. The latter provides proper process conditions, whereas the former gives valuable information of the thermodynamics (i.e. melt-stability) under the desired processing conditions.

For example, the results of TMA investigations of two structurally very similar precursors \([\text{B}[\text{C}_2\text{H}_4\text{SiR(NMe)}]_3]_n\) \((R = \text{H, Me})\) are shown in Figure 51.

The polymers only differ in their silicon-bonded substituent \(R\), which is either hydrogen (1M) or methyl (3M). Nevertheless, the different functionalities of these groups have a significant influence on the processability of the precursors. Compound 1M does not melt below 160 °C and is thus not suitable for melt spinning, whereas softening of \([\text{B}[\text{C}_2\text{H}_4\text{SiMe(NMe)}]_3]_n\) takes place at 75–85 °C [339,341]. For color version of this figure, the reader is referred to the online version of this book.

![Figure 51](image.png)

**FIGURE 51** TMA of \([\text{B}[\text{C}_2\text{H}_4\text{SiR(NMe)}]_3]_n\) (1M(Me): \(R = \text{H}\); 3M(Me): \(R = \text{Me}\)). \([\text{B}[\text{C}_2\text{H}_4\text{SiH(NMe)}]_3]_n\) does not soften below 160 °C and is thus not suitable for melt spinning, whereas softening of \([\text{B}[\text{C}_2\text{H}_4\text{SiMe(NMe)}]_3]_n\) takes place at 75–85 °C [339,341]. For color version of this figure, the reader is referred to the online version of this book.

Melt spinning was performed using a laboratory scale piston extrusion equipment and a windup device setup in a nitrogen-filled glove box. The polymer melt was forced through a 200-μm single-hole spinneret and the filament stretched to approximately 50 μm in diameter and collected on the spool (50 m/min). To avoid degradation during the polymer-to-ceramic conversion and to maintain fiber integrity, the green fibers were subsequently cured in an ammonia atmosphere at 200 °C for 60 min. Such a process requires sufficient latent reactivity of the polymer, which is provided by a small amount of silicon-bonded terminal NHMe units, which are replaced by NH₂ groups in a transamination reaction. The newly formed NH₂ units condense, thus forming Si–N–Si bridges, which cause a surface hardening of the fibers.

Pyrolysis in an N₂ atmosphere releases ceramics in an approximately 30% yield. Remarkably, the yield may be increased to 57% when applying the abovementioned curing process. Figure 52 shows scanning electron micrographs of green fibers and as-obtained ceramic fibers. The fibers have smooth surfaces. Cracks or pores are not visible, neither in the green fibers nor in the ceramic fibers. During thermolysis, the fiber diameter decreases from approximately 50 μm to approximately 20–25 μm. This shrinkage corresponds to a decrease in volume of 75% (since the fibers were pyrolyzed on the spool on which they were collected, no shrinkage in fiber direction occurred). This value corresponds to the weight loss during thermolysis in combination with an increase in relative density from approximately 0.9 g/cm³ to approximately 2.1 g/cm³.

Such Si–B–C–N fibers possess excellent thermal stability. High temperature TGA in an N₂ atmosphere points to the fact that no decomposition takes place at temperatures up to 1700 °C. At 1750, degradation with volatilization of gaseous species and crystallization occurs, which accelerates at higher temperatures. The change in the fiber morphology with temperature can nicely be monitored by Scanning electron microscopy (SEM) (Figure 53).
The as-obtained 3M ceramic fiber shows typical glass-like fracture. The fracture surface is smooth, and there is no evidence for the presence of pores, cracks, or inhomogeneities. The microstructure is preserved after annealing for 2 h at 1700 °C in a N₂ atmosphere. An increasing temperature results in a beginning degradation. The fracture surface of a fiber annealed at 1750 °C displays a remarkable coarsening of the microstructure and extensive grain growth. The beginning degradation is also visible on the surface of the fiber, which is still smooth, yet it reveals certain irregularities. The fiber annealed at 1900 °C is totally deteriorated. It is characterized by a very dense shell of approximately 0.5-μm thickness, which encloses a porous and crystalline-coarse interior. Such fibers are extremely brittle and no longer qualified as reinforcement in fiber-reinforced ceramics.

5.6. Fiber-reinforced Ceramics

Fiber-reinforced ceramic matrix composites (CMCs or FRCs) have been shown to exhibit extraordinary thermomechanical properties combined with low density. They are consequently highly interesting candidates for manifold applications under severe conditions especially in aerospace [307].

CMCs can be obtained by different methods. The most established procedures are liquid phase routes, gas phase routes, or their combination, referred to as hybrid processes [307].

Liquid phase routes occur by infiltration of liquids such as melts, solutions, or slurries (which will not be discussed here in detail) into a fiber preform. The first one, usually a reactive melt infiltration, requires matrix materials with sufficiently low melting point, such as silicon (1410 °C) or aluminum (660 °C). During the infiltration step, the molten elements react either with the fiber preform, for example, carbonaceous fiber prepregs, or the atmosphere. Even though these methods allow for a rapid preparation of near net shape CMCs, there are some drawbacks, which have to be considered carefully. For example, the preparation of carbon-reinforced SiC composites is performed by capillary infiltration of liquid silicon. Molten silicon readily wets carbon and under the applied conditions reacts—on the expense of the carbon fibers—to SiC, even in the presence of amorphous carbon. To avoid fiber degradation, it is essential to protect the fibers, for example, by depositing boron nitride coatings, which is an additional time-intensive and expensive procedure.

The preparation of Si-C-based CMCs with inert fibers, such as Nicalon, is also possible. However, in this case, it is inevitable to carefully adjust the ratio of amorphous carbon in the prepreg and silicon, which is infiltrated in order to avoid segregation of residual nonreacted elemental silicon, since such a segregation usually worsens thermomechanical properties of the CMC dramatically.

Precursor-derived CMCs, which can be obtained by polymer impregnation and pyrolysis (PIP) process, do not suffer from these drawbacks. Although rather simple, the
process cannot be applied for every kind of polymeric precursor. Prerequisites for an effective polymer infiltration and loading of the preforms are adequate fluidity, that is, low viscosity of the polymer solution to allow for an efficient penetration into the preform, sufficiently high concentration of the solution to prevent extensive shrinkage during the drying step, appropriate crosslinking chemistry of the polymeric precursor, and high polymer-to-ceramic conversion yields. Even if a precursor fulfills all requirements, a large number of infiltration/thermolysis cycles may be required to obtain fully dense CMCs.

Weinmann et al. published the first approaches for the fabrication of ultrahigh temperature CMCs with precursor-derived Si–B–C–N ceramic matrix using two-component precursors. In contrast to the abovementioned system, this route avoids the use of solvents [142–144]. According to Eqn (35) (Section 2.5.3), the polymeric matrix is obtained by a thermally induced hydrosilylation of OVS, with tris(hydridosilylethyl)boranes, B(C2H4Si(CH3)3)nHn]3 (n = 1–3, in the literature referred to as Tris-SiH1–Tris-SiH3). Remarkably, polymer synthesis occurs quantitatively without requiring catalysts and without the formation of byproducts. The physical/chemical properties of the starting compounds and the straightforward synthetic procedure—the starting compounds were simply mixed and reacted at moderate temperature—allowed for the preparation of fiber-reinforced Si–B–C–N green parts using the RTM process. The single steps of the RTM process, as was performed for the preparation of the ceramic fiber-matrix composites, is shown schematically in Figure 54.

Since the monomeric starting compounds are extremely sensitive to oxygen and/or moisture, special equipment, which is shown in Figure 55, was developed in order to allow for a handling of all substances under inert gas conditions.

It consists of a Schlenk flask that serves as a reservoir for the precursor mixture and a brass mold, which contains the preforms. The flask and mold are linked via an interconnection glass tube. The valves attached allow separate evacuation of the two main parts. Initially, the brass mold is filled with the fiber material and the whole system is thoroughly evacuated. The central valve is closed and the Schlenk flask flooded with inert gas. The reaction mixture is introduced and the Schlenk flask evacuated again. After the central valve is opened, the precursor mixture is injected into the brass mold by slowly introducing argon through the valve attached to the Schlenk tube. After injection, the central valve is closed again and the brass mold heated within an oil bath to crosslink the precursors. The fiber-reinforced green body is removed from the brass mold within a glove box and subsequently pyrolyzed.

Suitable conditions during the crosslinking process are of major importance with respect to obtaining a sufficiently dense matrix. Otherwise, an excessive amount of defects will appear, degrading the mechanical properties of the final composite.

The initially applied two-step heating program reported for the crosslinking chemistry of OVS and Tris-SiH2 (cf. Figure 56), that is, heating to 120 °C within 2 h, isothermal annealing for 60 min and further heating to 200 °C using a heating rate of 20 °C/h, caused a foaming of the low-viscous reaction mixture. Accordingly, the fiber-reinforced composites displayed bubbles and voids and possessed an unsatisfactory green density (Figure 56a) [342].

For this reason, systematic experiments were performed to improve the crosslinking process with respect to
obtaining a dense crack-free polymer matrix. Optimized heating schedules were explored empirically. According to the drawn-through line in Figure 56, a temperature regime was chosen, in which the samples were heated to 170 °C within 5 h, isothermally annealed for 6 h, and subsequently heated to 175 °C and 180 °C (isothermal annealing each 3.5 h). Crosslinking was completed by heating the precursor to 200 °C for 6 h. Finally, a dense glass-like and crack-free polymer matrix was obtained (Figure 56b).

Thermolysis in an Ar atmosphere at 1400 °C released ceramic matrix composites. Generally, their density is a function of various parameters such as the volume fraction of fiber and matrix, the efficiency of the impregnation step, weight loss during crosslinking and the ceramic yield of the precursor. However, even if assuming full impregnation and very high ceramic yields, a significant matrix shrinkage during pyrolysis appears, which is due to an increasing matrix density from approximately 1 g/cm³ (precursor) to approximately 2.1–2.5 g/cm³ (bulk ceramic). Accordingly, cracks and pores appear during the heat treatment. Such voids, however, can be subsequently filled by repetitive polymer impregnation and pyrolysis (PIP).

The microstructure of a specimen obtained after 16 PIP cycles (89% relative density) is shown in Figure 57. It displays a dense, crack- and pore-free matrix. The integrity of the carbon fibers, which are visible as elliptical gray...
spots, is maintained (note: fibers are spherical but appear oval because cutting and polishing was not performed perpendicular to the fiber direction). Neither oxidation on the fiber surface nor within the fiber-matrix interface was visible. Moreover, it is evident that the voids between the individual carbon strands were filled by the matrix material with sufficient homogeneity.

Bending strength of as-obtained FRC was 255 MPa. The thermal stability of the C/\(\text{Si-B-C-N}_{\text{matrix}}\) composite was investigated by TGA and XRD. It retained its microstructure and mechanical properties up to 1500 °C. After annealing at higher temperatures, that is, 1700, 1900, and 2000 °C, a weight loss appeared and the matrix crystallized. The strength thereby decreased to <100 MPa. Brittle fracture of the FRC did not take place, regardless of the heat treatment conditions applied; the as-prepared sample and the sample investigated at 1500 °C displayed a very well-developed fiber pull-out. In contrast to Si-B-C-N bulk ceramics in which the creep rate under compression decreases continuously [283], the C/\(\text{Si-B-C-N}_{\text{matrix}}\) composite possessed secondary creep up to at least 60 h at 1400 °C and 100 MPa load. However, the creep strain, which measured 0.55%, was significantly lower than values typically measured for FRC made by chemical vapor infiltration (CVI) method. Most probably, transverse microcracks in the transverse bundles and interply microcracks mainly contributed to the creep strain (Figure 58).

6. CONCLUSIONS

PDCs are unique materials with excellent properties such as high temperature stability against decomposition, crystallization, oxidation, corrosion, and creep. Furthermore, the chemical and phase composition as well as the microstructure and the properties of the PDCs can be tailored by appropriate designing of the preceramic polymers and by choosing specific crosslinking and pyrolysis/annealing conditions.

In this chapter, the synthesis and properties of PDCs has been considered with an emphasis on the design of the preceramic polymers, their processing and their transformation into ceramics. Also the high-temperature behavior of PDCs has been reviewed and discussed in the light of the strong relationship between the architecture of the preceramic polymer and the microstructure and properties of the resulting ceramic.

Despite the extensive research performed in the past decades, there are still open questions related for instance to the microstructure-property relations of PDCs. Thus, further studies are needed, which will require novel interdisciplinary approaches to both basic research and materials development.

REFERENCES

Precursor-Derived Ceramics


Chapter | 11.1.10 Precursor-Derived Ceramics


[192] Polysilazanes with general structure [RSi(NH)₁.₅]ₙ, e.g. commercially available poly(vinylsilsesquiazane) VT50, Hoechst AG, Germany.


[198] PCS is a commercially available product of Nippon Carbon Co. Ltd., Type S, Tokyo, Japan.

[199] PHPS is a commercially available product of Tonen Co., N-N110, Tokyo, Japan.


Handbook of Advanced Ceramics


[322] Nicalon™ and Hi-Nicalon™ are trademarks of Nippon Carbon Co., Tokyo, Japan.


Combinatorial Nanoscience and Technology for Solid-State Materials

Hideomi Koinuma,* Ryota Takahashi,† Mikk Lippmaa,‡ Se-Young Jeong,‡ Yuji Matsumoto,‡ Toyohiro Chikyo** and Setsu Suzuki**

*Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan, †Pusan National University, Miryang, Gyeongnam 627-706, Korea, ‡Comet Inc., 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan, **Institute for Solid State Physics, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8581, Japan, ‡‡Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Chapter Outline

1. The Why and How of Combinatorial Technology in Materials Research 1103
2. Instrumentation of Combinatorial Technology 1105
3. Combinatorial Methods Applied to Ceramics Research 1107
4. Discoveries Made by Combinatorial Technology 1115
5. High-Tech Venture: Combinatorial Material Business 1122
6. Conclusion 1123
Acknowledgments 1123
References 1123

1. THE WHY AND HOW OF COMBINATORIAL TECHNOLOGY IN MATERIALS RESEARCH

The modern world of technology is changing rapidly. For materials science, the rapid change creates intense competition in developing new manufacturing processes and new products. In contrast to the conventional discrete materials development process that depends on experience and intuition, the combinatorial approach enables us to quickly synthesize and characterize systematic sample libraries. Numerous possibilities can be tested by rapid screening, resulting in the discovery of novel materials as well as in the reduction of the amount of energy and raw materials used for research and development. Pioneered by the pharmaceutical industry, combinatorial chemistry has prevailed in the synthesis of new chemicals, especially new drugs. In an automated parallel chemical reactor containing immobilized reaction sites on a bead in a solution, various chemicals can be fed and rinsed out as programmed to leave an intermediate reaction product anchored to the bead. Subsequently, the product library is characterized in parallel or by rapid scan analyzers to pick up candidate(s) for a desired product.

Based on the insight that such organic combinatorial chemistry is substantially composed of downsizing of individual reactors and integrating multiple reactors into arrays, we proposed an extension of the combinatorial concept to solid-state materials, inclusive of ceramics, semiconductors, and metals [1,2]. For downsizing and integration, we primarily used physical vapor thin film deposition, combined with movable shadow masks to grow many different thin films at specific locations on a single substrate, creating a parallel thin film library. As an example, modern systems for high-throughput exploration of advanced materials and devices are being used to solve the problems related to energy conversion, storage, and transmission, thereby reducing energy consumption and CO2 emissions, contributing to a sustainable future of the Earth. A substantial downsizing and series integration of chemical reactions can be traced back to the 1960s when Merrifield invented a scheme shown in Figure 1(a) for sequential peptide synthesis [3,4]. A series of amino acids is connected in a preprogrammed order to a reaction site on an insoluble polymer bead, resulting in the formation of the desired proteins. This method can not only miniaturize the chemical reactor but can also significantly promote the efficiency of separation and purification of products by the spatial separation of the products from the uniform solution by the heterogeneous solid polymer beads. In the 1980s, Merrifield’s method was extended to parallel synthesis and to systematic screening of raw material compositions and reaction conditions, and was named “combinatorial
chemistry,” becoming a standard method for new drug development [5]. The combinatorial chemistry concept was further extended to solid materials and device development in the mid-1990s when it was used for parallelizing the traditional ceramics process of sintering, producing thermodynamically controlled libraries [6–8]. Soon after that, we developed a more sophisticated technology for parallel integration of sequence-controlled and nanostructured thin film libraries [9]. For solid-state material development, especially in the form of thin films and for devices, the control of atomic-layer stacking direction and stacking order is important. True combinatorial solid-state chemical technology was created by coupling automatically controlled layer-by-layer growth of ceramic thin films with the parallel integration concept. Fabrication and characterization tools were designed for precise control of the stacking structure and periodicity of oxide crystal lattices and for the optimization of synthesis conditions in a three-dimensional coordinate space.

The combinatorial synthesis strategy of solid materials brings together three important components. First of all, as illustrated in Figure 2, combinatorial parallel synthesis has to be combined with rapid characterization to prevent the one-by-one analysis bottleneck in experimental throughput. Even when high experimental throughput is achieved, it is necessary to analyze the results in order to design new synthesis experiments. The combinatorial workflow circle is therefore completed by adding the necessary data processing, visualization, and simulation tools, collectively known as Materials Informatics [10].

In a typical parallel workflow, the material development process starts with the synthesis of a combinatorial library. For a thin film library, for example, in situ shadow masking may be used to integrate a collection of individual small samples on a single substrate, either in a parallel deposition or by growing individual samples in a single sequence. The library is usually designed so that there is a systematic variation of either material compositions or synthesis conditions between the individual thin film samples in the library. Second, the prepared library is rapidly characterized by high-throughput characterization tools, where the primary purpose is usually to perform rough screening of material properties. Individual compositions or even libraries may then be used for an in-depth study, if necessary. The characterization tools typically limit the size and the density of the library by setting the minimum useful sample area and the maximum

FIGURE 1 (a) Material synthesis as a multistep process. (b) Comparison of the Merrifield method and combinatorial chemistry. (c) A combinatorial library of solid materials in a two- and three-dimensional coordinate space.
measurable library area. The minimum size of a single sample in the library may be set by the size of an X-ray beam used for diffraction measurements, while the maximum library area could be set by the scanning area of an instrument, as is often the case for various scanning probe microscopes. In addition to library size and density, the synthesis and characterization tools also need to be matched in terms of throughput. Especially for scanning-type characterization tools, it should be possible to characterize the whole library area on a timescale that is much less than the time spent on the library synthesis. The requirements for fast and high-resolution characterization may conflict with the ability to perform quantitative measurements of material properties. A tradeoff is often used to obtain purely qualitative measurements for screening and possibly using a lower-density library for more accurate and quantitative measurements.

Rapid parallel characterization can yield very large data sets. To maintain the throughput of the whole material development workflow, including synthesis, characterization, and analysis, special analysis tools are also necessary. A variety of software tools are used to accumulate, manage, process, and visualize the experimental results. Collectively, these tools are known as Materials Informatics [10]. Informatics tools are well established in combinatorial life sciences, for example, for storing, processing, and distributing genetic sequences. Software tools for solid-state synthesis are less well developed, largely due to the wide variety of characterization tools that are often experiment specific and not standardized. Software solutions for storing and distributing large data sets are therefore mostly site specific, limiting the effectiveness of data mining and modeling. Regardless of the need for specialty software, informatics tools are essential for maintaining the total throughput of the combinatorial workflow.

2. INSTRUMENTATION OF COMBINATORIAL TECHNOLOGY

The instrumentation used for synthesizing combinatorial libraries is dependent on the type of materials that are studied. Here, we review one technique, pulsed laser deposition (PLD) of oxide thin films, as this example covers many of the issues found in the synthesis of various other library types as well, including in situ definition of the library structure, synthesis process monitoring by scanning real-time tools, and characterizing two-dimensional sample array libraries.

Figure 2(a) schematically illustrates the combinatorial PLD method [11, 12]. In the basic PLD technique, a deep-ultraviolet excimer laser is used to evaporate the desired material in vacuum or in the presence of a background gas. The evaporated plasma rapidly expands, forming a directed plume that carries the evaporated atoms to the surface of a substrate, which is usually at an elevated temperature. The main advantages of using PLD for growing oxide thin films are the nearly stoichiometric transfer of material from a target to the film and the ability to control the growth rate and film thickness with a submonolayer accuracy. Although the presence of multiple selectable targets in a PLD system is generally a matter of convenience, it is an essential feature of a combinatorial PLD system, as it allows for mixing two or more different target materials on the surface of the growing film. A similar multiasource capability is usually present in almost all combinatorial synthesis tools.

The purpose of combinatorial synthesis is to produce a library of samples, each grown under different conditions or having a different composition. In the case of PLD, this is achieved by placing computer-controlled stencil and shadow masks in the path of the ablation plume. The role of
the stencil mask is to provide library element separation, as is often necessary for transport measurements of individual samples in a library, while a shadow mask is used to selectively open certain cells in a library for deposition. By controlling the movement of the masks synchronously with the ablation laser pulses and target motion, it is possible to grow a variety of different types of thin film libraries.

A common purpose of a combinatorial experiment is the mapping of material properties as a function of composition, that is, rapidly generating a phase diagram for a particular physical property [13]. As the PLD method transfers the target material directly to the film, variation of the film composition is usually achieved by mixing material from several targets in predetermined ratios. This is possible, if the growth rate is accurately controlled. For example, a 50:50 mix of two targets can be obtained by alternately evaporating an equal amount of material from two targets. As long as the layer thickness grown from each target is below a single growth unit, typically a single unit cell, surface diffusion ensures mixing of the materials and the formation of a solid solution film. In situ analysis of growth rates in different parts of the library from each target can be done by diffraction techniques. Due to the geometry of the PLD process, the most convenient and surface sensitive technique is in situ scanning reflection high-energy electron diffraction (RHEED) [12,14].

A schematic illustration of an RHEED setup used in combinatorial experiments is shown in Figure 3(a). In this implementation, two sets of electron beam deflection coils are used and controlled by a computer. RHEED diffraction patterns are usually observed along a high-symmetry direction of a crystal. While a simple beam deflection system with a single set of coils would be sufficient to position the electron beam at any point in the library, the incident angle of the beam would also vary, complicating the pattern and diffraction intensity analysis. For this reason, a double coil deflection setup is used, which allows for parallel movement of the electron beam across the library surface. By synchronizing the electron beam movement with the diffraction pattern imaging camera, real-time imaging of crystal growth is possible at a number of points in the library during film growth [12]. Typical diffraction patterns obtained by scanning the electron beam across a cubic perovskite surface, with the beam along the [001] direction, are also shown in Figure 3(a).

By using specially designed shadow masks and corresponding motion sequences, it is possible to fabricate a wide variety of combinatorial libraries with discrete cells, with binary [13] or ternary [15] composition spreads, superlattice libraries [12], etc. Crystal growth parameters can also be studied. Among the most useful libraries for this purpose are the temperature-gradient films that are fabricated by using a specially designed sample holder and a tightly focused heating laser [16]. In this case, a stable temperature gradient, often covering a range of 300° or more exists in the substrate. By combining the temperature-gradient method with a composition spread, it is also possible to obtain a two-dimensional parallel film library in which the growth temperature varies along one axis and the film composition varies along another direction on the

![FIGURE 3](image-url)  
**FIGURE 3** Schematic illustrations of combinatorial characterization methods: scanning-beam RHEED (a), concurrent XRD (b), and magnetic characterization with a scanning SQUID microscope (c).
substrate surface [17,18]. This combinatorial synthesis technique is not specific for PLD but can also be used in libraries grown by sputtering or molecular beam epitaxy (MBE). The combination of composition and temperature gradients is not limited to oxides either, but it can be used for mapping phase diagrams of a variety of materials, including metal alloys and organic thin films.

Rapid characterization of libraries after synthesis is just as important as the parallel synthesis itself in a combinatorial workflow. Among the most useful tools for guiding materials development is the rapid analysis of crystal structures. However, structural characterization by X-ray diffraction (XRD) is usually a rather time consuming part of experiments. For achieving high-throughput characterization of integrated thin film libraries, a concurrent X-ray diffractometer has been developed, as shown in Figure 3(b) [19,20]. The system is composed of a convergent X-ray beam source (Cu Kα1), a conventional powder diffraction goniometer, and a two-dimensional detector. A dispersed X-ray beam from a source is focused onto a line of 10 × 0.1 mm² on a 10 × 10 mm² library chip, generating a two-dimensional intensity image on the detector, one axis of which corresponds to a diffraction angle range of about 2° and the other axis represents a position in the library, that is, covering a single row of samples in a library. The measurement of a single cut through a library takes only a few seconds and a reasonably wide range of diffraction angles, spanning from 0 to 90°, can be obtained in a few minutes by scanning the detector camera position, which is more than ten times faster than with a conventional X-ray diffractometer and competing with synchrotron-based microbeam setups. The resolution of the diffraction angle and position are 0.01° and 100 μm, respectively.

In cases where a single library covers a significant fraction of a phase diagram, rapid property mapping can be used to obtain complete phase diagrams in just a few experiments. This information is critical for materials modeling and for providing a reference for combinatorial computational chemistry. It can also lead to serendipitous finds, such as the discovery of ferromagnetic order in transparent Co-doped chemistry. It can also lead to serendipitous finds, such as the discovery of ferromagnetic order in transparent Co-doped.

3. COMBINATORIAL METHODS APPLIED TO CERAMICS RESEARCH

3.1. Composition Spread Libraries

As was discussed earlier, combinatorial thin film libraries are designed to match the characterization tools that are used to map the desired physical properties. Because scanning local probes with various spatial resolutions are the most efficient tools for rapidly characterizing material properties, it is not always necessary to physically separate individual samples, or cells, in a library. This simplifies the film growth process, as a cell-defining stencil mask can be omitted and a library can be grown by using only movable shadow masks. In a composition-spread library, as the name implies, the composition of the film varies continuously across a substrate, and it is thus the preferred technique for mapping phase diagrams. The technique of using films with continuously varying compositions of a solid solution system on a substrate and can be traced back to the codeposition technique reported by Kennedy in 1965 [24] and Miller in 1967 [25], as illustrated in Figure 4(a). Subsequent reports came from Hanak [26,27], Koinuma [28], and several other groups. The method is still in use in various physical vapor deposition systems and can be used for either binary or ternary composition spreads. A recent report from van Dover describes a study of dielectric properties of various ternary oxide systems, such as SrO−SnO₂−TiO₂ [29]. This codeposition composition spread technique is a simple process that uses the composition gradient that occurs naturally across a wafer. The main drawback is that the distribution of the composition on a substrate cannot be directly controlled and must be determined independently by ex situ analytical techniques, such as electron probe microanalysis. This additional analysis step has a negative impact on the throughput, as other property mappings cannot be used before the composition has been measured. Another drawback of this method is the difficulty of covering the entire phase diagram, because of inevitable intermixing of chemical species evaporated from different sources. Natural composition spreads are therefore not useful for studying the regions of phase diagrams close to end compositions. To avoid this problem, Xiang and Schultz fabricated spatially addressable materials library chips [6–8], illustrated in Figure 4(b). These libraries resembled the
multiwell plates used in life sciences but required each cell in a library to be individually grown or mixed. This can be viewed as a second-generation technique, and while it addressed the problem of mapping complete phase diagrams, the method has a much lower throughput than the natural composition spread technique. Subsequently, a general methodology was developed for systematic synthesis of inorganic materials libraries using a precursor multilayer thin film technique. Different combinations of amorphous layers (one precursor layer for each element) are put down at different sites, and afterward, materials at different sites are simultaneously thermally processed. Thermal diffusion of amorphous layer and nucleation of alloy phases determine the phase formation in this process.

The addressable composition spread technique for binary [13] and ternary [15] phase diagrams, shown in Figure 4(c), can be used for both amorphous and crystalline materials and can, in principle, cover a full phase diagram, including the end compositions. The basic procedure of fabricating an addressable binary composition spread library is shown in Figure 5 [13]. In a simple composition spread system, a single movable shadow mask may be used to limit the area of a substrate exposed to the physical vapor from a PLD, sputtering or a molecular beam source. As shown in Figure 5(a), if the mask is moved slowly to the left during evaporation of material A, gradually covering the substrate surface, a film with a thickness gradient is obtained. In the next step, the evaporation target is switched to material B, and the mask is slowly shifted to the right, as shown in Figure 5(b). This will produce another thickness gradient layer but with the opposite gradient orientation. The total film thickness in this two-step process is usually close to a single unit cell, guaranteeing efficient mixing of precursors A and B, forming a uniform alloy phase. Steps shown in Figure 5(a) and (b) are repeated many times, until the desired film thickness is achieved, as shown in

**FIGURE 4** Schematic illustrations of three generations of composition spread libraries: (a) natural composition spread grown by a codeposition technique, (b) spatially addressable library, (c) addressable continuous binary and ternary composition spread libraries.

**FIGURE 5** A schematic diagram of a binary continuous composition-spread deposition. (a) A shadow mask moves from the right to the left during the deposition of material A. On the substrate, the thickness of the deposited precursor A layer increases monotonically from the right to the left. (b) Similarly for depositing material B, the mask is moved in the opposite direction, producing another gradient layer. (c) For each deposition step cycle, one monolayer is deposited on a substrate. Steps (a) and (b) are repeated many times until the desired film thickness is achieved, yielding the addressable binary composition spread library (d).
Figure 5(c). For high-temperature growth, an alloy phase is formed during growth. If amorphous precursors are deposited, an additional annealing step may be used to produce an alloy film, as shown in Figure 5(d).

The use of addressable binary composition spreads has been demonstrated for many different materials. Figure 6(a) shows a concurrent XRD mapping of a Ba,Sr,TiO$_3$ composition spread library close to the SrTiO$_3$ substrate (002) reflection [19,30]. The vertical and horizontal axes correspond to the sample position, that is, composition parameter $\alpha$ and the 20 diffraction angle, respectively. The out-of-plane lattice parameter of the film is plotted in Figure 6(a) as a straight vertical line at a diffraction angle of 46.5°. The film peak can be seen on the left-hand side of the substrate peak, and it is clear that the film diffraction peak position shifts to lower angles as a function of the film composition up to $\alpha = 0.6$. At $\alpha = 0.6$, the peak position abruptly shifted and the peak broadened in the 0.6 < $\alpha$ < 1.0 range. Figure 6(b) shows the out-of-plane and in-plane lattice constants derived from the XRD mapping. For 0 < $\alpha$ < 0.6, the out-of-plane lattice constant increased while the in-plane lattice constant remained constant with increasing $\alpha$, consistent with fully-strained growth. For $\alpha > 0.6$, the out-of-plane lattice constant abruptly decreased at $\alpha = 0.6$ and remained almost constant close to the BaTiO$_3$ end of the composition spread. A gradual increase of the in-plane lattice parameter shows that the gradual substitution of Ba for Sr resulted in an expansion of the unit-cell volume. For $\alpha < 0.6$, coherent epitaxy took place despite the cell volume increase, resulting in an in-plane lattice constant of the film that was equal to that of the substrate and the lattice was thus compressively strained. At $\alpha = 0.6$, the increasing in-plane lattice mismatch between the film and the substrate generated misfit dislocations at the interface. Hence, the compressive stress was abruptly relaxed, as seen by the abrupt change of the in-plane and out-of-plane lattice constants and the broadening of the film diffraction peak. This binary composition spread experiment was used to obtain a growth mode map of the alloy phases and was used to map the strain-dependent dielectric properties. Similar high-throughput characterization methods have recently been used to study (La,Sr)MnO$_3$ [13], Sr$_2$(Rh,Ru)O$_4$ [17], and (Zn,Co)Fe$_2$O$_4$ [31] systems.

3.2. Ternary Phase Diagrams

Formation of addressable ternary composition spreads with predetermined composition distributions has been achieved by controlling the film deposition time or thickness at each position on the substrate with the aid of physical shadow mask movement and substrate rotation [15,32–34]. Figure 7 schematically illustrates the preparation process for an addressable ternary composition spread library. A ternary composition spread library can be grown by superimposing three linear gradients that follow the three axes of an equilateral triangle, as illustrated in Figure 7(a)–(d). This can be easily achieved by using a single shadow mask and 120° substrate rotations. The total thickness of the three gradient films deposited in a single cycle is kept constant at one unit cell. The elemental diffusion length should exceed the vertical film thickness of each deposition cycle on a heated substrate, whereas it should be short enough to maintain lateral spatial resolution of the library. This condition is always satisfied in thin films. The sequence of growing linear gradients and library rotations is repeated until the desired total film thickness is obtained, as shown in Figure 7(e), yielding an addressable continuous ternary composition spread thin film library.

A variety of shadow mask designs have been used in combinatorial thin film growth systems. In many cases, the type of masks used is dictated by the vacuum system geometry. In general, shadow masks used for preparing thickness gradient films use either a single edge or two edges, which pass over the substrate surface once or twice in each deposition cycle. Figure 8 shows a schematic illustration of the preparation procedure of a ternary composition spread film by using a single-edge mask that can be used without rotating the substrate [34]. Three separate masks, shown in Figure 8(a)–(c), are used, with working edges oriented at 30°, 90°, and −30° relative to the mask movement direction. The three masks are used to grow three film layers with the thickness gradients of the layers separated by 120° rotations, as shown in Figure 8(d)–(f). The actual mask shape used in a deposition chamber is shown in Figure 8(h). In principle, the whole ternary phase space shown in Figure 8(g) can be

![FIGURE 6](image-url) (a) Concurrent XRD mapping of a Ba,Sr,TiO$_3$ composition spread library. (b) Out-of-plane and in-plane lattice constants derived from the concurrent XRD mapping.
covered, although typically a single library covers only a subset of the full phase diagram.

Figure 9 shows a schematic illustration of a masking technique that also produces a ternary composition spread library but uses double-edge masks [34]. An arbitrary angle $\alpha$ of the double edge strongly affects the shadow mask motion speed and the film deposition speed. The characteristic point in the double edge is that a linear thickness gradient is available along the 90° rotated direction with respect to the mask movement direction, whereas this gradient direction is impossible when depositing a library with a single edge mask. When the double edge mask is employed for ternary composition spread fabrication, the shadow masks move as shown in Figure 9(a–c). The composition gradients are rotated by 120° with respect to each other, as shown in Figure 9(d–f). Figure 9(h) illustrates the shadow mask pattern when double edges are used, suggesting that the total mask length is shorter but the deposition time is longer than with a single edge mask in Figure 8(h). The choice of the mask depends partly on the desired deposition time and deposition rate. This is particularly important for the PLD growth technique, which may set limits on the number of discrete deposition pulses and pulse rates that give optimal film growth but may be less than optimal for operating masks.

The ternary composition spread technique has been demonstrated for ReCa$_4$O(BO$_3$)$_3$ system where Re is a rare
earth metal, Tb, Sc, or Pr. GdCa₄O(BO₃)₃ and YCa₄O(BO₃)₃ are known as nonlinear optical materials with transmittance in the range of 210–2600 nm, a high damage threshold, and good chemical stability [32]. Tb atoms at the Re site are known to induce luminescence in the green region of the spectrum, at 542 nm as a result of a ⁵D₄ → ⁷F₅ transition under ultraviolet excitation. A ternary composition spread library of epitaxial 300-nm Tb₀.₆Sc₀.₄Pr₀ₓCa₄O(BO₃)₃ films grown on a YCa₄O(BO₃)₃ (100) substrate is shown in Figure 10. The photoluminescence intensity of the composition spread film was investigated by taking a color photograph of the library under ultraviolet excitation at 254 nm. The intensity of white light emission is shown in Figure 10. A composition region with the highest light emission intensity was identified in the luminescence intensity map. The Tb₀.₆Sc₀.₄Pr₀ₓCa₄O(BO₃)₃ region had the brightest emission, and there was no positive effect of Pr in this system. A similar photoluminescence intensity dependence on the composition was confirmed in the conventional photoluminescence spectra of ⁵D₄ → ⁷F₅ emission (542 nm) excited with a frequency-doubled Ar laser (488 nm). The ternary composition spread technique has also been used for high-throughput screening of many materials systems, including dielectrics HfO₂–Y₂O₃–Al₂O₃ [15], photocatalysts Sr(Ti, Cr, V)O₃ [35], phosphors Eu₁₋ₓScₓPrₓCa₄O(BO₃)₃ [36], (Eu, Tm, Tb)-doped Y₂O₃ [33], high-mobility oxide semiconductors SrTiO₃–LaAlO₃–LaTiO₃ [37], and flux materials [38–40].

3.3. Superlattice Libraries

To carry out parallel fabrication of layered structures by an atomic layer-by-layer process, combinatorial laser molecular beam epitaxy (CLMBE) was developed [12]. The CLMBE allows concurrent fabrication of a number of artificial lattices or heterojunctions where the composition and thickness of each layer and the deposition sequence are manipulated at an atomic scale. This represents a considerable improvement in the ability to design new materials systems: libraries can now be constructed where
arrangements of atoms are essentially controlled in three dimensions. Three key features of a CLMBE system are illustrated in Figure 11. It is a unique combinatorial synthesis tool, as it includes a set of physical masks for defining the film growth site, a scanning RHEED system [14] for in situ diagnostics of film growth mode at various sites during the growth, and a fiber-guided Nd:YAG laser for substrate heating. By monitoring the growth mode with RHEED and by synchronizing the target switching with mask movements, the synthesis of a number of atomic-layer controlled materials can be coordinated on a single substrate. This form of lattice engineering was first verified in the simultaneous fabrication of perovskite oxide (ABO$_3$) superlattices. In such an experiment, clear intensity oscillation in scanning RHEED, each of which corresponds to the growth of a unit-cell layer, is observed on the surfaces defined by the mask pattern in Figure 11(b). The RHEED beam is scanned across the substrate surface by a pair of coils, and the scanning and image acquisition are synchronized with the moving of shadow deposition masks to control the growth of as many as ten thin film strips in parallel.

Figure 12 (a) and (b) show concurrent XRD patterns for a library of ten [(SrTiO$_3$)$_n$/BaTiO$_3$)$_{30}$ superlattices with
$n = 12, \ldots, 30$. In addition to the specular reflection band near 0.70° and the SrTiO$_3$(001) reflection near 22.75°, many satellite peaks are visible due to the superstructure. The positions of the reflection peaks agree well with the simulated positions and show that the actual periodicities of the superlattices match the design.

### 3.4. Temperature-Gradient Epitaxy

Besides mapping phase diagrams, combinatorial thin film growth can also be used for rapidly optimizing the growth of known phases. The obvious advantage of using parallel growth techniques is the increased experimental throughput. However, another very important advantage is the increased reliability of the data from a growth parameter optimization experiment. For example, in a typical one-by-one experiment procedure, a sequence of samples may be grown to study the effects of film microstructure on electronic properties, such as breakdown fields of thin film capacitors. Such measurements generally have large error margins, largely because the electronic parameters are very sensitive to minute changes in the film microstructure. As each deposition run is slightly different, a sequence of individual samples often results in a large scatter of data points. In a combinatorial experiment, all samples within a single library are grown in a single deposition run on a single substrate. For growth parameter mapping, such as growth temperature, only that single parameter is intentionally varied. This can lead to a significant improvement in the reliability of the characterization data, which means that a smaller number of samples are sufficient, and all samples can be prepared in a single experiment.

Temperature is the most common parameter used for controlling crystal growth, either in bulk or thin film form. Mapping film microstructures, growth modes, and physical properties as a function of the growth temperature is therefore a very common and important task. A parallel technique for generating a lateral temperature gradient over a library surface during thin film growth has been achieved by using a special sample holder, illustrated in Figure 13 [16]. A continuous wave Nd:YAG laser is used to heat the sample holder, with the laser beam focused on the backside of the free end of the cantilever-like center part of the sample holder. A $5 \times 15$-mm$^2$ substrate is attached to the central part of the holder with silver paste. The heating laser beam is focused to a 3-mm diameter spot on the free-standing end of the sample holder while heat is lost mainly by conduction through the other end, which is attached to the main sample.

![Figure 13](image)

**FIGURE 13** (a) Schematic illustration of a temperature-gradient sample holder for mapping growth temperature effects on film properties. (b) Variation of local temperature on the sample surface as a function of position. (c–f) AFM images before (c) and after the growth of (La,Sr)MnO$_3$ epitaxial thin films at 650 °C (d), 750 °C (e), and 900 °C (f).
3.5. Combinatorial Screening of Flux Materials for Single-Crystal Film Growth

Flux materials are often used for the synthesis of compounds that have high melting temperatures and/or decompose incongruently. A flux is needed mainly to improve the crystallinity by reducing the growth temperature and suppressing thermodynamic phase separation in bulk crystal syntheses [36]. The presence of flux materials have been found useful in improving the crystallinity of many different films, for example, thin films grown by vapor phase epitaxy, such as the chemical vapor deposition growth of Si whiskers in the liquid phase by the vapor—liquid—solid (VLS) method [43]. The VLS process has recently been extended to triphase epitaxy (TPE) [44,45] of oxides, using PLD to deposit a liquid layer and film precursors from gas phase by laser ablation of stoichiometric solid targets, as shown in Figure 14. The advantage of using liquid phase growth for oxide thin films has been demonstrated by the growth of near-perfect single-crystal thin films of cuprate superconductors, NdBa2Cu3O7−δ [44] and Ca-doped NdBa2Cu3O7−δ [45]. The main advantage of TPE for superconductor growth is the lack of grain boundaries and improved crystallinity due to the reduction of misfit dislocations. Flux-mediated epitaxy, both VLS and TPE, are therefore receiving more attention as methods of fabricating very high-quality oxide thin films.

The growth of Y-type magnetoplumbite (Ba2Co2−Fe2Y2O22:Co2Y) thin films is a good demonstration of the capabilities of flux-mediated epitaxy [46–48]. Magnetic oxides can be used as core materials for thin film inductors and other microwave devices operated in the GHz frequency range. The synthesis of magnetoplumbite thin films is complicated by incongruent melting that leads to a decomposition into BaFe2O4 and Co2Y. If magnetoplumbite films are deposited by conventional methods from stoichiometric target, a segregated BaFe2O4 impurity phase grows on the substrate. To suppress the formation of the BaFe2O4 impurity phase, flux-mediated epitaxy has been used with the aid of CoO self-flux. The CoO self-flux layer is first grown directly on an MgAl2O4(111) substrate. Subsequently, a stoichiometric magnetoplumbite film can be grown through the predeposited flux layer. Figure 15(a) schematically depicts the cross-sectional library structure for combinatorial optimization of the CoO flux layer thickness from 800 to 1,000 Å. Concurrent XRD patterns for the peaks of the impurity phase and Co2Y are shown in Figures 15(b) and (c), respectively. The impurity peaks completely disappeared, and the Co2Y peak intensity increased when the thickness of the CoO layer exceeded 850 Å, reaching a maximum for a 950 Å CoO flux layer thickness. By optimizing the thickness, a single-crystal

![Figure 14](image.png)

**Figure 14** Schematic illustration of conventional thin film growth (a) and the flux-mediated epitaxy (b−d). In flux-mediated epitaxy, a liquid flux layer is deposited on the substrate (b). Subsequently, the film material is deposited through the liquid flux layer. The film nucleates and grows at the substrate interface, resulting in better crystallinity than in direct gas-phase growth (c). The film growth continues at the flux−film interface until the desired film thickness has been reached (d).
magnetoplumbite film could be grown on an MgAl2O4(111) substrate, as shown by transmission electron microscopy and XRD analysis [46–48].

4. DISCOVERIES MADE BY COMBINATORIAL TECHNOLOGY

4.1. Quantum Size Effect in Photocatalytic TiO2

Photocatalysis has attracted considerable attention since the discovery of the Honda–Fujishima effect [49] due to the possibility of hydrogen fuel production by water splitting and photodeposition of various materials. To optimize the catalytic properties, it is essential to understand the surface reaction on a nanometer scale. For this purpose, photocatalytic properties of epitaxial TiO2/Nb:TiO2 heterostructures were investigated [50]. The use of heterostructures in catalysts has the advantage of eliminating most of the surface morphology effects on catalytic activity, because the uniform atomically flat film surfaces are nearly identical to the original single-crystal substrates. Another advantage of thin films over powder catalysts is the ability to accurately control the layer thickness, which corresponds to the particle size of a conventional powder catalyst.

Most of the structural parameters of heterostructures, in particular layer thicknesses, can be easily mapped in combinatorial experiments. To study the process of photocatalysis on a titania surface, TiO2 thin films have been fabricated on atomically smooth Nb:TiO2(110) (Nb:0.5 wt %) single-crystal substrates by PLD [51], which allows for accurate in situ growth condition optimization by real-time RHEED intensity oscillation monitoring. Films with flat surfaces normally show sharp RHEED patterns and a step and terrace surface morphology that is equivalent to that of the substrate. Thickness gradient films can be used to investigate the effect of the surface layer thickness on the photoactivity. A convenient way for measuring the absolute activity and spatial variation of catalytic activity on an oxide surface is to use the photochemical deposition of Ag particles. When thin film samples are immersed in a 0.01 N AgNO3 electrolyte under ultraviolet illumination, as illustrated in Figure 16(a), the amount of photodeposited Ag particles can be evaluated by scanning X-ray fluorescence (XRF), giving an accurate measure of the catalytic activity with a spatial resolution of the XRF probe. A photograph of the catalyst surface after a photochemical reaction is shown in Figure 16(b). The bright white line close to the left edge of the library shows that the photodeposited Ag particles accumulate in a narrow thickness region of the TiO2 cap layer. Figure 16(c) shows the Ag concentration on the surface as a function of the TiO2 layer thickness. A strong photoactivity peak can be seen at a film thickness of about 5 nm, with a full width at half maximum of just 3.9 nm. By conventional one-by-one process, it is difficult to find such a narrow thickness range, which is the only region where strong photoactivity occurs in this material system. The combinatorial technique is very efficient at finding such narrow regions of interest in broad phase diagrams and is therefore suitable for the detection of photocatalytic size effects in single-crystal heterostructures. The same technique has also been used to study (Ti,V)O2 [50], (Ba,Sr) TiO3 [52], and Sr(V, Cr, Ti)O3 [35] systems and confirmed to be useful for the investigation of the film thickness effects as well as the film composition dependences.

4.2. Room-Temperature Transparent Ferromagnetism in Co-Doped TiO2

A class of ferromagnetic semiconductors that are obtained by doping magnetic impurities into host semiconductors is one of the key materials for spintronics in which the correlation between charge and spin of electrons is used to bring about spin-dependent electronic functionality such as giant magnetoresistance and a spin field effect transistor switching. Among the materials that had been reported
until 2001, Mn-doped GaAs had the highest Curie temperature, \( T_c \approx 100 \, \text{K} \), and is still presumed to be a promising candidate, with a gradual increase of \( T_c \) up to 250 K \([53]\), for practical applications. However, there had been no report on any ferromagnetic semiconductor in oxides at that time.

An exploration of a novel ferromagnetic oxide semiconductor was thus a challenge to verify the capability of our combinatorial thin film method. ZnO \([54]\), rutile, and anatase TiO\(_2\) \([21,25,55]\), the latter two phases of which could be controlled by using a different lattice-matched substrate of sapphire (0001) or SrTiO\(_3\)(001), were selected as a host semiconductor, and each was systematically doped with all the 3d transition metal ions. The “one-month combinatorial doping experiment” yielded 243 distinct samples as shown in Figure 17, and the solubility limits of each different transition metal ion in anatase, rutile and ZnO films were quickly determined from XRD measurements \([54,55]\).

For a quick screening of magnetic properties of these 3d-transition metal-doped TiO\(_2\) and ZnO film libraries, a scanning SQUID microscope technique was employed, leading to a serendipitous discovery of room-temperature transparent ferromagnetism in both of Co-doped anatase and rutile TiO\(_2\) films \([21,22]\). Figure 18 shows a series of scanning SQUID microscope images taken at 3 K for anatase thin films with different Co contents of 0\% (a), 3\% (b), and 6\% (c) of Co. In all of the Co-doped TiO\(_2\) films, magnetic domain structures with an approximately 20-\( \mu \text{m}\) size can be seen, while none of the nondoped TiO\(_2\) films show magnetic structure within an experimental error of around 0.5 \( \mu \text{T}\). With an increase of the Co content in the film, the magnitude of the observed magnetic field is also enhanced as a result of the increased spontaneous magnetization. These observations are evidence for the emergence of a ferromagnetic long-range order in the Co-doped TiO\(_2\), irrespective of the crystal phase being rutile or anatase. The \( T_c \) was estimated to be >400 K at that

![Figure 17](image)

**FIGURE 17** Combinatorial fabrications of TiO\(_2\) and ZnO films doped with all 3d transition metal ions.
time. Figure 18(d) is an Ultraviolet—Visible spectrum for a Co-doped anatase thin film, together with a photograph of the sample. The Co-doped TiO$_2$ film is transparent in the visible and near-infrared regions, exhibiting a band gap at 400 nm (3.1 eV).

On the other hand, no signs of such a magnetic domain structure were observed in any other combinatorial TiO$_2$ and ZnO film libraries, though some research groups claimed that Co-doped ZnO also became ferromagnetic, which is still somewhat controversial. In fact, to date, there have been reports of many types of ferromagnetic oxide semiconductors including Co-doped TiO$_2$ and ZnO, none of which has, however, been commonly accepted as a ferromagnetic semiconductor. In this controversial situation, it is quite recently that hydrogen was found to play a key role in the emergence of ferromagnetism in Co-doped ZnO [56] and the origin of ferromagnetism in Co-doped TiO$_2$ was verified to be intrinsic and rationalized by a scheme of carrier-induced ferromagnetism. Fukumura and coauthors have successfully demonstrated that the magnitude of ferromagnetism in Co-doped TiO$_2$ can be well scaled by the carrier density, which can be chemically controlled by the degree of oxygen deficiency and also physically modulated by the field effect even at room temperature [57–61]. Therefore, Co-doped has now become one of the most promising and representative ferromagnetic semiconductors for a possible realization of transparent semiconductor spintronics devices operable at room temperature.

### 4.3. Growth of Single-Crystal Oxide Films by Flux-mediated Epitaxy

The key point in the successful application of flux-mediated epitaxy is the selection of an appropriate flux material. It is not sufficient to examine the well-known fluxes predicted from the bulk phase diagrams, because thin film growth conditions are definitely different from those of the bulk process. In fact, the volatile Bi$_2$O$_3$ self-flux of Bi$_4$Ti$_3$O$_{12}$, which is predicted from the phase diagram of Bi$_2$O$_3$ and TiO$_2$, has proven to be useless for the Bi$_4$Ti$_3$O$_{12}$ thin film growth [38]. It is therefore necessary to find a suitable impurity flux individually for each case and without the ability to rely on obvious guiding principles that are known for flux composition selection in the case of corresponding bulk crystal growth. Owing to the large diversity of possible flux materials, this is another area where combinatorial parallel synthesis can greatly speed up the optimization process. It is inevitably laborious and time consuming to optimize the composition of the flux materials in a conventional one-by-one procedure.
The combinatorial approach offers a novel and unique way to overcome the throughput problem in flux optimization. To explore a flux material for Bi$_4$Ti$_3$O$_{12}$, addressable ternary composition spread libraries were fabricated by the use of a combinatorial PLD system. The libraries were composed of Bi$_{4-x}$Ti$_3$O$_{12}$ (0 < x < 3) self-flux and impurity fluxes. Candidates of impurity fluxes were oxides of V, W, Cu, Mo, Bi—P, and Ba. All these oxides are well-known flux materials for the growth of various oxide bulk crystals.

Figure 19 schematically illustrates the combinatorial process of the flux material search for Bi$_4$Ti$_3$O$_{12}$ thin film growth [38,39]. At first, a ternary composition spread of the flux layers was deposited on a SrTiO$_3$(001) substrate, followed by a Bi$_4$Ti$_3$O$_{12}$ film grown through the flux layer. After growth, the crystallinity of the Bi$_4$Ti$_3$O$_{12}$ film in the library was characterized by concurrent XRD [19], by taking several cuts through the ternary phase diagram library and combining the XRD data in a crystallinity phase diagram plot, as shown in Figure 19(b).

Figure 20(a) shows an XRD pattern of a Bi$_4$Ti$_3$O$_{12}$ thin film prepared in a conventional one-by-one process. The Bi$_4$Ti$_3$O$_{12}$ thin film on an SrTiO$_3$(001) c-axis oriented. Using concurrent XRD, all the c-axis lattice parameters and the Bi$_4$Ti$_3$O$_{12}$ (0014) peak intensities were mapped over the entire library area for 500-nm-thick Bi$_4$Ti$_3$O$_{12}$ films, including a Bi$_4$Ti$_3$O$_{12}$ thin film grown without a flux. By plotting the c-axis parameter for each measurement point vs. the (0014) peak intensity, it is possible to evaluate the crystallinity of the film as a function of the flux composition. The results of such a mapping are shown in Figure 20(b—h). Each ternary library contained 5,000 measurement points in a triangle with 8 mm sides. A total of 35,000 data points were collected to find the best flux composition for Bi$_4$Ti$_3$O$_{12}$ film growth. Although the c-axis parameters and peak intensities showed slight fluctuations from one position to another even in the case of a Bi$_4$Ti$_3$O$_{12}$ film grown without a flux, detailed statistical analysis of the flux library made it possible to establish a clear tendency: the series of Bi$_4$Ti$_3$O$_{12}$ films grown by using a CuO-containing flux showed higher X-ray peak intensities than any other library, while the c-axis lattice parameter values approached the bulk Bi$_4$Ti$_3$O$_{12}$ value of 3.283 nm. This combinatorial screening suggested that a CuO-containing flux would be the best candidate to assist Bi$_4$Ti$_3$O$_{12}$ film growth. After further optimization of several growth parameters, stoichiometric and very high-quality Bi$_4$Ti$_3$O$_{12}$ single-crystal films could finally be grown in a CuO—BiO$_3$ flux [38].

Figure 21 shows a comparison of transmission electron microscope (TEM) images of Bi$_4$Ti$_3$O$_{12}$ films grown without and with a CuO-containing flux. Without using the CuO flux, many out-of-phase boundaries originating from the step structure of the substrates could be observed. In contrast, no boundaries were seen in a high-quality Bi$_4$Ti$_3$O$_{12}$ film grown with a flux containing CuO. It appeared that the CuO flux suppressed boundary growth at the substrate steps by maintaining thermodynamic equilibrium conditions across substrate step edges. Figure 21(c) shows an AFM image of a Bi$_4$Ti$_3$O$_{12}$ film grown with the CuO flux. Atomically smooth terraces can be seen with all steps having a uniform 1.6-nm height, corresponding to half a unit of the Bi$_4$Ti$_3$O$_{12}$ crystal structure. The large terrace width is a direct result of the equilibrium conditions prevailing at the liquid—solid interface in the flux [62]. The uniformity of the step height can be attributed to the uniformity of the surface termination layer, which was determined to be the (Bi$_2$O$_2$)$_{2+}$ layer from cross-sectional high-resolution TEM analysis of the surface. AFM images of flux-grown Bi$_4$Ti$_3$O$_{12}$ films showed no 0.4-nm step structures originating from out-of-phase boundaries, supporting the TEM analysis, which also failed to find such boundaries. The effect of decreased structural defect density can be seen in the leak currents of the two types of films. Figure 21(d) shows a comparison of the leak current density, and it is clear that flux-mediated growth resulted in a drop of the leak current density from $10^{-3}$ to $10^{-4}$ A/cm$^2$ to $10^{-7}$ to $10^{-8}$ A/cm$^2$. The leak current density improvement is also related to reduced Bi deficiency, based on secondary ion mass spectrometry (SIMS) analysis. Without using a CuO flux, SIMS measurements showed a high Bi deficiency in the films. In contrast, the composition of flux-grown Bi$_4$Ti$_3$O$_{12}$ single-crystal films was very close to stoichiometric: the
FIGURE 20 (a) XRD pattern of a Bi$_4$Ti$_3$O$_{12}$ thin film fabricated by a conventional one-by-one process. Closed and open circles mark the SrTiO$_3$ substrate and Bi$_4$Ti$_3$O$_{12}$ film, respectively. (b–h) Mapping of the c-axis parameter vs. the Bi$_4$Ti$_3$O$_{12}$ (0014) peak intensity for 35,000 (5,000 × 7) different flux compositions. From a statistical point of view, Bi$_4$Ti$_3$O$_{12}$ films grown in the presence of a CuO-containing flux had the highest X-ray intensities and the closest lattice parameter to the bulk value.

FIGURE 21 Cross-sectional TEM images of the Bi$_4$Ti$_3$O$_{12}$ film grown with no flux (a) and in the presence of a CuO-containing flux (b). Many out-of-phase boundaries running across the Bi$_4$Ti$_3$O$_{12}$ film can be seen in (a), while none are visible in (b). (c) AFM image of a Bi$_4$Ti$_3$O$_{12}$ film grown by flux-mediated epitaxy. (d) Leak current density vs. applied electric field for Bi$_4$Ti$_3$O$_{12}$ films prepared with and without a flux.
films had Bi concentrations corresponding to a Bi$_4$Ti$_3$O$_{12}$ powder.

Combinatorial screening sometimes yields serendipitous discoveries [40]. As a result of accumulating 35,000 flux composition data points, a CuO-based flux was found to assist Bi$_4$Ti$_3$O$_{12}$ thin film growth, with the crystallinity of the film being the driving parameter for the optimization process. However, it is known that different flux materials can also affect other aspects of crystal growth, such as crystal orientation, besides crystallinity. Because the XRD data sets included more information than just the crystallinity data, the data set of 6 ternary composition spread libraries was reanalyzed. The concurrent XRD mappings showed that only the V oxide flux library showed a shoulder peak around the SrTiO$_3$(002) substrate reflection, suggesting the growth of an unexpected phase. Moreover, the same library color was visually yellowish in the V oxide rich region. To understand these observations, a conventional powder XRD pattern was taken from an entire region of the ternary composition spread library containing V oxide, as shown in Figure 22(a). In addition to the c-axis oriented phase, (110)-oriented Bi$_4$Ti$_3$O$_{12}$ was found to be present in the V oxide flux library but not in any other libraries. This was very surprising, because Bi$_4$Ti$_3$O$_{12}$ films grown directly on the SrTiO$_3$(001) surface are known to have a unique c-axis orientation. In order to look further into this intriguing phenomenon, the (220) peak intensity was mapped over the entire library by concurrent XRD, as shown in Figure 22(b). For most flux compositions, the (220) peak was not detected, with the (00l) peaks from the growth of c-axis oriented Bi$_4$Ti$_3$O$_{12}$ being dominant. Only the V oxide flux library exhibited elevated (220) peak intensity, with the highest intensity in the pure V oxide flux region in the entire library shown in Figure 22(b). From this first screening, it was concluded that V oxide may be a promising flux to assist orientation control from the common (001) direction to the (110) direction during the growth of Bi$_4$Ti$_3$O$_{12}$ on the SrTiO$_3$(001).

To understand the relationship between a flux containing V oxide and Bi$_4$Ti$_3$O$_{12}$(110)-oriented growth, the V oxide film thickness dependence was investigated by preparing a combinatorial library illustrated in Figure 23(a). A pure V oxide flux film was deposited with a thickness gradient of 0–50 nm directly on a SrTiO$_3$(001) substrate by moving a shadow mask during deposition. Subsequently, a 500-nm-thick Bi$_4$Ti$_3$O$_{12}$ film was deposited on the V oxide thickness gradient layer. Scanning X-ray microdiffraction analysis was used for mapping the X-ray peak intensity as a function of the 2θ angle in the 37°–49° range as a function of the V oxide thickness flux layer thickness (0–50 nm). It was found that even a small amount of V oxide was sufficient to promote the growth of a (110)-oriented Bi$_4$Ti$_3$O$_{12}$ film, and the (220) peak intensity gradually increased in accordance with the increase of the V oxide flux layer thickness. The preferential growth of the (110)-oriented Bi$_4$Ti$_3$O$_{12}$ film was therefore attributed to the V oxide flux effect. Here, the (0014) and (220) peak intensities were evaluated from a Gaussian fitting of the XRD mapping data, as shown in Figure 23(b), and plotted as a function of the sample position in Figure 23(c). It was found that a 30-nm V oxide layer is

![Figure 22](image-url)
A cross-sectional scanning electron microscope (SEM) image of a (110)-oriented Bi$_4$Ti$_3$O$_{12}$ film grown on a Nb-doped SrTiO$_3$(001) substrate is shown in Figure 24. Many rectangular skyscraper-like Bi$_4$Ti$_3$O$_{12}$ platelets with a characteristic size of 1,000 × 1,000 × 100 nm$^3$ grew all over the substrate. Each platelet has two equivalent in-plane orientations along the (100) direction of the substrate. The in-plane growth orientation of the Bi$_4$Ti$_3$O$_{12}$ nanoplatelets was investigated by the use of transmission electron microscopy. Figure 24(b) shows diffraction patterns of an isolated Bi$_4$Ti$_3$O$_{12}$ nanoplatelet, observed along directions normal to the narrower or wider sides. From these observations, the in-plane relationship between the Bi$_4$Ti$_3$O$_{12}$ platelets and the substrate was determined as follows:

$$\begin{align*}
\text{Bi}_4\text{Ti}_3\text{O}_{12} \parallel & \text{SrTiO}_3(010)(100), \\
\text{Bi}_4\text{Ti}_3\text{O}_{12} \parallel & \text{SrTiO}_3(100)(101) \\
\end{align*}$$

That is, the Bi$_4$Ti$_3$O$_{12}$ nanoplatelet crystal was found to grow epitaxially on the SrTiO$_3$ substrate, as illustrated in Figure 24(b). The epitaxial growth of the (110)-oriented Bi$_4$Ti$_3$O$_{12}$ film was also found on other substrates, such as LaAlO$_3$(001) and (LaAlO$_3$)$_{0.3}$(Sr$_2$AlTaO$_6$)$_{0.7}$(001). The piezoelectric property of an isolated epitaxial nanoplatelet was examined by scanning probe piezomicroscopy. The lattice displacement and the effective $d$ value have been measured along the [110] direction by applying a voltage. Figure 24(c) and (d) show a typical butterfly curve and a hysteresis loop, implying that the oriented platelets are ferroelectric.

FIGURE 23 The sample position dependence: (a) V oxide containing flux layer thickness, (b) XRD profile, (c) the integrated XRD intensities of Bi$_4$Ti$_3$O$_{12}$ (004) and (220) peaks.

FIGURE 24 (a) Cross-sectional SEM image of Bi$_4$Ti$_3$O$_{12}$(110) orientation nanoplatelets on a conductive Nb-doped SrTiO$_3$(001) substrate. (b) Schematic drawing of the epitaxial orientation of a Bi$_4$Ti$_3$O$_{12}$ nanoplatelet, and diffraction patterns of wide and narrow planes. (c) and (d) Piezoelectric displacement and $d_{eff}$ as a function of the bias voltage.
5. HIGH-TECH VENTURE: COMBINATORIAL MATERIAL BUSINESS

Combinatorial technology is a powerful tool to bring about innovation in many kinds of materials development, not only oxides typically represented and described above but also nonoxide ceramics, metals and alloys, semiconductors, polymers, and composites. It has also been applied to high-throughput research and development of layer-structured materials such as superlattices and heterojunction devices as well as for interface phenomena. Conventional sputtering, PLD, vacuum evaporation, and MBE can be adapted to a combinatorial model by inserting a scheme of shadow masking to enable the fabrication of a material chip on which a variety of different materials are grown in a single-run experiment. Although the spatial resolution of library cells is not as high as in a high-vacuum process, combinatorial technology can further be extended to other dry (chemical vapor deposition, etc.) and wet (inkjet, sol-gel, spray-pyrolysis, melt-quench, etc.) material processes. Matching of machine performance and experimental design can improve the research throughput by a factor of \( \frac{C}{21} \) as compared with the conventional one-by-one process.

The introduction of combinatorial function does not necessarily require special machines and/or expensive attachment to existing equipment. Nevertheless, increasing interest in advanced materials of complex composition and nanostructures inevitably ask us to screen a wider space of compositions and preparation conditions in a time span that is as short as possible. To fulfill the increasing needs for higher performance combinatorial material technology, it is desirable to have specially designed and computer-controlled easy-operation hardware, navigators, and skilled operators who can extract the full power of the machine. This relationship is comparable to that of a racing team composed of a tuned-up racing car, talented driver, and professional mechanics.

Venture companies using combinatorial methods as a business tool emerged in the US and Europe at the end of the 1990s. Represented by Symyx and Wildcat, they were focusing on the discovery of new catalysts and on the informatics to supply software for handling large quantities of data acquired from combinatorial synthesis and characterization. In 2004, another type of venture company, Intermolecular Inc., started business, focusing solely on electronic materials. The main customers targeted are electronic companies to substitute their R & D division for Intermolecular in developing new industrial products by using a 300-mm wafers as standard substrates because modern large-scale integrated (LSI) circuit industries use 300-mm wafers for LSI manufacturing.

In Japan, on the basis of fundamental combinatorial material research supported by Japan Science and Technology Agency for 1996–2001 and jointly by National Institute for Material Science (NIMS) and Tokyo Institute of Technology (TITECH) for 1999–2005, COMET Inc. was

**FIGURE 25** Combinatorial materials R & D Venture (Comet Inc.).
founded in 2007 as a NIMS authorized venture company [63]. COMET, which stands for Combinatorial Materials Exploration and Technology has the main office inside NIMS, located in Tsukuba, Japan, and a branch in Kashiwa campus of Tokyo University, about 20 km south of Tsukuba. According to the business model, schematically illustrated in Figure 25, COMET provides a service of new material discovery on the request mainly to industry and also academia in the public sector. The laboratory is equipped with originally designed combinatorial thin film deposition systems including several types of pulsed laser MBE and multitarget (~6) sputtering for inorganic solid-state materials and an infrared laser MBE for organic and soft matter. Recently installed combinatorial sputtering equipment handles a 4-inch wafer with 6 cathodes:3 cathodes for 2nd ternary matrix elements and the other 3 cathodes for 1st sputtering guns for doping or radical sources. Combinatorial cluster chambers were designed for all-in-one process from library fabrication to characterization without taking out (exposure to air) the sample, which is useful for device development. These original and patented combinatorial machines are on sale from COMET. In addition to the basic electronic property measurement in COMET, advanced structural analyses and property characterization are done in cooperative research laboratories in NIMS, Tokyo University, and TITECH. The focus of combinatorial materials research targeted by current activities at COMET covers energy saving, high-efficiency energy conversion, and energy storage technologies. From this point of view, we hope that combinatorial high-throughput technology not only opens new business opportunities but also serves the needs of all humanity by rapidly bringing new materials to the level of commercial use.

6. CONCLUSION

Combinatorial materials technology can be used to develop a wide variety of functional materials and for device process optimization. With increasing interest in nanostructured and smart (e.g. multiferroic, spintronic) materials, the combinatorial library design has been expanded from one-dimensional (binary composition spreads, temperature gradients) to two-dimensional (ternary phase diagrams, composition + preparation parameters) and three-dimensional (superlattices, heterojunction device, etc.) cases. Ceramic materials show a variety of characteristics, such as superconductivity, ferromagnetism, ferroelectricity, and light emission. Combinatorial techniques have contributed to the exploration of new ceramic materials as well as the optimization of the process parameters and discovery of new material properties. Combinatorial strategy with flexible consideration and ingenuity will continue to spread and is becoming a powerful tool for quickly solving urgent materials-related tasks that humanity is facing, such as energy efficiency, atmospheric pollution, and the emission of greenhouse gases. Combinatorial research can contribute to solving many of these problems, for example, by helping with the design of catalysts and energy harvesting devices.

ACKNOWLEDGMENTS

The research in this article was supported in part by JST (Japan Science and Technology agency)-CREST (Core Research for Evolution of Science and Technology) program and by PNU (Pusan National University)-WCU (World Class University) program on Cogno-Mechatronics Engineering through the National Research Foundation of Korea (#R31-20004).

REFERENCES

Stereo Fabric Modeling Technology in Manufacturing Ceramics

Hideki Kita

Department of Molecular Design and Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8603, Japan, Former Affiliate: National Institute of Advanced Industrial Science and Technology

Advanced ceramics are superior to conventional materials and, for this reason, have been adopted throughout the materials processing industry, including for steel, aluminum, chemicals, and semiconductors.

There has been a strong and continuous growth in user confidence due to wide confirmation of the energy savings, increased productivity, and reduced maintenance requirements with the use of these materials. Thus, to match the growing demand, there is also a strong thrust toward scaling up ceramic parts while continuing to improve upon performance and simultaneously lowering costs. Take, for example, ceramic parts used in the nonferrous casting industry, where low heat loss and reduced wettability are key performance parameters. For such parts, the use of a lightweight design can greatly improve productivity. In addition, to increase panel size in the liquid crystal display industry and to increase wafer size, and thus throughput, in the semiconductor industry, it is necessary to realize a corresponding increase in the sizes and specific rigidity of the ceramic members used for microfabrication.

Generally, in the case of structural parts, the performance of the part is determined by its geometrical design. This implies that greater flexibility in the possible geometry of parts will allow to greater functionality. The techniques currently used in ceramic modeling include extrusion molding, injection molding, and press molding, depending on the shape, size, and required precision of the intended components. It is, however, difficult to fabricate a large, complex-shaped component with a very high precision, and this restricts the scope of possible designs for production.

For these reasons, we have been developing a new modeling concept to simplify the fabrication of complex-shaped and large ceramic components. The concept involves the integration of multiple compact, highly

![FIGURE 1Stereo fabric concept. For color version of this figure, the reader is referred to the online version of this book.](image-url)
functional, and hollow units — such as LEGO blocks — to improve the energy efficiency of manufacturing plants and also product quality (Figure 1). The key feature is that it increases the degree of freedom in the structural design of ceramics, particularly for large ceramic components. The concept is referred to as stereo fabric modeling [1].

It is possible to produce a wide variety of large components by integrating small units having complex shapes. The expected effects are as follows:

1. Hollow structure → Weight reduction
2. Decreased material and energy consumption
3. Downsizing of manufacturing facilities
4. Scope for novel shapes and structures

Several such units have so far been designed. Figure 2 shows a few prototype models that can be built using the stereo fabric concept. Several kinds of ceramic members used in manufacturing systems were considered and model parts were fabricated with the proposed concept.

Figure 3 presents an example of a large, yet light, board with small surface protrusions of diameter 0.5 mm and height 0.68 mm. Each unit is hollow and has a truss structure to reduce weight and ensure rigidity. The protrusions reduce the contact area with the work piece, which in this case will be molten or semisolid metal.

The units have interlocking structures on adjacent faces, formed during the green stage, that allow them to be assembled very easily. After assembly during the green stage, the units are dewaxed and sintered in nitrogen gas. Adhesion tests were conducted using the sessile drop method on the final board. A barrel-shaped copper sample was placed near the center of the ceramic substrate with surface protrusions, and the sample—substrate system was inserted into the apparatus. As the temperature was gradually increased, the copper sample melted at 1100 °C, became spherical, and then stabilized. The sample was maintained at this temperature for approximately 10 min before the temperature was increased to the maximum temperature of 1185 °C. During this final stage, images of the sample were acquired using a CCD camera (Figure 4).

The stereo fabric modeling concepts constitute the core technology of the new Japanese National Project called “Innovative Development of Ceramics Production
Technology for Energy Saving project,” which was initiated in 2009 and will continue through to 2014. This project involves the development of joining techniques and near-net forming processes, aiming at reducing energy consumption throughout both the use and manufacture of ceramics.

It is hoped that through this project this stereo fabric concept will be used, in collaboration with commercial enterprises, to design a novel large-scale ceramic member for application to various fields and thus contribute to efficiency improvements in the manufacturing industry.

REFERENCES

Porous Ceramic Materials

Tatsuki Ohji
National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Chapter Outline
1. Introduction 1131
2. Partial Sintering 1133
3. Sacrificial Fugitives 1135
4. Replica Templates 1139
5. Direct Foaming 1140
6. Gas Permeability 1141
7. Summary 1142
References 1143

1. INTRODUCTION

Porous ceramics are now used for wide variety of industrial applications from filtration, absorption, catalysts, and catalyst supports to lightweight structural components. In these decades, a great deal of research efforts has been devoted for tailoring deliberately sizes, amounts, shapes, locations, and connectivity of distributed pores, which have brought improved or unique properties and functions of porous ceramics [1–12]. The merits in using porous ceramics for these applications are generally a combination of intrinsic properties of ceramics themselves and advantages of dispersing pores into them. The former include heat and corrosion resistances, wear and erosion resistance, unique electronic properties, good bioaffinity, low density, and high specific strength, and the latter are low density, low thermal conductivity, controlled permeability, high surface area, low dielectric constant, and improved piezoelectric properties [13,14].

This chapter intends to deals with the recent progress of porous ceramics. Porous materials are classified into three classes depending on the pore diameter, $d$: macroporous ($d > 50$ nm), mesoporous ($50$ nm $> d > 2$ nm), and microporous ($d < 2$ nm), according to the nomenclature of IUPAC (International Union of Pure and Applied Chemistry). Figure 1 shows this classification along with typical applications and fabrication processes specific to the pore diameters. One of the most representative applications of porous materials is filtration or separation of matters in fluids. Filtration is roughly classified into several classes depending on pore diameter, $d$, and molecular weight cut-off of the matters (MWCO): filtration (typically $d > 10$ μm), microfiltration ($10$ μm $> d > 100$ nm), ultrafiltration ($100$ nm $> d > 1$ nm, MWCO $= 10^3$–$10^6$), nanofiltration ($d \approx 1$–$2$ nm, MWCO $= 200$–$10^3$), and reverse osmosis ($d < 1$ nm, MWCO $\approx 100$). In filtration and microfiltration where pore size is relatively large, the separation is principally made by the sieving effect where matters whose size is larger than the pore size is trapped. On the other hand, in ultrafiltration, nanofiltration, and reverse osmosis where pore size is relatively small, fluid permeability depends on the affinity of solute and solvent to the porous materials as well.

Because of the great deal of research work reported in this field these days, this chapter mainly focuses on macroporous ceramics; micro- and mesoporous ceramics whose pore size is below $50$ nm are not included here. Representative applications of macroporous ceramics are briefly described. Ceramic filters are now widely loaded in diesel engines to trap particulate matters in the exhaust gas stream, so called, diesel particulate filters (DPFs). Since the high combustion efficiency and low carbon dioxide emission of diesel engines, the demand of DPF is also expected to further increase the world over [15–17]. Ceramic water purification filters are used for eliminating bacteria and suspension from wastewater, because of their higher flux capability, sharper pore-size distribution, better durability, and higher damage tolerance than those of organic hollow fibers [18]. Ceramic foam filters have been employed for removing metallic inclusions from molten metals such as cast iron, steel, and aluminum, as well as rectifying flow of
the molten metals [19]. Since the metallic inclusions result in defects in cast metals, this filtration process substantially improves the performance of the products. Porous ceramics with high specific surface area are adopted for absorptive and catalytic applications, where larger contact area with reactants is preferred, particularly in high temperature or corrosive atmospheres. Bioreactors are devices or systems that provide a biologically active environment, where microorganisms and enzymes are immobilized, and biochemical reactions are performed in porous beds, and porous ceramics are often used as such bioreactor beds due to chemical stability of ceramics and accommodative function of porous structure [20]. Recently, porous bioceramics with open-pore structures have attracted great attention for bioimplant applications including bone regeneration [21]. Bone cells are impregnated through the open pores and grow on their biocompatible walls resulting in bone ingrowth. Many electrodes used in electro-chemical devices including gas purifiers, gas sensors, fuel cells, and chemical analyzers are porous ceramics [22]. Some porous electrodes require two mode distributions of pore sizes; small pores are for the electro-chemical reactions while large ones are for flow paths of reactants. Properties of electroceramics also depend substantially on the porosity content and morphology and therefore porous ceramics are also applied or expected to be used in various electro-chemical devices. For example, porous piezoelectric ceramics show good piezoelectric property and are expected to be used for ultrasonic transducers, etc. [23] A variety of porous ceramics have been applied as materials for refractory bricks of kilns and furnaces in various industrial fields, due to their low thermal conductivity and high thermal shock resistance [24,25]. On the other hand, some porous materials of conductive ceramics like zirconia and silicon carbide have been utilized in heat exchangers and heaters [13].

As is known from Figure 1, the representative processes for making macroporous ceramics are (1) partial sintering, (2) sacrificial fugitives, (3) replica templates, and (4) direct foaming. A number of innovative techniques which have been developed recently for critical control of pores are introduced, divided into these four categories, together with some important properties of porous ceramics obtained in these processes. It should be noted however, that many new approaches for macroporous ceramics such as phase separations [26–30] have been developed other than the processes shown here. Figure 2 shows schematic illustrations of these processes, each of which will be interpreted in its section. We then discuss gas permeability of these porous ceramics with different pore sizes and structures in

![FIGURE 1](image1)

**FIGURE 1** Classification of porous materials by the pore size and corresponding typical applications and fabrication processes.

![FIGURE 2](image2)

**FIGURE 2** Representative fabrication processes of macroporous ceramics.
Section 6. Finally, a summary will be given with the issues to be solved for further realizing the potential of porous ceramics and for expanding their applicability.

2. PARTIAL SINTERING

Partial sintering of powder compact is one of the most conventional and frequently employed approaches to fabricate porous ceramic materials. Particles of powder compact are bonded due to surface diffusion or evaporation–condensation processes enhanced by heat treatments, and a homogeneous porous structure forms when sintering is terminated before being fully densified (see Figure 2 (a)). Pore size and porosity can be controlled by the size of starting powders and degree of partial sintering, respectively. Generally, in order to provide the desired pore size, the size of raw powder should be geometrically in the range two to five times larger than that of the pore. Porosity decreases with increased forming pressure, sintering temperature, and time. In addition, processing factors such as the type and amount of additives, green densities, and sintering conditions (temperature, atmosphere, pressure, etc.) also greatly affect the microstructures of porous ceramics [31]. The mechanical properties depend significantly on degree of neck growth between grains, as well as porosity and pore size. For example, the formation of necks between touching particles by surface diffusion without densification can increase the elastic modulus to 10% of the fully dense value [32,33]. The porosities of porous materials obtained by partial sintering are usually below 50%. In industry, this method has been utilized for various applications including molten metal filters, aeration filters (gas bubble generation in wastewater treatment plants) [13], and water purification membranes [18].

Several processing approaches have been developed to enhance neck growth between grains and improve strength of porous ceramics. Oh et al. [34], Jayaseelan et al. [35], and Yang et al. [36] fabricated porous Al2O3 and Al2O3-based composites by the pulse electric current sintering (PECS) technique and found that the strength was substantially improved due to the formation of thick and strong necks. During sintering, the discharge between the particles is thought to promote the bridging of particles by neck growth in the initial stages of sintering. This strong neck growth leads to substantially high strength compared to those of the conventional porous materials. For example, the flexural strength of porous alumina-based composites via PECS reached 250 and 177 MPa, with 30% and 42% porosity, respectively, which are considerably high compared to those of porous alumina fabricated by conventional partial sintering, e.g., ≈100 MPa at 30% porosity [35] (see Figure 3). Using PECS, Akhtar et al. [37] also fabricated porous ceramic monoliths from diatomite powders, which are known as a cheap and renewable, natural resource.

PECS that rapidly heats diatomite powder successfully bonds the particles together into relatively strong porous bodies, without significantly destroying the internal pores of the diatomite powder. The microstructural studies revealed that consolidation proceeds by the formation of necks at temperatures around 700—750 °C, which is followed by significant melt phase formation around 850 °C, resulting in porous ceramics with a relatively high strength.

Deng et al. [38,39] tried to obtain strong grain bonding through the combination of partial sintering and powder decomposition. A mixture of α-Al2O3 and Al(OH)3 was used as the starting powder to make porous Al2O3 ceramics, and because Al(OH)3 experiences a 60% volume contraction during decomposition and produces fine Al2O3 grains, the fracture strength of obtained porous Al2O3 was substantially higher than that of the pure Al2O3 sintered specimens because of strong grain bonding that resulted from the fine Al2O3 grains produced by the decomposition of Al(OH)3. Similar improvement of mechanical properties was also identified for ZrO2 porous ceramics fabricated by adding Zr(OH)4 [40].

Partial sintering through reaction bonding techniques have been frequently used for making porous ceramics,
where reaction products form or precipitate epitaxially on grains, resulting in well-developed neck growth between grains [41,42]. In combination with the reactive sintering process, Suzuki et al. [43,44] synthesized a CaZrO₃/MgO porous ceramic with three-dimensional grain network structure; using reactive sintering of highly pure mixtures of natural dolomite (CaMg(CO₃)₂) and synthesized zirconia powders. CaMg(CO₃)₂ decomposes into CaCO₃, MgO, and CO₂ (g) at ~500 °C, and CaCO₃ then reacts with ZrO₂ to form CaZrO₃ and CO₂ (g) at ~700 °C. Through liquid formation via LiF doping, these reactions and liberated CO₂ gas result in formation of a homogeneous open-pore structure with strong grain bonding as shown in Figure 4. The pore-size distribution is very narrow (with typical pore size: ~1 μm), and the porosity was controllable (~30–60%) by changing the sintering temperature. The relatively high flexural strength (~40 MPa for 47% porosity) was observed over the temperature range of R.T.–1300 °C. The similar approach has been applied to other materials systems such as CaAl₄O₇/CaZrO₃ and CaZrO₃/MgAl₂O₄ composite systems [45,46].

She et al. [47] used an oxidation-bonding process for the low-temperature fabrication of porous SiC ceramics with superior resistance against oxidation. In such a process, the powder compacts are heated in air instead of an inert atmosphere. The heating temperature was kept below 1300 °C in order to suppress formation of cristobalite. Figure 5 shows microstructures of two porous silicon carbides fabricated through the oxidation-bonding process (1300 °C, 1 h), using fine (0.6 μm) and coarse (2.3 μm) α-SiC powders, together with their pore-size distributions. Because of the occurrence of surface oxidation at the heating stage, SiC particles are bonded to each other by the oxidation-derived SiO₂ glass. The difference of the pore-size distributions arises from the different starting powders. Mechanical strength is strongly affected by particle size; the flexural strength attained as high as 185 MPa at a porosity of 31%, when using the fine powder (Figure 5 (a)).
while it was 88 MPa at 27% porosity for the coarse powder (Figure 5(b)). The oxidation-bonding technique has been applied to other materials including silicon nitride [48], SiC/mullite composites [49], and SiC/cordierite composites [50].

The partial sintering technique has been also applied for making porous nonoxide ceramics such as porous silicon nitride with fibrous grains of high aspect ratios [51–58]. Compared with oxide ceramics, the densification of silicon nitride ceramics is difficult because of strong covalent bonding between silicon and nitrogen atoms. This difficulty of sintering silicon nitride ceramics is beneficial for controlling density or porosity through adjusting the additives and the sintering process. In order to suppress densification, oxides with high melting point and high viscosity such as Yb₂O₃ are frequently used as sintering additives [52]. Depending on the sintering temperature, the addition of Yb₂O₃ also accelerates the fibrous grain growth of β-Si₃N₄, which substantially affects the mechanical properties of porous silicon nitrides [52,53]. Due to differences in the melting points and the preferential absorption sites of cations, the porous structure is substantially affected by the types of sintering additives as well [54].

Yang et al. [55] directly synthesized porous β-Si₃N₄ ceramics by carbothermal nitridation of silica, using carbon black as the carbon source and α-Si₃N₄ as seeds. The complete reaction results in a large weight loss and high porosities (about 65–70%) after sintering. Fine elongated fibrous β-Si₃N₄ grains were developed in the seeded samples while only large equiaxial grains were observed in the seed-free samples, as shown in Figure 6. The former sample exhibited relatively high flexure strength close to 40 MPa for the high porosity of ~65%, which is five times higher than that of the latter one. Tuyen et al. [56] fabricated porous reaction-bonded silicon nitride (RBSN) by nitridation process at 1350°C and post-sintering at 1550–1850°C, which provides similar fibrous microstructure and high porosity. The sintering time had a significant effect on the microstructure and grain morphology, and porous structure with fibrous grains of high aspect ratios was obtained by adjusting the time even at the comparatively low temperature (1550°C). These techniques offer the possibility of synthesizing highly porous and strong Si₃N₄ materials at considerably lower cost.

One of the unique processing routes for porous ceramics with anisotropic microstructure is tape-casting fibrous seed crystals or whiskers. For porous silicon nitrides, β-Si₃N₄ seed crystals were mixed with sintering additives as starting powders, and the green sheets formed by tape casting were stacked and bonded under pressure, followed by sintering at 1850°C under a nitrogen pressure of 1 MPa [57,58]. The microstructural observation for the porous silicon nitrides revealed that the fibrous grains are aligned toward the casting direction, and the plate-like pores exist among the grains. Because of the enhanced crack shielding effects of aligned fibrous grains, the anisotropic porous silicon nitrides showed excellent mechanical behavior, when a stress is applied in the alignment direction. More detailed discussion on the microstructure changes and mechanical properties (strength, fracture toughness, and thermal shock resistances) for the isotropic and anisotropic porous silicon nitrides will be given in the Chapter “Microstructural Control and Mechanical Properties.”

3. SACRIFICIAL FUGITIVES

Porous ceramics are often fabricated by mixing appropriate amounts of sacrificial fugitives as pore-forming agents with ceramic raw powder and evaporating or burning out them before or during sintering to create pores (see Figure 2(b)). Frequently used pore-forming agents are polymer beads, organic fibers, potato starch, graphite, charcoal, salicylic
acid, carbonyl, coal, and liquid paraffin. The pore-forming agents are generally classified into synthetic organic matters (polymer beads, organic fibers, etc.) [59—89], natural organic matters (potato starch, cellulose, cotton, etc.) [67,68,90—105], metallic and inorganic matters (nickel, carbon, fly ash, glass particles, etc.) [49,78,106—114], and liquid (water, gel, emulsions, etc.) [115—155]. Porosity is controllable by the amount of the agents and pore shape and size are also affected by the shape and size of the agents, respectively, when their sizes are large in comparison with those of starting powders or matrix grains. The agents, however, need to be mixed with ceramic raw powder homogeneously for obtaining uniform and regular distribution of pores. Solid fugitives like organic materials are usually removed through pyrolysis, which often requires long-term heat-treatment and generates a great deal of vaporized, sometimes harmful, by-products.

Polymethylmethacrylate (PMMA) beads and microbeads have been frequently employed for sacrificial fugitives [8,59—64,77,79,82—85]. Colombo and his co-workers [8,59—61] fabricated SiOC ceramic foam by dry mixing the silicon resin powder with a sacrificial template constituted by PMMA microbeads, and subsequent heat treatments (Figure 7). Descamps et al. [63,64] produced macroporous β-tricalcium phosphate (TCP) ceramics using PMMA. An organic skeleton was formed by interconnecting the PMMA balls through a chemical superficial dissolution, and was impregnated by the TCP slurry. PMMA was then eliminated by a thermal treatment at low temperature, followed by sintering for final porous structure. This process allows a total control of the porous architecture; the porous volume can vary from 70 to 80% and the interconnection size from 0.2 to 0.6 times the average macro-pore size.

Andersson et al. [82] used expandable microspheres as a sacrificial template to produce macroporous ceramic materials by a gel-casting process. The microspheres consist of a co-polymer shell and are filled with a blowing agent (isobutane), which allows rapid and facile burn out. By controlling the amount and size of the expandable microspheres, it is possible to tune the porosity up to 86% and the pore-size distribution from 15 to 150 μm. Up to 1—2 wt.% of the microspheres leads to a final porosity above 80 vol%. Expandable microspheres as sacrificial templates, rather than other templates such as PMMA microbeads, are advantageous because of lower levels of gaseous by-products generated during pyrolysis, and lower cost of the overall materials. Kim and his co-workers [83,84] used hollow microspheres as sacrificial templates to make porous silicon carbide ceramics synthesized from carbon-filled polysiloxane and others. Using preceramic polymer and organic microspheres for fabricating porous ceramics allows use of the low-cost and/or near-net-shaped processing techniques like extrusion and direct casting [84]. Song et al. [85] produced microcellular silicon carbide ceramics with a duplex pore structure using expandable microspheres and PMMA spheres, which resulted in the large pores and the small windows in the strut area, respectively. This porous ceramics showed excellent air permeability as shown in Section 6.

Diaz et al. [93,94] fabricated porous silicon nitride ceramics using a fugitive additive, cornstarch (particle size: 5—18 μm). In order to obtain homogeneous dispersion of the fugitives, the mixture slurry was kept in agitation using a magnetic stirrer for a while, and then was frozen and dried under vacuum for sieving. Kim et al. [95] mixed various amounts of cornstarch to (Ba, Sr) TiO3 powder to obtain (Ba, Sr) TiO3 porous ceramics. They found that depending on the porosity, the PTCR effect was 1—2 orders of magnitude improved in comparison with the dense reference.

Chen et al. [66] produced porous silicon nitride of equiaxed α-grains by using phosphoric acid (H3PO4) as the pore-forming agent and relatively low temperature (1000—1200°C) sintering. On the other hand, Li et al. [80] fabricated porous silicon nitride with fibrous β-grain structure, using naphthalene powder as the pore-forming agent and gas-pressure sintering of high temperatures above 1700 °C. The bending strength of the former materials was 50—120 MPa in porosity range of 42—63%, while that of the latter was 160—220 MPa in porosity range of 50—54%. This substantial difference in strength is attributable to the microstructural difference (equiaxed vs. fibrous), similar to the case of Figure 5.

Ding et al. [49] used graphite as the pore-former to fabricate mullite-bonded porous silicon carbide ceramics in air from SiC and α-Al2O3 through an in-situ reaction.
bonding technique. Graphite is burned out to form pores and SiC is oxidized at high temperatures to SiO$_2$, which further reacts with $\gamma$-Al$_2$O$_3$ to form mullite (3Al$_2$O$_3$·2SiO$_2$). SiC particles are bonded by mullite and oxidation-derived SiO$_2$.

Long fibers such as cotton thread [96], natural tropical fiber [97], and metal wires [109] are often used as pore-forming agents for obtaining porous ceramics of unidirectional through channels. Zhang et al. [96] fabricated porous alumina ceramics with unidirectionally aligned continuous pores (diameter: $\sim$160 $\mu$m) via the slurry coating of mercerized cotton threads. The pore size can be adjusted, using cotton threads of different diameter, and the porosity can be controlled by changing the solids’ concentration of the slurry. In this case, excellent permeability can be achieved for porous ceramics with unidirectional through channel pores, because gas can flow directly through the pores. However, the preparation of such ceramics is complex because handling long fibers such as thin wire or cotton thread is difficult. Using short fibers or whiskers as the pore-forming agent is an alternative that combines the advantages of partially sintered porous ceramic and those of unidirectional pores. Yang et al. [65] demonstrated formation of rod-shaped pores in silicon nitride ceramics, using slip casting of aqueous slurries of silicon nitride powder and sintering additives with 0–60 vol% fugitive organic whiskers. Rheological properties of slurries were optimized to achieve a high degree of dispersion with a high solid-volume fraction. Samples were heated at 800 $^\circ$C in air to remove the whiskers and sintered at 1850 $^\circ$C in nitrogen atmosphere to consolidate the matrix. Porosity was adjusted in 0–45% by changing the whisker content in 0–60 vol%. The obtained porous silicon nitride contained uniform rod-shaped pores with random directions as shown in Figure 8, and therefore exhibited relatively high gas permeability in comparison to porous silicon nitride containing equiaxed pores [156]. Isobe et al. [81,110] and Okada et al. [86,87] used carbon fibers (14 $\mu$m diameter and 600 $\mu$m length) or Nylon 66 fibers (9.5–43 $\mu$m diameter and 800 $\mu$m length) as a pore-forming agent, and tried to align them by an extrusion technique to produce porous alumina [81,110] and mullite [86,87] ceramics with unidirectionally oriented pores. The pore sizes and porosities can be controlled by varying the fiber diameter and fiber content, and the obtained samples showed better air permeability than the conventional porous materials used for filter applications [110]. This technique can allow the production of highly oriented porous ceramics by an industrially favored extrusion method.

Liquids such as water and oils, which are readily sublimated or evaporated, are often used as pore-forming agents [115–156]. One of the most frequently employed approaches in recent years is freeze-drying the water- or liquid-based slurry to produce porous ceramics of unique structure [118–154]. Figure 9 shows a schematic
Illustration of the procedures which Fukasawa et al. [118–120] employed, and a porous silicon nitride body obtained thereby. When the bottom part of the slurry is frozen, ice grows macroscopically in the vertical direction, and pores are generated subsequently by sublimation of the ice. Sintering this green body results in a porous ceramic and pores are generated subsequently by sublimation of the frozen, ice grows macroscopically in the vertical direction, which contains smaller pores in the internal walls (Al2O3) [118,119] or fibrous grains protruding from them (Si3N4) [120]. The advantages of this approach include a simple process without materials to be burnt out, a wide range of porosity (30–99%) controllable by the slurry concentration, applicability to various types of ceramics, and environmental friendliness without emitting harmful products. Particularly porous scaffolds with ice-designed channel-like porosity fabricated by this method have been intensively studied for a wide variety of applications including biomedical implants and catalysis supports.

The porosity of the porous materials obtained using this technique is a replica of the original ice structure. The porous channels run from the bottom to the top of the samples (when the bottom part is first frozen), and the pores most frequently exhibit an anisotropic morphology in the solidification plane. Deville et al. [127–130] investigated the relationships between the freezing conditions and the final porous structures in freeze casting of ceramic slurries. It has been clarified that the morphology of the porous structures including the dimensions, shape, and orientation of porosity are adjustable by varying the initial slurry compositions and the freezing conditions. For highly concentrated solutions, the particle–particle interactions lead to the formation of ceramic bridges between two adjacent lamellae. Using this technique, they fabricated sophisticated porous and layered-hybrid materials such as nacre-like structure with lamellar dendrites and high compressive strength (four times higher than those of materials currently used for implantation). Munch et al. [131] emulated nature’s toughening mechanisms by combining two ordinary compounds, alumina and poly-methyl methacrylate, into an ice-templated structure, and succeeded in obtaining toughness more than 300 times (in energy terms) that of their constituents.

Araki and Halloran [132–134] used camphene, C10H16, as a vehicle for producing porous ceramics via a freeze-drying process, to realize a freezing process at room temperature. Slurries containing ceramic powder in the molten camphene were prepared at 55 °C, and were quickly solidified (frozen) when they were poured into poly-urethane molds at room temperature. The obtained porous ceramics have pore channels of nearly circular cross sections (unlike ellipsoidal ones obtained via conventional aqueous freeze casting). Koh and his co-workers used a similar camphene-based freeze-casting approach to fabricate highly porous Al2O3 [135–137], SiC [138,139], PZT-based ceramics [140,141], hydroxyapatite [142,143], glass-ceramics [144], and ZrO2 [145,146], etc., having interconnected pore without noticeable defects. Many researchers have tried to combine the freeze-dry process and the gel-casting technique to produce porous ceramics with refined microstructure [147–155]. It has been shown that the use of an organic polymer in the freeze-casting route affects the pore size and morphology by controlling ice crystal growth during freezing. Chen et al. [147,148] used alumina slurries containing tert-butyl alcohol (TBA) and acrylamide (AM) for the freeze-dry process. TBA freezes below 25°C and volatilizes rapidly above 30°C, acting as the freezing vehicle and template for forming pores, while AM is polymerized in the slurry as the gelation agent, strengthening the green bodies substantially. The sintered porous ceramics have high compression strength (~150 MPa at 60% porosity) because the pore channels formed by the TBA template are surrounded by almost fully dense walls without any noticeable defects. Ding et al. [149] also employed a gel freeze-drying process to fabricate porous mullite ceramics with porosity up to 93%. Alumina gel mixed with ultrafine silica was frozen isotropically, followed by sublimation of ice crystals. Porous mullite ceramics were prepared in air at 1400–1600 °C due to the nullification between Al2O3 and SiO2. Porous yttria-stabilized zirconia (YSZ) [150] and porous alumina [151] were fabricated by the freeze-drying process with addition of polyvinyl alcohol (PVA), which suppresses ice crystal growth and reduces the pore sizes substantially. Porous alumina with oriented pore structures has been also fabricated by the freeze-casting technique with a water-soluble polymer such as polyethylene glycol (PEG) [152]. Using precursor silica hydrogels, Nishihara et al. [153] fabricated ordered macroporous silica (silica gel micro-honeycomb) using freeze-dry methods, where micrometer-sized ice crystals are used as a template. The pore sizes can be controlled by changing the immersion rate into a cold bath and the freezing temperature. The average pore size can be as small as 4.7 μm with the rate of 20 cm/h at 77 K, and the thickness of the honeycomb walls can be adjusted by the SiO2 concentration.

Fukushima et al. [154,155] fabricated porous cordierite or silicon carbide ceramics with porosity from 80 to 95% using a gel-freezing method; unidirectionally oriented cylindrical channels are uniformly distributed over relatively large bulk samples (typically several centimeters). Gelatin was used as the gelation agent, which was mixed with water for the freezing vehicle and raw powder. The gel was frozen at −10 to −70°C, dried under vacuum, and degreased before sintering was carried out (at 1200–1400 °C for cordierite and at 1800 °C for silicon carbide). The cell size and cell wall thickness both decreased with decreasing the freezing temperature, from 200 to 20 μm and from 20 to 3 μm, respectively. The
4. REPLICA TEMPLATES

Macroporous ceramics having interconnected large pores, or channels, of high porosity have been frequently fabricated by the replica techniques (Figure 2(c)). The first step is typically impregnation of a porous or cellular template with ceramic suspension, precursor solution, etc. Various synthetic and natural cellular structures are used as the templates. The most frequently used synthetic template is porous polymeric sponge such as polyurethane. They are soaked into a ceramic slurry or precursor solution to impregnate the templates with them, and the surplus is drained and removed by centrifugation, roller compression, etc. In this process, appropriate viscosity and fluidity depending on the cell size, etc., are required in order to obtain uniform ceramic layer over the sponge walls. The ceramic-impregnated templates are dried and then heat-treated to decompose the organic sponges, followed by sintering the ceramic layers at higher temperatures. Porosity higher than 90% can be obtained with cell sizes ranging from a few hundred micron meters to several millimeters. The open cells are interconnected, which allows fluid to pass through the foams with a relatively low pressure drop. Figure 11 shows a typical example of alumina foam prepared by slurry infiltration of polyurethane templates, which has been reported by Faure et al. [157]. However, due to cracking the struts during the pyrolysis, the mechanical properties of ceramic reticulated foams are generally poor. In order to avoid strut crack formation, various approaches have been made [157–163]. For examples, Vogt et al. [159] employed vacuum infiltration of ceramic slurry to fill up the struts in the presintered foam. The hollow struts caused by burnout of the polyurethane template were completely filled up, resulting in a considerable increase of compressive strength. Jun et al. [161,162] produced hydroxyapatite scaffolds coated with bioactive glass-ceramics using the polymer foam replication method, to enhance their mechanical properties and bioactivities. Highly porous ceramics can be also obtained from preceramic polymers after pyrolysis above 800 °C in inert atmosphere [8]. One of the typical methods is dissolving the silicone resin preceramic polymer into a suitable solvent and adding appropriate surfactants and catalysts, followed by pyrolysis. The advantages are a wide control of pore sizes (typically 1 μm to 2 mm), well-defined open-cell structures, and macro-defect-free struts [164–167]. Travitzky et al. [168] also succeeded in fabricating single-sheet, corrugated structures, and multilayer ceramics by using various paper replica templates.

Natural resources of porous structures such as woods, corals, sea sponge, etc., can be also used as replica templates. The woods are transformed to carbonaceous preforms by heat-treatment in inert atmosphere. They are then infiltrated with molten metals [169–179], gaseous metals [174,180–185], alkoxide solutions [186–188], and others [189,190]. The advantages include a wide variety of obtained porous structures (depending on the type of wood selected), low-cost starting materials, near-net and complex shape capabilities, and a relatively low-temperature fabrication process. Locs et al. prepared porous SiC ceramics from pyrolyzed pine wood samples via impregnation with SiO2 sol and heat-treatment at 1600 °C for 4 hours, as shown in Figure 12 [190]. The longitudinal pore size in SiC is 10–20 μm and the wall thickness is 3–5 μm. The oriented vessels of the woods provide unique anisotropic porous structure of aligned unidirectional through channels, which is suitable to applications such as filtration and catalysis supports. Porous biomimetic silicon carbides obtained through this approach have been also studied for medical implant materials [191].

Biomorphic porous silicon nitride was produced from natural sea sponge via replication method. The sponges were impregnated with silicon-containing slurry via dip coating, and were heat-treated to delete the bio-polymers, leading to a Si-skeleton. Subsequent thermal treatment under flowing nitrogen promoted the nitridation of the silicon, porous α/β-silicon nitride with a porosity of 88%, and the original morphology of the sea sponge [192].
5. DIRECT FOAMING

In direct foaming techniques, the ceramics suspension is foamed by incorporating air or gas and stabilized and dried, followed by sintering to obtain a consolidated structure (Figure 2(d)). The advantage of this technique is low-cost and easy fabrication of highly porous ceramic materials (95% or higher porosity). Porous ceramics with unidirectional channels can be produced using continuous bubble formation in ceramic slurry [193,194].

However, due to the thermodynamic instability, the gas bubbles easily coalesce in order to reduce the total Gibbs free energy of the system, which results in undesirable large pores. It is, therefore, critically needed to stabilize the air or gas bubbles in ceramic suspension. One of the most frequently approaches for the stabilization is to use surfactants reducing the interfacial energy of the gas–liquid boundaries. Surfactants used for stabilization are classified into several types including nonionic, anionic, cationic, and protein, and the pore size of the produced porous body ranges from below 50 μm up to the millimeter scale, depending on the used surfactants. A variety of effective surfactants have been developed for direct foaming of porous ceramics [195–201].

Barg et al. [198–200] developed a novel direct foaming process by emulsifying a homogeneously dispersed alkane or air–alkane phase in the stabilized aqueous powder suspension. Foaming is made by evaporation of the emulsified alkane droplet, leading to high performance ceramic foams with porosities up to 90% and cell sizes ranging from 3 to 200 μm. This autonomous foaming process also allows high flexibility in the production of ceramic parts with gradient structures and complex shaping. Foaming proceeds as a consequence of the evaporation of the alkane phase resulting in the growth of the stabilized alkane bubbles and in a volume increase of the foam. Figure 13 shows an example of the alumina foam which was produced through sintering (1550°C/2 h) from high alkane phase emulsified suspensions containing 45 vol% particle content [200].
Preceramic polymer solution has been also used instead of ceramic suspension for direct foaming. Colombo et al. [202] produced porous ceramics by dissolving preceramic polymers (silicone resins) into a suitable solvent with blowing agent, surfactant, catalyst, etc., and heat-treating them at 1000–1200 °C in an inert atmosphere. Due to the suppressed defect formations in the struts, the strength of the obtained porous ceramics was relatively high in comparison to those of conventional reticulated foams [203]. Kim et al. [204,205] fabricated porous ceramics with a fine microcellular structure from preceramic polymers using CO₂ as a blowing agent. A mixture of polycarbosilane and polysiloxane was saturated with gaseous CO₂ under a high pressure and bubbles were introduced using a thermodynamic instability via a rapid pressure drop, followed by pyrolysis and sintering.

It has been shown that particles with tailored surface chemistry can also be used efficiently to stabilize gas bubbles for producing stable wet foams [206–214]. Gonzenbach et al. [210–214] have developed a novel direct foaming method that uses colloidal particles as foam stabilizers. The method is based on the in-situ hydrophobization of initially hydrophilic particles to enable their adsorption on the surface of air bubbles. In-situ hydrophobization is accomplished through the adsorption of short-chain amphiphiles on the particle surface. The obtained ultra-stable wet foams show neither bubble coalescence nor disproportionation over several days, as opposed to the several minutes typically required for the collapse of the surfactant-based foams. Because of their remarkable stability, the particle-stabilized foams can be dried directly in air without crack formation. The macroporous ceramics obtained after sintering have porosities from 45% to 95% and cell sizes between 10 and 300 μm. The compressive strength of the sintered foams with closed cells is relatively high in comparison with those of foams prepared with other conventional techniques (for example, 16 MPa at a porosity of 88% in alumina foams). It has been observed that the surface-modified particles which originally cover the air bubble in wet foams become a thin surface layer of single grains after sintering. Macroporous ceramics with open porosity can be also fabricated using this technique when decreasing the concentration of stabilizing particles.

6. GAS PERMEABILITY

Gas permeability is one of the most important properties of porous ceramics which are expected to be used for gas filters such as diesel particulate filter (DPF), since large pressure drops should be avoided in such applications.
Highly porous ceramics with aligned unidirectionally through pore channels are expected to provide excellent permeability, and as already stated, the freeze-dry technique is one of the most representative processes for producing such porous ceramics. This section deals with Darcian air permeability of porous ceramics with different pore sizes and structures. Figure 14 shows the Darcian permeability as a function of pore size for porous ceramics fabricated by freeze-dry processes [151,155], organic spherical fugitives [61,85], graphite fugitives [215,216], extruded organic fibrous fugitives [81,87], direct foaming [195,217], and replica templates [217,218]. The pore structures are classified into three categories of “Spherical (Connected)” [61,85,195,215]e [218] , “Cylindrical (Connected)” [81,87] and “Cylindrical” [151,155], as schematically shown in the figure. The Darcian permeability, \( K \), is determined from pressure drop and flow rate of air by the Darcy’s law [219].

Based on the capillary model, \( K \) is expressed by

\[
K = \frac{\phi D_p^2}{C}
\]  

where \( \phi \) is the porosity, \( D_p \) is the pore diameter, and \( C \) is a constant depending on the pore structure [219,220].

\( K \) values of Figure 14 are those adjusted from the reported values at the porosity of 0.85 using Eqn (1) for comparison. The inertial contribution (non-Darcian permeability) was considered in addition to the viscous one (Darcian permeability) in Ref. [61,152,217–219], which results in high values of \( K \) in comparison to the cases of neglected inertial effect [81,85,175,195,215] (the ratio of viscous contribution in total is 60–90% [151]). The solid line of the figure shows the case that the fluid flows through the unidirectional cylindrical pores penetrating in parallel (\( C = 32 \)) [220]. The porous ceramics fabricated by the freeze-dry processes [151,155] showed permeability very close to this solid line, indicating the unidirectional alignment of the cylindrical pores. The permeability required for a commercially available DPF is \( 10^{-11} \) to \( 10^{-12} \) m² [221], and most of the freeze-dry-processed materials exceed this criterion. The porous ceramics prepared with extruded organic fibrous fugitives [81,87] showed lower permeability values than those of the freeze-dry-processed ones, most likely because of limited contact area among the short fibers.

7. SUMMARY

During the last decade, tremendous efforts have been devoted to research on porous ceramics, resulting in better control of the porous structures and substantial improvements of the properties. This chapter reviewed these recent progresses of porous ceramics. Because of the vast amount of research works reported in this field these days, the chapter mainly focused on macroporous ceramics whose pore size is larger than 50 nm. Followed by giving a general classification of porous ceramics, a number of innovative processing routes for critically controlling pores were described, along with some important properties. They were divided into four categories including (1) partial sintering, (2) sacrificial fugitives, (3) replica templates, and (4) direct foaming.

The partial sintering, which is one of the most common techniques for making porous ceramics, has been substantially sophisticated in recent years. Very homogeneous porous ceramics with extremely narrow size distribution has been successfully prepared through sintering combined with in-situ chemical synthesis. Porous silicon nitrides with aligned fibrous grains and pores have demonstrated excellent mechanical properties, which are equivalent, or sometimes superior, to those of the dense materials. The sacrificial fugitives have an advantage that pore shape and size are controllable by the shape and size of the agents, respectively. The fugitives are generally removed through pyrolysis, generating a great deal of vaporized, sometimes harmful, by-products, and a lot of research has been conducted to reduce or eliminate them. The freeze-dry processes using water or liquid as fugitive agents are advantageous in this viewpoint and have been very intensively studied in recent years. Careful control of
ice growth leads to unique porous structures and excellent performances of porous ceramics. The replica template techniques have been widely used to fabricate highly porous ceramics with interconnected large pores. Porous polymeric sponge such as polyurethane is the most typical synthetic template used for this process. However, due to cracking struts during pyrolysis of the sponge, the mechanical property is generally low; a variety of approaches have been used to avoid strut crack formation. Natural template approaches using wood, for example, as positive replica, have been frequently studied in these years and have realized highly oriented porous open-porous structure with a wide range of porosity. The direct foaming technique is low-cost and an easy fabrication process of porous ceramics with high porosity volume. In order to suppress coalescence of gas bubbles in ceramic suspension that results in large pores in the final porous bodies, various methods which stabilize the bubbles have been developed; they include use of effective surfactants, evaporation of emulsified alkane droplets, and use of surface-modified particles. Finally, we discussed gas permeability (Darcian air permeability) of porous ceramics with different pore sizes and structures. It has been demonstrated that the freeze-dry-processed porous ceramics with cylindrical through channels have excellent permeability, which is close to the ideal case that the fluid flows through the unidirectional cylindrical pores penetrating in parallel.

REFERENCES

[31] (For example) Fukushima M, Zhou Y, Miyazaki H, Yoshizawa Y, Hiro K, Iwamoto Y, et al. Microstructural characterization of


Spark Plasma Sintering (SPS) Method, Systems, and Applications

Masao Tokita
NJS Co., Ltd. 301 Office Shinyokohama, 2-14-8, Shinyokohama, Kohoku, Yokohama, Kanagawa 222-0033, Japan

Chapter Outline
1. Introduction 1149
2. Historical Background 1149
3. Suitable Materials For SPS Process 1152
4. Principles of the SPS Process 1152
5. Examples of SPS Process Applications 1160
6. Summary and Outlook 1174

Chapter 11.2.3

1. INTRODUCTION

Spark plasma sintering (SPS), also called the pressure-assisted pulse energizing process or the pulsed electric current sintering (PECS) process, is a promising technology for innovative processing in the field of new materials fabrication in the 21st century.

SPS is a synthesis and processing technique which makes possible sintering and sinter bonding at low temperatures and in short periods by charging the intervals between powder particles with electrical energy and effectively applying a high-temperature spark plasma generated at an initial stage of energizing momentarily, and an electro-magnetic field and/or joule heating by continuous ON-OFF DC pulsed high electric current with a low voltage. As shown in Figure 1, the method is a solid compressive and a large pulse electric current energizing sintering technique that has recently drawn considerable attention as one of the newest rapid sintering methods with accurate energy density control. It is a novel sintering process featuring energy saving and high-speed consolidation and has a low power consumption of between 1/5 and 1/3 compared with conventional sintering techniques such as pressureless sintering (PLS), hot press (HP) sintering, and hot isostatic pressing (HIP) [1-6]. The system outlook is similar to a conventional hot press apparatus without outer heating element. However, the SPS has demonstrated the different superior sintering results for instance a structurally tailoring effect, minimizing grain growth, enhancement of electro-migration and strong preferential orientation effect in SPS processing [4]. This paper introduces the recent SPS technology, method, development of SPS systems, and its applications.

2. HISTORICAL BACKGROUND

Since two decades ago, spark plasma sintering (SPS) method is of great interest to the powder and powder metallurgy industry and to material researchers of academia for both product manufacturing and advanced material research and development. It is generally well known that the SPS is an advanced processing technology to produce a homogeneous highly dense nanostructural sintered compacts, functionally graded materials (FGMs), fine ceramics, composite materials, new wear-resistant materials, thermo-electric semiconductors, and biomaterials. Today in Japan, a number of SPSed products for different industries have already been realized. The SPS is now moving from the scientific academia and/or R&D prototype materials level to practical industry-use product stage in the field of mold and die industry, cutting tools industry, electronics industry, and automotive industry. A technique similar to SPS was first studied in Germany around 1910 that was an electric energizing applied technique to consolidate a powder material. In the USA, G.F. Tayler patented the first resistance sintering method for sheet metals in 1933 [7]. Thereafter, G.D. Cremer obtained US patent for the method of sintering for copper, brass, or aluminum powder materials [8]. They were...
considered as the origin of a current hot pressing (HP) technique that commonly applies a high-frequency induction heating method. The SPS was originally invented in Japan as “spark sintering (SS)” in 1962 by Dr. Kiyoshi Inoue of Japax Inc. [9,10]. In 1989, the present SPS was introduced by Sumitomo Coal Mining Co., Ltd (Japan) [11], and was developed as the third-generation sintering technique to advance the first generation of spark sintering and the second generation of plasma-activated sintering (PAS) from Inoue-Japax Research Inc. As shown in Figure 2, the historical progress of SPS technology is indicated by the relationship between size effect and shape effect containing functionality, reproducibility, and productivity. Research and development for the implementation of advanced SPS methods and systems were initiated to design practical hardware and software for industrial applications. Following the development of a box-type experimental-use SPS systems for new materials preparation and a single-head open-type introductory limited-production sintering systems, from 2001 to 2009 as the fourth-generation technology to accommodate for the product manufacturing field, five
basic styles of SPS production systems from the medium to mass-production scale were developed [12,13]. As a result, the fabrication of various advanced new materials and industrial products has been presented. The 4th-generation system performed to replace existing traditional fabrication processes to powder used processes, and to divide the SPS technology into four kinds of SPS processing and methods, namely the confirmed availability, were not only a sintering, but also a solid-phase diffusion bonding and joining [14], a surface modification (treatment) [15–17], and a synthesis technique for an example of a single-crystal fabrication as the SPS field. After 2010, it is to be called the start of a more practical manufacturing era, the progress of SPS technology is now getting into the 5th generation of “advanced SPS” with customized SPS apparatus. Figure 3 shows an example of an SPS job-shop center facility in Japan, a large ceramic sample, and total number of SPS machine systems produced and installed from 1990 to as of early 2010. Due to the increase of SPS manufacturers internationally in recent years, more than 550 units of SPS machines have already been working in the world approximately.
3. SUITABLE MATERIALS FOR SPS PROCESS

Table 1 represents an example of suitable materials for SPS processing. Figure 4 is a typical example of SPS sintering effect on nano-SiC ceramic material in grain growth and Al₂O₃ in terms of Hv hardness. Table 2 shows a comparison of characteristics on SPS method and conventional hot press (HP) sintering. The SPS process features very high thermal efficiency because of the direct heating of the sintering graphite mold and compressed powder materials by a large DC pulsed current. It can easily consolidate a homogeneous, high-dense high-quality sintered compact because of the uniform heating, surface purification, and activation made possible by dispersing the spark points (early stage) and/or joule heat points during sintering.

4. PRINCIPLES OF THE SPS PROCESS

4.1. Basic Configuration of the SPS System

Figure 5 shows the basic configuration of a typical SPS system. The system consists of an SPS sintering press machine with a vertical single-axis pressurization mechanism, specially designed punch electrodes incorporating a water cooler, a water-cooled vacuum chamber, a vacuum/air/argon-gas atmosphere control mechanism, a special DC pulse sintering power generator, a cooling-water control unit, Z-axis position measuring and control unit, temperature measuring and control units, an applied pressure

TABLE 1 Suitable Materials for SPS Process

<table>
<thead>
<tr>
<th>Classification</th>
<th>Materials for SPS processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metals</td>
<td>Fe, Cu, Al, Au, Ag, Ni, Cr, Mo, Sn, Ti, W, Be virtually any metal possible</td>
</tr>
<tr>
<td>Ceramics Oxides</td>
<td>Al₂O₃, mullite, ZrO₂, MgO, SiO₂, TiO₂, HfO₂</td>
</tr>
<tr>
<td>Carbides</td>
<td>SiC, B₄C, TaC, TiC, WC, ZrC, VC</td>
</tr>
<tr>
<td>Nitrides</td>
<td>Si₃N₄, TaN, TiN, AlN, ZrN, VN</td>
</tr>
<tr>
<td>Borides</td>
<td>TiB₂, HfB₂, LaB₆, ZrB₂, VB₂, MgB₂</td>
</tr>
<tr>
<td>Fluorides</td>
<td>LiF, CaF₂, MgF₂</td>
</tr>
<tr>
<td>Cermets</td>
<td>Si₃N₄ + Ni, Al₂O₃ + Ni, ZrO₂ + Ni</td>
</tr>
<tr>
<td>Al₂O₃ + Ti, ZrO₂ + SUS, Al₂O₃ + SUS</td>
<td></td>
</tr>
<tr>
<td>WC + Co, WC + Ni, TiC + TiN + Ni, BN + Fe,</td>
<td></td>
</tr>
<tr>
<td>Intermetallic compounds</td>
<td>TiAl, MoSi₂, Si₃N₄, NiAl</td>
</tr>
<tr>
<td>NbCo, Nb₃Al, LaBaCu₄S₉, Sm₂Co₁₇</td>
<td></td>
</tr>
<tr>
<td>Other materials</td>
<td>Organic materials (polyimide, etc.), FRM, FRC, CNT composite materials</td>
</tr>
</tbody>
</table>

FIGURE 4 Nano-SiC compact by SPS and Al₂O₃ Hv hardness distribution behavior sintered by various sintering methods.

TABLE 2 Comparison of Characteristics on SPS Vs HP

<table>
<thead>
<tr>
<th></th>
<th>SPS sintering</th>
<th>HP sintering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature gradient sintering</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>Grain boundary controlled sintering</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>Fine crystalline structure controlled sintering</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>Temperature rise rate</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>Sintering time</td>
<td>Temperature rise time</td>
<td>Fast</td>
</tr>
<tr>
<td>Holding time</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>Homogeneous sintering</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Expandability</td>
<td>○</td>
<td>Δ</td>
</tr>
<tr>
<td>Productivity</td>
<td>○</td>
<td>Δ</td>
</tr>
<tr>
<td>Investment in equipment</td>
<td>○</td>
<td>Δ</td>
</tr>
<tr>
<td>Running cost</td>
<td>○</td>
<td>Δ</td>
</tr>
</tbody>
</table>

(○ excellent, ○ good, △ fair, × difficult)
display unit, and various safety interlock devices. Figures 6 and 7 show an example of a medium-sized box-structure experimental-use SPS apparatus and the inside of the water-cooling chamber during SPS sintering.

4.2. Mechanism of SPS Process

Despite many years of research work concerning the SPS mechanism by many material researchers, the SPS effect, in other words, the effect of pulsed high current on the generation of spark plasma, peculiar properties in consolidated materials, still remains unclear [1, 18–24]. The SPS process is a dynamical nonequilibrium processing phenomenon that varies from early stage to later stage of sintering along with reacted material characteristics. However, following is one of the most basic proposed ideas on the mechanism of SPS processing. The ON–OFF DC pulse energizing method generates (1)
spark plasma, (2) spark impact pressure, (3) joule heating, and (4) an electrical field diffusion effect. In the SPS process, the powder particle surfaces are more easily purified and activated than in conventional electrical sintering processes, and material transfers at both the micro- and macro-levels are promoted; so a high-quality sintered compact is obtained at a lower temperature and in a shorter time compared with conventional processes. Figure 8 illustrates a typical ON–OFF DC current path and how pulse current flows through powder particles inside the SPS sintering using conductive powder material, die, and punches made of graphite material. Conventional electrical hot press processes use DC or commercial AC power, and the main factors promoting sintering in these processes are the joule heat generated by the power supply ($I^2 R$) or high-frequency induction heating elements and the plastic flow of materials caused by the hydraulically or mechanically driving pressure. The SPS process is an electrical sintering technique which applies an ON–OFF DC pulse voltage and high current from a special pulse generator to a powder of particles, and, in addition to the factors prompting sintering described above, it effectively discharges between particles of powder occurring at the initial stage of the pulse energizing for sintering. When a sparking occurs, high-temperature field with sputtering phenomenon generated by spark plasma and spark impact pressure eliminates adsorptive gases and oxide films and impurities existing on the surface of the powder particles. The action of the electro-magnetic field enhances high-speed diffusion due to the high-speed migration of ions. The application of the pulse voltage induces various phenomena as shown in Figure 9 in terms of effects of ON–OFF DC pulse energizing method. In case of WITHOUT sparking due to non-conductive material used or the pulse energizing conditions, however, it is considered that there still exists the effect of ON–OFF DC pulse current energizing which provides enhancement of sinterability and densification rate on the material. The large pulsed energy generates an electro-magnetic field effect such as an electro-migration and preferential orientation of crystalline [3,4,25].

4.3. Neck Formation and Densification by SPS

When a spark discharge appears in a gap or at the contact point between the particles of a material at early stage of
sintering, a local high-temperature state (discharge column) of several to ten thousands of degrees centigrade is generated momentarily. This causes evaporation and melting on the surface of powder particles in the SPS process, and “necks” are formed around the area of contact point between particles. Figure 10 shows a basic mechanism of neck formation by spark plasma sintering. Figure 11 is an SEM micrograph showing the result of SPS experiments performed at normal atmospheric pressure (no-load) and 29 MPa applied in vacuum, with a sintering die and punches made of a graphite and a spherical bronze alloy powder (Cu90/Sn10 wt%, particle size 45 μm under). Figure 11 (left) shows the behavior in the initial stage of neck formation due to sparks in the plasma. The heat is transferred immediately from the center of the spark discharge column to the sphere surface and diffused so that intergranular bonding portion is quickly cooled. As seen in Figure 11 (center) photo, which shows several necks, the pulse energizing method causes spark discharge one after another between particles. Even with a single particle, the number of portions where necks are formed between adjacent particles increases as the discharges and/or joule heating are repeated. The right of Figure 11 shows the condition of an SPS sintered grain boundary which is plastic-deformed after the sintering has progressed further. This state is the result of processing conditions in which the applied pressure was 29 MPa, the SPS sintering temperature was 773 K (measured in the wall of graphite mold), the holding time was 120 sec., the SPS current 850 A, and the voltage was 3.9 V. Figure 12(a) (b) (c) and 13 indicate ultra-high-temperature field existence between particles as collateral evidence examples of SPS effects on metallic and ceramic materials locally. It is suggested that the rippled surfaces observed in Figure 12 (a) were caused by the high-temperature state resulting from thermal and sputtering effects of spark plasma and impact pressure by SPS on atomized cast iron at 973 K. Figure 12 (b) shows sintering by pressureless normal sintering at the same temperature affecting only thermal effect. However, as shown in Figure 12(c) and Figure 13, the local high-temperature state can be observed as a bridging, evaporation, solidification and/or recrystallization phenomena at SPS sintering temperature of Fe/973K, Ni/673K, Al2O3/1173K, SiC/1973K respectively. These temperatures are 1/2-1/3 of much lower temperature than each material’s melting point temperature. Thus, an existence of high-temperature field is suggested [26–29] locally. It is noticeable that a design of sintering die and punch assembly made of graphite is extremely important subject to joule heating according to the sintering progress, interaction of powder material, system resistivity, and the function as direct heating elements in order to assume the role of maintaining the homogeneity of sintering temperature [28,30–32].
FIGURE 10  Basic mechanism of neck formation by spark plasma sintering.
4.4. DC Pulse Generator

Presently, there are two basic types of DC pulse generator for SPS apparatus which are thyristor-type and inverter-type power supply. The waveform, max./min. ON-time/OFF-time pulse width, peak current, frequency, duty factor settings, control system, and energy consumption are different; however, each system has different advantages individually. Therefore, pulse generator should preferably be chosen to meet the estimated purpose and usage of SPS processing. Majority of SPS systems working in universities, national institutes, and private companies use the thyristor-type pulse generator due to its rich reference database on SPS and higher reliability of power supply hardware. The inverter type with PWM (pulse width modulation) control provides a low power consumption and compact space-saving advantages, so that it has a high possibility as economical and of low-cost production. The development on pulse generator is still in progress. Both thyristor and inverter types,
ON–OFF pulse width, duty factor, and frequency are designed selectively. Figures 14 and 15 show examples of typical ON–OFF pulse waveforms and different pulse widths and ratios.

4.5. Vacuum Exhausting System

SPS processing is generally under a vacuum sintering condition. The system allows operating with an inert gas such as argon gas or nitrogen gas, and atmospheric air conditions. By using a rotary vacuum pump with or without mechanical booster pump, pressure ranging from $1 \times 10^5$ Pa (air pressure) to 5–6 Pa within 15 minutes will normally be applied. Depending on the sintering purpose and usage, SPS vacuum exhausting system can also employ a diffusion pump for $5–10 \times 10^{-3}$ Pa high vacuum or a dry pump against a sticky gas.

4.6. SPS Sintering Temperature Measurement and Difference

In SPS process, the sintering temperature is generally measured at the inside of the graphite die wall by thermocouple or surface by pyrometer, and not in the powder directly. Thus, there exist temperature differences between the measured value and an actual filled powder temperature. Figure 16 is an experimental example on nickel powder material indicating 110–120 K difference at 300 sec. The
heating rate was 2.4 K/sec and the open circle in the graph shows sliding surfaces of the graphite punch and die. In accordance with various experiments, it was found that the measured sintering temperatures in both of metals and ceramics by SPS were about 50–250 K lower than the inside of filled powder material. As to nonequilibrium rapid processing, therefore, SPS sintering temperature in SPS data implies this measured temperature [28,30].

4.7. SPS System Machine Layout and Installation

Figures 17 and 18 show a medium-size SPS system machine overview and 1-MN large-size SPS apparatus machine dimensions and layout drawings. As observed, it does not require a large space and special foundations for installation. Approximately 5.5 m (width) by 5.5 m (depth) in length and 3.5 m in height are sufficient. Cooling water, compressed air, and electric facilities are necessary in operation. Cooling water requires city water pressure level of 0.15–0.3 MPa with a flow rate higher than 20–35 l/min. For a DC pulse current output max. 10,000 A, the AC input power requires 175 or 250 kVA (3-phase, 200/220 V, 600 A). Compressed air supply facility requires a pressure level of 0.3–0.5 MPa (3.0–5.0 kgf/cm²). It is usually recommended to make every effort to prevent the generation of electro-magnetic noise around SPS systems that could result in faulty operation such that images on CRT computer screens and electron microscopes located nearby may be affected by interference. Approximate weights of the 1-MN system units are sintering machine 6500 kg and DC pulse generator 2300 kg. The system was designed with special consideration complying with the specifications required for ease of use and safety in comprehensive research on new materials.
4.8. SPS Production Systems

The development of production-type SPS apparatus involves various considerations regarding production strategies which are cycle time, cost and development of optimum systems, technologies to support scale expansion, automation, mass production and NC (numerical control) system, development of process technologies for high functionality, reproducibility, and uniformity, and structure control and development of 3D near-net shape forming. Figure 19 shows the large-size SPS apparatus for the fabrication of sputtering target materials. As shown in Figure 20, in order to verify SPS productivity, five basic styles of production-type SPS machine systems have been developed: (1) multi-head SPS systems, (2) batch-type SPS systems, (3) tunnel-type SPS systems, (4) rotary-type SPS systems, and (5) shuttle-type SPS systems. In addition to these, scaling-up process, automatic handling and powder stacking equipment for materials, and processing optimizer with LCD were also developed. As a result, SPS has become an applicable technique for use as an influential means of industrial manufacturing with high-value-added products [33–38].

5. EXAMPLES OF SPS PROCESS APPLICATIONS

5.1. SiC with Al2O3/Yb2O3 Consolidated by SPS and its Mechanical Properties

Rapid sintering is one of the remarkable features of SPS. In order to clarify the advantages and mechanism of SPS method, it was examined on a monolithic ceramic material containing additives. As starting powder materials, 0.28 μm β-SiC with 3 mass%Yb2O3–5 mass%Al2O3 powder was used. Densification behavior and mechanical properties of SPS-consolidated SiC were compared with those of hot-pressed SiC. Figure 21 shows that nearly full densification was attained at about 2023 K for SPS-consolidated SiC and at about 2173 K for hot-pressed SiC, respectively. In Figures 22, 23 and 24, the optimal bending strength, Vickers hardness, and fracture toughness (not shown) of SPS-consolidated SiC were 720 MPa, 25 GPa, and 4.0 MPa·m^{1/2}, respectively, and these values were higher than those (640 MPa, 23.5 GPa, and 3.5 MPa·m^{1/2}) of hot-pressed SiC, indicating that SPS consolidation can improve the bending strength without degradation of fracture toughness. Pressurization shows a good effect on the densification enhancement remarkably as given in Figure 24. From XRD and SEM (Figures 25, 26) and Raman scattering analysis (not shown), it is suggested that
Chapter | 11.2.3 Spark Plasma Sintering (SPS) Method, Systems, and Applications

**FIGURE 21** Dependence of relative density on the sintering temperature in the SPS- and HP-consolidated SiCs.

**FIGURE 22** Dependence of bending strength on the sintering temperature in the SPS- and HP-consolidated SiCs.

**FIGURE 20** Typical automated SPS production systems.
3C-type disordered structure is preserved in the SPS-consolidated SiC and the improvement of the bending strength without degradation of fracture toughness is attained by the preservation of the 3C-type disordered structure as the SPS effect; even enhancement of grain growth of SPSed SiC was larger [39]. When different additives of Al₂O₃/Y₂O₃ were used, the SPSed SiC compact had shown a similar tendency as the SPS effect [40].

5.2. Consolidation of Nano-Al₂O₃, Phase Transformation, and Grain Growth

It is known that the SPS features provide a microstructure-controlled sintering. Structurally tailoring effect in SPS processing was verified in the consolidation of alumina. A capability of SPS processing for the generation of dense nanostructure was investigated by consolidating γ-alumina.

FIGURE 23 Dependence of Vickers hardness on the sintering temperature in the SPS- and HP-consolidated SiCs.

FIGURE 24 Dependence of relative density on applied pressure in the SPS- and HP-consolidated SiCs.

FIGURE 25 XRD patterns obtained from raw powder, SPS- and HP-SiCs.
nano-powder and z-alumina powders. Up to the present time, SPS consolidation of z-alumina has been performed for various purposes such as the attainment of high bending strength [41], the estimation of microstructure inhomogeneity [42], and of the relation between process factors and densification behavior [20]. Further, generation of nanostructure due to suppression of grain growth has been indicated in z-alumina SPS-consolidated at very rapid heating conditions [43,44]. Here, it is shown that the formation of nanostructured dense alumina is possible in conventional SPS conditions by using nano-z-alumina powder. In this study, first, γ-Al2O3 powder particle size of 37 nm, and z-alumina of 0.2 and 0.5 μm were consolidated by the SPS method at a heating rate of 160–197 K/min under the pressure conditions of 49 MPa or 690 MPa. Relative density and powder grain size against the consolidation temperature are shown in Table 3. In Figure 27, the z-alumina was fully densified at 1673 K under the pressure condition of 49 MPa, while γ-alumina full densification was attained at 1773 K. Figure 29 shows the phase transformation of γ-Al2O3 prepared at 1173–1473 K under 49 MPa. When the pressure increased to 690 MPa (Figure 28), nanostructured 98% dense alumina with the grain size less than 500 nm was obtained under the conditions of 1173 K.

As shown in Figures 28 and 30, it was found that density decreased at 973 K due to phase change to δ- and θ-phase. Under a high pressure of 690 MPa, the sintering was done at nearly full densification of γ-alumina at 600 K lower temperature [45]. Second, z-alumina powder with a grain size of 100 nm used was consolidated by SPS method at a heating rate of 123 K/min under the pressure conditions of 30 MPa or 100 MPa. Nanostructured dense alumina with the grain size less than 300 nm, 99.18%TD, was synthesized under the conditions of 1423 K and 100 MPa. At the same 1423 K with 30 MPa, the sintered one was reached at 86.9% TD. To obtain 98.3%TD dense alumina compact

### Table 3 SPS Sintering Conditions of Alumina

<table>
<thead>
<tr>
<th>Al2O3 starting powder materials</th>
<th>SPS sintering pressure (MPa)</th>
<th>SPS sintering temperature (K)</th>
<th>Sintering keeping time (sec)</th>
<th>Sintering heating-up (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ-Al2O3</td>
<td>49</td>
<td>1173</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>1273</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1473</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1573</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1673</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1773</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>690</td>
<td></td>
<td></td>
<td>180</td>
<td>600</td>
</tr>
<tr>
<td>923</td>
<td></td>
<td></td>
<td>180</td>
<td>600</td>
</tr>
<tr>
<td>973</td>
<td></td>
<td></td>
<td>180</td>
<td>600</td>
</tr>
<tr>
<td>1073</td>
<td></td>
<td></td>
<td>180</td>
<td>600</td>
</tr>
<tr>
<td>1173</td>
<td></td>
<td></td>
<td>180</td>
<td>600</td>
</tr>
</tbody>
</table>

| α-Al2O3 (0.5 μm)               | 1473                        | 60                            | 540                          |
|                                | 1573                        | 60                            | 540                          |
|                                | 1673                        | 60                            | 540                          |
|                                | 1773                        | 60                            | 540                          |

| α-Al2O3 (0.2 μm)               |                             |                               |                              |                             |
|                                |                             |                               |                              |                             |

**FIGURE 26** SEM photos of etched surface (a) SPS-consolidated SiC (2073 K, 30 MPa) and (b) HP-consolidated SiC (2273 K, 30 MPa).
with 30 MPa, a higher temperature of 1473 K was required. In Figure 31 (fractured surfaces of FE-SEM), it is indicated that the difference in the applied pressure at a consolidation temperature of 1423 K drastically changes the microstructure in the SPS-consolidated \( \alpha\)-Al\(_2\)O\(_3\); that is, full densification and nanostructure can be attained by SPS consolidation when the preferable pressure condition is selected without selecting a very high heating rate. The heating rate is more important for the synthesis of dense nanostructured and highly transparent alumina, although applied pressure is also correlated with generation of the structural tailoring. The findings suggest that it is important to select an optimum combination between SPS parameters and heating rate, which is delicately effective for generating new micro/nanostructures in the SPS-consolidated ceramics [46]. High-temperature short-period SPS sintering is expected to provide almost all ceramic materials with new characteristics and sintered effects which are different from those obtained by the HP and HIP processes.
5.3. Examples of Synthesized Functionally Graded Materials (FGMs) by SPS

SPS has been used successfully to synthesize a wide range of bulk FGM materials with three to eight or up to 19 intermediate mixed layers, including systems of ZrO$_2$/SUS (stainless steel), ZrO$_2$/TiAl, ZrO$_2$/Ni, Al$_2$O$_3$/SUS, Al$_2$O$_3$/Ti, Al$_2$O$_3$/Ti-6Al-4 V, WC/SUS, WC/Co, WC/Ni, Cu/SUS, SiO$_2$ glass/SUS, Al/polyimide resin, Cu/phenol resin, Cu/polyimide resin materials, etc. [47,48]. All were sintered in ten to twenty-five minutes heating-up and keeping time for a diameter of 20–50 mm after the start of the process. Figure 32 shows a schematic illustration of SPS temperature-gradient die assembly. Figure 33 shows typical examples of bulk FGM compacts synthesized by SPS: from the left, a ZrO$_2$(3Y)/SUS stainless steel compact with 6 interlayers, a ZrO$_2$(3Y)/nickel compact with 7 interlayers, and an aluminum/polyimide compact with 3 interlayers; on the right, an Al$_2$O$_3$/titanium compact with 3 interlayers. Full dense and no microcracks were detected in the sintered compacts obtained by this process. An example of ZrO$_2$(3Y)/SUS stainless steel compact, a 3 mol% yttrium partially stabilized zirconia (PSZ) powder, a SUS410L stainless steel (SUS) powder, and mixed powders of the two as intermediate layers were stacked on a graphite temperature-gradient die of 20-mm internal diameter. The SUS powder has an average particle size of 9 μm, while the PSZ powder is of granulated particles with an average particle size of 50 μm (its crystalline size is 350 Å). Sintering pressures used were from 20 to 40 MPa, SPS temperatures of 1243–1293 K, with a temperature rise rate of 50 K/min. Temperatures were measured near the SUS layer [49–53].

5.4. Demands from Industries on SPS Processing Techniques

5.4.1. Sputtering Target Material and Fabrication of Large-size Metal and Ceramic Compact

With the SPS process, highly dense sintered products can be fabricated at a much lower temperature and with shorter heating-up and holding times compared with hot pressing.
Figure 34 shows typical examples of large-sized metallic sputtering target material with a diameter of φ 350 mm produced by large-size SPS apparatus. The productivity is approximately 7–8 times higher than conventional sintering of HP and HIP processes. Because of the obtained finer grain size of the SPSed sputtering target materials, it can provide superior performance such as no splashing phenomenon in the sputtering process. Large-sized oxide (Al₂O₃, ZrO₂, and SiO₂), carbide (WC, SiC, and B₄C), nitride (Si₃N₄), and boride (TiB₂) ceramic materials were also fabricated homogeneously with finer grain size and almost full density [54–58]. The samples shown in Figure 35 were investigated by SEM observation that nearly no residual micropores or cracks were detected in the large-size disk-shaped sintered compacts.

5.4.2. Si₃N₄/Al₂O₃ Composite Compacts for Homogenizer Component

Figure 36 is an example of a fine ceramic sintered compact for homogenizer’s wear-resistant parts. Rapid sintering process can minimize grain growth and manufacture a harder sintered compact with a high relative density of 99–100% and a hardness of over 20 GPa. Pressurized sintering allows for the direct sintering from powders to near-net shape dimensions of ring or cylindrical shape, and eliminating the green body compaction process of normal pressureless sintering. The sintering conditions were an SPS temperature of 1673–1873 K with heating-up and holding time approximately 15–20 minutes totally and an applied pressure of 30–50 MPa. The tolerances of as sintered parts ensure within 0.2–0.3 mm, so that it is easy to obtain a final finishing accuracy by conventional mechanical grinding as post-process.

5.4.3. Pure WC (tungsten carbide) Aspheric Glass Lens Mold

Figure 37 is an example of SPsed component in an actual commercial use in the optic industry which is aspheric glass lens molds made of a binder-less pure-tungsten carbide (WC single phase, Hv2600) material without additives and not containing the W₂C phase. They were homogeneously consolidated in nanostructured fine grain size. By using ultra-fine grinding machine, the super finishing of mirror surface roughness of Ra5–10 nm can be obtained to fit
with digital camera lens application. The aspheric glass
lens mold consists of three pieces: upper punch, lower
punch, and sleeve die part. The advantages of SPSed pure
WC material are solid-phase sintering, attaining finer grain,
and higher oxidation resistance compared to conventionally
produced other binder-less WC materials. Under 973 K in
atmospheric furnace with 10 hours running test, it resulted
in a 30–60% better oxidization in volume (g/cm²).

5.5. 3-Dimensional Complex Near-Net-
Shape Forming of Al₂O₃ Blasting Nozzle

Figure 38 shows an example of an Al₂O₃ ceramic nozzle for
sandblasting machine. The actual lifetime was investigated
under the same blast operating conditions. The SPSed
Al₂O₃ nozzle achieved 10 times longer life than a conven-
tionally sintered one made by atmospheric pressureless
sintering furnace. As sintered by SPS, the Vickers hardness
of 2100–2200, and 3-D net-shape accuracy with Ra0.64-
μm surface roughness can be obtained. In order to eliminate
the usual mechanical grinding process by SPS, the
production method has been developed.
FIGURE 37  Examples of pure WC aspheric glass lens mold materials.

FIGURE 38  Comparison of SPSed nozzle and other commercial products.

<table>
<thead>
<tr>
<th>Material: Al₂O₃ Length: 60mm Outer dia. (Tapered): ø 30/15mm Inner dia. (Straight): ø 6mm</th>
<th>General Commercial Products</th>
<th>SPS Sintered Nozzle</th>
<th>Advantage of SPS sintered Nozzle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vickers Hardness (HV)</td>
<td>900–1100</td>
<td>2100–2200</td>
<td>High-Hardness</td>
</tr>
<tr>
<td>Relative Density (%)</td>
<td>91</td>
<td>100</td>
<td>High-Density</td>
</tr>
<tr>
<td>Surface</td>
<td>Rough</td>
<td>Mirrored surface</td>
<td>Mirror finished surface by Near Net Shape</td>
</tr>
<tr>
<td>Lifetime</td>
<td>1 day</td>
<td>10 days</td>
<td>10 times longer Lifetime</td>
</tr>
</tbody>
</table>
5.1. Improvement of the Super-plasticity and Specific Strength of Al–high-Si Alloy Materials

The use of SPS process can easily attain nano-structured dense high-silicon/aluminum alloys. The rapidly solidified Al-Si powder (average particle size 120–150 μm, silicon 12–17% or higher and containing a nano-crystalline structure) produced by water atomization process, was successfully consolidated into large sized cylindrical compact with a diameter of 60–120 mm and a thickness of 40–60 mm. By using a steel die for high-pressure SPS, the consolidation process is rapidly performed and maintains nanometric-sized crystals. A sintered compact with a relative density of almost 100% was obtained under atmospheric SPS sintering conditions at SPS temperatures of 723–773 K, applied pressures of 100–150 MPa, and a temperature rise and holding time of about 20 minutes. The TEM micrograph of nanostructure compact in Figure 39 shows that the grain diameter of the sintered compact was between 600 and 800 nm. When the size of 60 mm diameter and 40 mm in thickness nano-structured bulk sintered compact was formed into three dimensional (3-D) shape of automotive engine piston component (Figure 39) by a high-speed forging press machine, it was possible to perform a full forming for lengths of 75 mm within 15 to 20 seconds per stroke and the strain rate of $10^{-3}/s$ or higher. These results indicate that it performs the phenomenon of super-plasticity in the material, and improving the ductility and elongation of the sintered compact. The SPSed material attained a tensile strength of about 350 MPa at normal room temperature, and it is approximately 1.5 times stronger than a conventionally forged component. This Al–Si alloy with nanostructural material is expected to find wide applications in electrics, electronics, automotive, and other industrial fields [59].

5.6. Porous Materials by SPS

SPS also demonstrates good features on a porous material fabrication in both of metallic and ceramics. Figure 40 shows pure titanium beads and ZrO$_2$(3Y) beads used in porous samples prepared by SPS. They are expected in commercial applications for a bioreactor, a filter, an artificial bone joint, a vent core material for plastic molding die, EV (electric vehicle)/HEV (hybrid electric vehicle) cars, SOFC (solid oxide fuel cell) batteries, thermo-electric semiconductors, a heat spreader of thermal conductive component, and others. Applying lower sintering pressure range of 0–5 MPa, it is possible to attain high porosity with higher bonding strength at neck portion even on an active material of Ti or Al wearing TiO$_2$ or Al$_2$O$_3$ oxide films on the surface of particles compared with other processes [60,61].

5.7. Development of Large-size Functionally Graded Materials (FGMs)

5.7.1. Fabrication of φ 150 mm ZrO$_2$(3Y)/Stainless Steel Large-size FGMs

A ZrO$_2$(3Y) submicron-size partially stabilized zirconia powder containing 3 mol% Y$_2$O$_3$ and SUS410L stainless steel powder with average particle sizes of 9 μm and 60 μm were used as starting materials in this study. The composition of the stainless steel powder was 0.03 wt% C, 0.89 wt% Si, 0.13 wt% Mn, 0.011 wt% P, 0.012 wt% S, 0.09 wt% Ni, and 13.06 wt% Cr. By stacking nine kinds of mixed-composition powders with ZrO$_2$(3Y)/stainless steel ratios of 90/10, 80/20, 70/30, 60/40, 50/50, 40/60, 30/70, 20/80, and 10/90 vol.% as interlayers between the 100% front and back layers, a total of 11 gradient layers were laminated (Figure 41). As a result of the optimization of sintering conditions, with applied pressure of 30 MPa, SPS temperature at 1253 K, and holding time of 1 minute, the healthy SPSed compact was obtained within 58 minutes, including heating-up time, holding time, and cooling time. A large-sized ZrO$_2$(3Y)/stainless steel system FGM with a diameter of 150 mm and thickness of 15 mm having 9 interlayers was synthesized free of cracks, no distortion, and no delamination, as shown in Figure 42. In Figure 43, it is suggested that the optimized SPS conditions of higher sintering pressure
and temperature are necessary to prepare homogeneous dense large-sized bulk FGM from the center (00%) to the near perimeter (75%) of the specimen [12,49].

5.7.2. Development of Fine-WC/Co Hard-alloy FGMs

In order to fabricate cobalt content compositionally graded cemented carbide materials, initial research focused on obtaining homogeneous large-size monolayer WC/Co hard alloys. Although fabrication of hard alloys usually takes many hours, with the development of a new automated SPS system, it is now possible to sinter such materials in a significantly shorter period of time taking advantage of the characteristics inherent in SPS rapid sintering technology. Figure 44 shows the system configuration and an outside view of the automated SPS system. Generally, dimension size and shape effects of SPS tend to fluctuate in
hardness around the outer portion. This degradation increases as the size of the sintered compact becomes larger. Utilizing this tunnel-type automated SPS system and optimized SPS conditions, square-shaped large-size WC/Co cemented carbide hard alloy with dimensions of 70 mm(width) × 100 mm (length) × 5–20 mm (thickness) were homogeneously fabricated in a shorter sintering time and with a finer grain size than with conventional sintering methods while maintaining high quality and repeatability every 10 minutes or less per piece, as shown in Figure 44 [62]. Table 4 shows the typical mechanical properties of WC/Co tungsten carbide hard alloys sintered by SPS. A fine tungsten carbide powder of an average particle size less than 0.5 μm containing 0–6 weight percent cobalt was used as a starting powder material. Figure 45 is a comparison with other commercial products. The fine-WC/Co hard
alloys obtained by SPS had higher hardness, transverse rupture strength, and fracture toughness than those obtained by conventional methods. Based on this study of the optimum SPS sintering conditions, the square-shaped large-size WC/Co FGMs with dimensions of 100 mm (width) \times 100 mm (length) \times 40–60 mm (thickness) were successfully synthesized within an hour. Figures 46 and 47 are the WC/Co FGM and the relation between micro-Vickers hardness and distance from surface to bottom on the cobalt content graded cemented carbide compacts fabricated by SPS.

TABLE 4 Typical mechanical properties of WC/Co hard alloys sintered by SPS

<table>
<thead>
<tr>
<th>Product cord name</th>
<th>Co content (wt%)</th>
<th>Grain size (μm)</th>
<th>Density g/cm³</th>
<th>Hardness mHV</th>
<th>Transverse rupture strength kg/mm²</th>
<th>Fracture toughness K1c</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC10</td>
<td>&lt;4</td>
<td>&lt;0.5</td>
<td>15.0</td>
<td>2150</td>
<td>270</td>
<td>6.5</td>
</tr>
<tr>
<td>TC20</td>
<td>&lt;6</td>
<td>&lt;0.5</td>
<td>14.8</td>
<td>2050</td>
<td>300</td>
<td>7.3</td>
</tr>
<tr>
<td>M78</td>
<td>0</td>
<td>&lt;0.2</td>
<td>15.4</td>
<td>2600</td>
<td>150</td>
<td>5.0</td>
</tr>
</tbody>
</table>

5.7.3. Examples of SPSed WC/Co and WC/Co/Ni FGM for Industrial Applications

As a typical high wear-resistant material, WC/Co or WC/Ni system cemented carbide is now widely used in various press-stamping dies and cutting tools for industrial applications. The FGMs by SPS process presents many excellent advantages including simple control of sintering energy, ease of handling and elimination of sintering skills, as well as high sintering speed and producibility, safety, space saving, energy saving, and low equipment cost. However, from the viewpoint of integrated process of non-equilibrium processing and dynamic sintering, it is suggested that the simultaneous development concept that of SPS software development technology involving the powder preparation and SPS machine hardware technology should be studied at a time to realize effective manufacturing method for the practical industrial applications. Figure 48 shows a square-shaped large-size WC/Co/Ni FGM fabricated by SPS. The sintered bulk FGM was machined by numerically controlled wire-cut electrical discharge machine to cut off from the large bulk material to the small pieces of FGM block. As shown in Figure 49, the small pieces of WC/Co FGMs were usually ground to form the specified profiles, accuracies, and tolerances as press-stamping dies.
and punches by a precision type of grinding machine. Concerning its mechanical properties, the SPSed fine-WC/Co FGM hard alloy had a higher hardness on topcoat layer and higher strength and fracture toughness in the bottom side than mono-WC/Co hard-alloy materials. When assembled as press-stamping progressive die and punch, it achieved approximately 3.5–10 times longer lifetime compared with conventional commercial cemented carbide products. In addition, weldable WC/Ni FGM for an extruding machine was also developed as shown in Figure 50. The lifetime was over 3 times longer which was more than 3000 hours than the conventional screw of 800 hours [63].

### 5.7.4. Other Industrial Applications

As shown in Figure 51, the typical example of the SPSed mass products has already been realized in practical use in the field of cutting tools and wear-resistant materials industry. By 2-shift or 3-shift day-and-night continuous working time in the production factory, multiple fabrication of 15–20 plates per batch were carried out simultaneously.

---

Figure 47: Large-size WC/Co FGM fabricated by SPS (100 mm × 100 mm × 40 mm) and profiles of micro-hardness on Co content graded cemented carbide by SPS.

Figure 48: Large-size WC/Co/Ni FGM and cut-off sample by wire-cut electrical discharge machine.

Figure 49: Example of press-stamping die and punch made of WC/Co FGM for electronic component.
The metal-bonded dicing blade (cutting tool) that of 100/150-mm diameter and 0.35/0.4-mm thickness was produced providing the flatness within \( \pm 20 \mu m \) with minimal residual stresses. Sintered WC/Co plate of relative density 99–100\% and Young’s modulus of 500–580 GPa were attained under SPS temperature of 1473–1523 K. To apply the SPS processing in their daily production, the results have clearly demonstrated its commercialization, a new functionality, versatile advantages, and cost-effectiveness as the actual product. On the other hand, it is expected that 3-D forming method and titanium/HAP will be applied for parts production and biomedical applications from R&D prototype level to practical use in the near future [47,64–67]. Although SPS research and development on other categories of materials in details were not introduced in this paper, such as FRC/FRM composite materials [68–70], electronic materials of SiGe, Bi₂Te₃, FeSi₂, CoSb₃, MnSi₂, and Mg₅Si systems for thermo-electric semiconductors [71–73], Nd–Fe–B, Sm₂Co₁₇, and ferrite for magnetic materials [74,75], MgB₂ superconducting materials [76], BaTiO₃ and PbTiO₃ dielectric materials [77,78], shape memory alloys, solid cell materials, glassy metals, optically functional materials, and nanocrystalline materials, they are a promising candidate for industrial SPS applications today [79–85].

6. SUMMARY AND OUTLOOK

In this paper, SPS fundamentals and wider availability of SPS and its related processing effects based on DC pulsed high current energizing were mainly introduced in the synthesis of various ceramics, metals, metal/ceramics, and ceramics/ceramics composites. Especially, typical advantages of SPS processing were indicated in the synthesis of functionally graded materials (FGMs) and nanocrystalline materials. In addition to these, actual product applications in industries were also introduced. Because of the versatility of SPS, it should be noted that the remarkable rapid growth in the number of presented papers and patents in the last two decades resulted from the worldwide spread of the SPS technology in both the scientific community and the industrial sector. SPS features an electrical energy concentration at areas where current flows easily. In terms of the high energy density and dynamical processing, the study on these characteristics will lead us to the successful advancement and expanded SPS applications for reasonable commercial production. As future prospects of SPS technology,
Figure 52 shows recent demands from industries on SPS technologies. The SPS has a high potential to be a major manufacturing method in the various industries, provided the developments necessary, therefore, are now focusing on both high-value-added small-scale and mass production.

REFERENCES

[8] Cremer GD. Sintering together powders metals such as bronze, brass or aluminum. U.S. Patent No.2,355,954; 1944.
Chapter | 11.2.3  Spark Plasma Sintering (SPS) Method, Systems, and Applications


[72] Chen LD, Huang XY, Zhou M, Shi X, Zhang WB. The high temperature thermoelectric performances of Zr0.5Hf0.5Ni0.8 Pd0.2Sn0.99Sb0.01 alloy with nanophase inclusions. J Appl Physics 2006;99:064305.


1. FUNDAMENTALS OF FGMs

1.1. Concept

The term “functionally graded materials” contains two important words: “functionally” and “graded.” The word “functionally” modifies “graded.” These refer to not only simple functional materials but also to graded materials. The general concept of functionally graded materials is materials with graded functions inside them. Functionally graded materials are abbreviated as FGMs hereafter.

Original FGMs were one-body materials in which their thermal expansion coefficient was changed gradually between the front and back sides. This was achieved by changing their composition from metallic to ceramic. The purpose of FGMs was to relax the internal thermal stress at elevated temperatures.

Figure 1 shows the difference in compositions and properties between an ordinary composite material and FGM. There is a distinct interface between metals and ceramics in an ordinary composite material, but not in an FGM. This difference corresponds to the distribution of properties such as thermal expansion coefficient, thermal conductivity, and thermal resistance. An ordinary composite material contains a sudden change in properties at the interface, while an FGM presents a gradual change inside it. The difference in thermal expansion coefficients at the interface causes internal thermal stress at elevated temperatures, sometimes leading to the destruction of the interfaces. As shown in Figure 1, an FGM reduces thermal stress by almost 30% and can prevent destruction of the interface.

The original concept of FGMs for reducing thermal stress was extended to general materials in 1988, and an organization named “Functionally Graded Materials Forum, Japan,” was founded. The expanded concept refers to one-body materials with a continuous or stepwise change in functions inside them [1]. FGMs have covered not only heat-resistant materials but also structural materials, biomaterials, semiconductors, and electrode materials.

1.2. History

The birth of FGMs was deeply related to space-plane plans, which was proposed by the Science and Technology Agency of Japan in the mid-1980s. The plan was that Japan take an important part in the global space exploitation by the beginning of the 21st century.

The space plane is a combination of a jet plane and rocket that shuttles between the earth and outer space. The greatest difference from the space shuttle of the United States is that both a jet engine and a rocket engine are equipped. The space plane is supposed to take off from the runway on the earth without any adaptor boosters and shuttle approximately 30 times between the earth and the outer space without repair. When it goes through the earth’s atmosphere at a supersonic speed, the nose of the body and the leading edge of the engine are heated up to high temperatures close to 1800 °C, as shown in Figure 2 [2]. The maximum temperature of the space shuttle is about 1500 °C at the leading edge of the vertical tail during atmospheric entry. Even though the space shuttle was exposed to lower temperatures by 300 °C than the space plane, problems of the body were sometimes reported.

The problems of the space shuttle mainly arose from separation and loss of the thermal protection tiles on the
body. The space shuttle is partially covered with thermal protection tiles of carbon-fiber-reinforced sponge-like carbon (C/C) composites, which prevent overheating of the space-shuttle’s interior during atmospheric entry. Black parts of the body are the tiles. The tiles are of a light weight (specific gravity is \( \approx 0.2 \)), and they have high thermal insulation.

What actually happens to the tiles during atmospheric entry? Even if the body is partially covered with the tiles, the body, which is made up of metals, get heated to high temperatures by heat generated by friction with the atmosphere. Generally, materials expand with temperatures; the thermal expansion coefficients of iron, aluminum, and C/C composites are \( 11.7 \times 10^{-6} \), \( 23.9 \times 10^{-6} \), and approximately \( 3 \times 10^{-6}/^\circ C \) near room temperature, respectively.

Changes in the metal body fixed with the tiles due to heating are shown schematically in Figure 3. The difference in thermal expansions causes thermal stress at the joined interface between the tiles and body. When the thermal stress is larger than the joining strength at the interface, the tiles separate from the body. This is the cause of problems in space shuttles.

For the purpose of preventing the tiles from separating, it is necessary for thermal stress not to concentrate at the joined interfaces between the tiles and the body but to distribute uniformly near the interfaces. FGM is a new concept for heat-resistant materials, which was proposed to address this need. The thermal expansion coefficient changes gradually from a large value for metals to a small value for ceramics by changing the composition continuously from metal to ceramic. The thermal stress induced by the difference in thermal expansion coefficients was reduced by \( >30\% \).

2. NATURAL MATERIALS WITH GRADED STRUCTURES

We have many kinds of materials with functional gradients, in nature. Especially most living bodies consist of materials with graded structures. This is the reason why biological cells are characterized by adapting to external stimulus. This adaptation has induced generation of graded structures against the external force in evolutional processes.

(1) Bamboo

Bamboos have a diameter of \(<20\, \text{cm}\), and they grow to a height of \(>10\, \text{m}\). The ideas that we can get from bamboos are that they are ‘strong,’ ‘lithe,’ and ‘easy to split.’ These characters are quite different from those of other trees.
Figure 4 shows that a bamboo consists of the epidermis, fundamental tissue, and bundles. The bundles are of bundle sheaths (black in the figure), inner tubes, and metaphloems. The bundle sheaths enclose the inner tubes and metaphloems to protect them. The bundle sheaths are known to have high elasticity and strength. The tensile strength is comparable with that of iron steels. As shown in Figure 4, the volume fraction of bundle sheaths increases from the inner part toward the epidermis. Almost 90% of the tissues near the epidermis are occupied by bundle sheaths. The high density of bundle sheaths makes the epidermis hard and strong, and the graded structure increases the toughness of bamboos to prevent them from splitting and getting crushed. The hard epidermis also protects the inner tissues from the outside. It can be said that the graded structure of bundle sheaths is a bamboo itself.

(2) Human bone

Figure 5 shows the structure of the human bone [3]. The human bone consists of compact and cancellous bones. There are no visible pores in the compact bone, which is the surface part of human bone. On the other hand, the cancellous bone is sponge-like with many small pores. The inner part of the human bone is the cancellous bone. The small pores offer spaces to the bone marrow that works to increase blood. In the case of birds, most of the pores in the cancellous bone are empty, which reduces the weight of the bone (called hollow bone).

The human bone changes from the cancellous bone to the compact bone in the direction of the inner region to the surface. This is a graded structure. What is the benefit of the graded structure of the human bone?

The graded structure of the human bone is similar to that of the bamboo. The compact bone with a high bone density contributes to the hardness and strength of the human bone. The cancellous bone is lightweight because of its honeycomb structure. The external stress should be transferred smoothly and uniformly from the compact bone to the cancellous bone. The graded structure also contributes to smooth stress transfer between the compact and cancellous bone.

3. FABRICATION PROCESSES

The functional change of FGMs can be realized by the compositional change according to the functional design. There are many kinds of processes applied to FGM fabrication. The processes are classified into three based on the starting materials: gaseous processes, liquid processes, and solid processes. In this chapter, the typical fabrication processes for ceramic system FGMs are introduced.

(1) Gaseous processes

Chemical vapor deposition (CVD), physical vapor deposition (PVD), plasma spraying, and surface reaction are included.

(a) CVD process

CVD is a process for depositing metal or ceramics on the substrate by the thermal decomposition of halide gases including metals or semimetals. The composition or
structure of deposits can be controlled by deposition parameters such as the kind of gas, gas flow rate, and thermal decomposition temperature. For example, the reaction products of SiCl$_4$ and CH$_4$ gases are SiC, C, and their mixtures according to the mixing ratio of gases. The SiC/C FGM can be obtained by changing the mixing ratio of gases during the deposition process [4].

(b) PVD process

PVD is a process for depositing metals or ceramics on the substrate by evaporating the source metal. When reaction gases such as C$_2$H$_2$, N$_2$, and O$_2$ are introduced into the deposition chamber during the deposition process, metal carbide, nitride, and oxide are formed. For example, the reaction products of Ti vapor with C$_2$H$_2$ are TiC, Ti, and their mixtures according to the C$_2$H$_2$ flow rate. The TiC/Ti FGM can be obtained by changing the flow rate of C$_2$H$_2$ during the deposition process [5].

(c) Plasma spraying process

Plasma spraying is a process for depositing semisolid metals or ceramics on the substrate by plasma jet. There are two methods: the one-gun method and the two-gun method. Figure 6 shows a schematic view of the one-gun method [6]. The mixture of metal and ceramic powders is fed into the one plasma gun and sprayed on the substrate. The FGM can be obtained by changing the mixing ratio of metal and ceramic powders during the spraying process. Figure 7 shows the schematic view of the two-gun method [7]. Metal and ceramic powders are fed into different plasma guns, and the two plasma sprayings are performed simultaneously. The FGM can be obtained by increasing the feeding rate of ceramic powders and decreasing that of metal powders during the spraying process.

The PSZ/Ni$_x$Cr$_x$Al$_x$Y FGM fabricated by the two-gun method is shown in Figure 8. The gray area in the figure represents PSZ and the white area represents Ni$_x$Cr$_x$Al$_x$Y. The volume fraction of PSZ to Ni$_x$Cr$_x$Al$_x$Y increases in the thickness direction toward the surface.

(d) Surface reaction process

Surface reactions are processes for diffusing reactive gases into the base material to form reaction products at the surface of the base material. Typical processes are carburizing and nitriding. These processes have been applied to prepare cutting tools with graded surface coatings. Figure 9 shows the distribution of the micro-Vickers hardness in the thickness direction for the nitrided hard metal of TiCN–WC–Co–Ni [8]. The surface area is mostly occupied with TiCN and the inner region is a mixture of TiCN, WC, and Co–Ni. The micro-Vickers hardness is high at the surface area and decreases with increasing distance from the surface.
process of sintering. Metal and ceramic slurries are prepared by blending metal and ceramic powders of various mixing ratios with appropriate binder materials and then deforming under low vacuum. The thickness of thin sheets is 10 µm to 2 mm. The thin sheets are layered in the order of composition and pressed to form a step-by-step graded compact. The FGM is obtained by debinding and sintering the compact. The overall process is shown in Figure 10 [9].

(b) Electrodeposition process

Electrodeposition is a process for depositing a mixed film of metals and ceramics on the electrode from a plating solution containing ceramic powders. The volume fraction of ceramics in the deposited film is controlled in the range from 0% to approximately 25% by the concentration of ceramic powders in the plating solution. The FGM can be obtained by changing the concentration of the ceramic powders in the solution during the deposition process [10].

(c) Centrifugal casting process

Centrifugal casting is a process for solidifying the mixture of molten metal and ceramic powders in a die under centrifugal force. When the relative density of ceramic powders is higher than that of the molten metal, the ceramic powders move in the centrifugal direction in the die under centrifugal force. In the opposite case, the powders move against the centrifugal direction. The gradation profile of ceramic powders in the FGM can be controlled by the difference in the relative density between molten metal and ceramic powders and the intensity of centrifugal force [11].

(3) Solid processes

Powder configuration, self-propagating high temperature synthesis (SHS) and spark plasma sintering (SPS) are included.

(a) Powder configuration process

The powder configuration process is used to fill the mixtures of metal and ceramic powders in a die as a preceding process of sintering. Sintering is achieved by hot pressing, hot isostatic pressing (HIP) and cold isostatic pressing (CIP). FGMs can be obtained by changing the mixing ratio of metal and ceramic powders in the depth direction of the die. When the mixing ratio is changed step by step or continuously, the compositional gradient of the FGM is stepwise or continuous, respectively. The imbalance of sintering conditions between metal and ceramic powders often causes the deformation of the FGM, crack initiation, and delamination. A graded temperature according to the compositional gradient in the die should be applied during the sintering process.
(b) SHS process

SHS is a process for forming compounds of boride, carbide, or nitride from the mixtures of metal powders and boron powders, nitrogen gas, or carbon powders by their exothermic formation reaction. The formation energy of boride, carbide, and nitride is >40 kcal/mol. When one part of the mixture is ignited, the compound formation starts. The formation reaction heats the neighboring mixtures and then self-propagates in the mixtures. Because the compounds obtained by SHS are generally powderlike or porous, HIP or CIP is sometimes applied to the mixtures during the SHS process. The FGM can be obtained by preparing mixtures with a compositional gradient using the powder configuration process [12].

(c) SPS process

SPS is a process for forming the compacts in a short time by passing a big pulsed current through the powders filled in the die. The FGM can be obtained from the mixtures of metal and ceramic powders with a compositional gradient by the powder configuration process. SHS is often used for the fabrication of FGMs, because the sintering time is short and the applied electric power is lower than for other sintering processes. The SPS system to fabricate FGMs with a size of 100 × 100 × 40 – 60 mm continuously has been developed [13].

4. TYPICAL APPLICATIONS OF FGMs

(1) Relaxation of thermal stress

Glass is generally difficult to join with a metal, since both their thermal expansion coefficients are so different that cracks are initiated at the joined interface between glass and metal after cooling. The values of the thermal expansion coefficients are 3.2 × 10⁻⁶/K for Pyrex glass and ≥10 × 10⁻⁶/K for metals. Kovar metal is an Fe–29Ni–17Co alloy, which is characterized by a low thermal expansion coefficient of approximately 5 × 10⁻⁶/K at room temperature. Kovar metal is used as an intermediate layer between glass and metal during the welding process. Thermal stress does not concentrate at the interface but distributes in the intermediate layer to prevent crack initiation after cooling down. This is a graded structure of the thermal expansion coefficient for the relaxation of thermal stress.

(2) Control of internal and external stress

The WC–Co hard metals, TiCN–Ni cermet, and hard metals coated with Ti series ceramics of <10-μm thickness have been practically applied to cutting tools. The characteristics required for cutting tools are abrasion resistance, defect resistance, and low cost. Hard metals are superior in defect resistance, while soft metals are inferior in abrasion resistance. On the contrary, cermet is superior in abrasion resistance. The hard metals coated with ceramics possess both excellent abrasion resistance and defect resistance, but the biggest problem is high price since an additional process to coat ceramics on hard metals is necessary.

The surface reaction process has been developed to fabricate cutting tools with graded surface coatings at the usual price. The surface structure of the tools is shown in Figure 11 [8]. High compressive internal stress is induced at the surface coatings by the diffusion of gas atoms from
A surface structure of TiCN/WC-Co-Ni FGM

The surface structure of TiCN/WC–Co–Ni FGM.

Figure 11

Deposition of the coatings on the surface. This prevents the coatings from cracking against the tensile stress loaded in the cutting process. The diffusion process forms a graded structure near the surface of the tools. The graded structure works well not to concentrate the internal compressive stress and external tensile stress at the interface between the coatings and hard metals. It makes the life longer by more than twice that of the ordinary hard metals coated with ceramics. High-performance cutting tools have been realized by the graded structure.

(3) Improvement of energy conversion

Thermoelectric energy conversion is a solid-state conversion between heat and electricity. It is an important technique to recover waste heat and supply electric power. Some oxides that are thermoelectric materials of Na–Co–O, Ca–Co–O, etc. have been researched on and developed mainly in Japan [14,15]. They are characterized by common element systems, light element systems, and heat resistance. Figure 12 shows the temperature dependence of the dimensionless thermoelectric figure-of-merit ZT for thermoelectric materials [16]. Here, T is the absolute temperature and Z is a thermoelectric figure-of-merit, which implies energy conversion efficiency when a temperature difference of 1 K is applied to a material.

As shown in Figure 12, all thermoelectric materials have specific temperature ranges to present high ZT. For example, Ca₃Co₄O₉ has a high ZT in the temperature range of >800 K. When thermoelectric energy conversion is applied to the temperature range between room temperature and 1000 K, a single thermoelectric material is not suitable to cover such a wide temperature range. A combination of Bi₂Te₃ and Ca₃Co₄O₉ is possible to cover the temperature range between room temperature and 1000 K. A new system of this combination has been developed to recover the waste heat from the steam generator. National Aeronautics and Space Administration (NASA) has developed a new power-generating device with a combination of Bi–Te and Co–Sb systems shown in Figure 13. This is to be used as a power device for the exploration of Mars. These are graded structures of thermoelectric figure-of-merit according to temperature distribution.

(4) Biocompatibility

Dental implants are a root to fix the artificial teeth. Figure 14 shows a schematic view of a dental implant, which is mounted directly in the jaw [17]. Sintered hydroxyapatite (HAP) and titanium have been used as constituent materials of dental implants. Since HAP is the main component of bones and teeth, it is superior in its affinity for bones, but the disadvantage is that it is easy to break because of low toughness. HAP-coated Ti implants were developed subsequently to overcome the low toughness of HAP, but another problem of separation of HAP coatings from Ti implants occurred.

A compositional gradient from Ti implants to HAP has been designed to resolve the separation problem. Figure 15 shows two types of compositional gradients in dental implants; one is a radial gradient and the other is the longitudinal gradient. In the case of the radial gradient, the...
The merit is that HAP is difficult to separate from Ti implants, but the demerit is that the implant easily becomes loose after HAP is taken into the bone. On the other hand, the longitudinal gradient is useful for fixing the implant at early stages of curing as well as the radial gradient. If the upper Ti part of the implant is fixed by the surrounding bone before HAP is taken into the bone, the implant does not become loose. The longitudinal gradient may be a promising structure, but more research and development are required.

5. FUTURE OF FGMs

When we observe the graded structures in nature, they are highly developed structures to adapt to the changes in the environment. The graded structures may be one of the final forms of natural materials. Is it possible that FGMs are one of the final forms of artificial materials?

FGMs are characterized by correspondence to the steep gradient of stress, temperature, and compatibility mentioned above. We can say that one of the keywords of FGMs is “correspondence to the steep gradient” [18]. We have many global problems such as those related to energy, environment, population, food, economy. Demands on materials have become advanced, multifunctional, and ultimate to resolve such problems. For example, the high energy conversion efficiency of heat power plants is essential for energy and environmental problems. Since the energy conversion efficiency increases with the operation temperature, ultrahigh heat-resistant, oxidation resistant, and adiabatic materials under a steep gradient of temperature are needed. Monolithic materials are almost impossible to answer these multidemands. FGMs with multifunctions will be one of the major candidates to “correspond to the steep gradient” of temperature.

Research and development of materials is actively being pursued, as a source of economic growth. Recently, the shift has been from monolithic materials to combined materials like composite materials. A new Boeing 787 airplane consists of many light parts of composite materials to reduce the weight and to improve the fuel efficiency. Nowadays, artificial materials are in the evolutionary process. When the demand for materials is more advanced and multifunctional in the near future, FGMs will get a better chance in our practical applications.

REFERENCES


1. LIFESTYLES AND TECHNOLOGIES BASED ON ENVIRONMENTAL ISSUES [1]

How do global environmental issues affect the human race? Climate change is not the only issue. What is more important is that we have given in to our insatiable material desires, and all of our modern technological advances have been created with these desires in mind. This has resulted in magnifying the impacts of human activity while depleting natural resources, degrading biodiversity, and increasing pressures caused by population, food shortages, and water shortages. If we continue on this path, no one knows where the breaking point will be. We do know, however, that all these issues are interrelated and have the potential to affect the balance of life.

The key question surrounding today’s environmental issues is whether or not we can stop this expansion of human activity and downshift our lifestyles. The answer lies in the new technologies that are most needed now.

If such a downshift is indeed possible, how much time do we have to create the new technologies that are needed? Considering the eminent depletion of natural resources, degraded biodiversity, and global warming, the situation is unfortunately expected to reach a critical stage by 2030.

One key to creating this new concept of technology can be found in the very natural systems that have been a part of nature’s perfect closed loop for the over 4.6 billion years of the earth’s history. Nature teaches us two things: one is that the natural, closed-loop system runs on minimal energy. Gecko lizards can run along the ceiling at full speed, leaves on trees change colors and fall in autumn, and maggots thrive in manure. Everything is supported by very simple, yet amazing mechanisms.

Why then, is there such a discrepancy between modern civilization and the natural system? The answer lies in the natural selection that has always been present in the insects’ world if their needs were not sufficiently met. In other words, although individual members of the natural system act for their own good, the natural system as a whole is altruistic. By contrast, human beings have used modern technology to avoid natural selection and have overconsumed natural resources. If we continue with our current way of life, the first natural selection in the human community may happen around 2030. This could potentially lead to the end of civilization.

Current fossil fuel-reliant technology is supported by the perception that nature is a tool for human use. The great philosopher, René Descartes, introduced the concept of nature expressed as a formula, and that it can be enslaved and freely exploited for our own benefits. This principle spread from Europe to every country in the world, becoming one of the fundamentals of modern civilization and technology. Certainly, this is not the fault of Descartes, but he and Francis Bacon explicitly showed a way to live lives separate from natural systems. This lack of recognition of the natural system allowed modern technology to get out of control.
On the other hand, we have an insatiable desire to gain added value from life. This desire is probably unique to human beings, and we find it difficult to adjust to the loss of comforts and conveniences that we have already obtained. It is this desire for added value, coupled with a lack of understanding of the natural system that drives the development of modern technology and encourages people to consume more in the name of convenience and efficiency.

Now, we need to develop a new method of manufacturing and lifestyle that can support both this desire for added value and a recyclable society. Contradiction between the two must be mitigated by transitioning from technologies that encourage consumption to technologies that encourage spiritual growth and respect for the natural system. We call such technology, based on nature’s perfect closed loop, “Nature Technology.”

2. A SYSTEM TO CREATE NATURE TECHNOLOGY [1]

In order to give form to Nature Technology, it is important to use the backcasting method. The three steps needed for its creation are listed below (Figure 1).

1. Imagine and draw a new, happy, excited, and spiritually satisfied lifestyle under the environmental constraints in 2030.

   First, it is necessary to clarify what is meant by irresistible convenience. There will be some irresistible conveniences that cannot be given up regardless of the environmental constraints. It is important to be clear as to why they are irresistible even if the reasons seem trivial.

   Second, it is necessary to clarify how much impact it would have on the environment if we continue enjoying the irresistible convenience available through the current technologies. (At this point, it would be helpful to critically discuss if they are truly irresistible conveniences.)

   Third, imagine a lifestyle that allows for irresistible conveniences but does not have large environmental impacts. One example is a marketing phrase used by an auto maker, “Cars. The more you drive, the cleaner the air will be.” This phrase was submitted in one of the children’s workshops at the 2005 World EXPO in Aichi, Japan. It is important to write down many imaginative ideas, forgetting about feasibility and common sense. It tends to work better to invite children rather than engineers for discussion at this stage, because the most imaginative ideas seem to appear impossible to engineers who are grounded in the resource-oriented civilization and wisdom.

2. Extract necessary elemental technologies from the design and find matching technologies from nature that

   First, it is necessary to clarify what is meant by irresistible convenience. There will be some irresistible conveniences that cannot be given up regardless of the environmental constraints. It is important to be clear as to why they are irresistible even if the reasons seem trivial.

   Second, it is necessary to clarify how much impact it would have on the environment if we continue enjoying the irresistible convenience available through the current technologies. (At this point, it would be helpful to critically discuss if they are truly irresistible conveniences.)

   Third, imagine a lifestyle that allows for irresistible conveniences but does not have large environmental impacts. One example is a marketing phrase used by an auto maker, “Cars. The more you drive, the cleaner the air will be.” This phrase was submitted in one of the children’s workshops at the 2005 World EXPO in Aichi, Japan. It is important to write down many imaginative ideas, forgetting about feasibility and common sense. It tends to work better to invite children rather than engineers for discussion at this stage, because the most imaginative ideas seem to appear impossible to engineers who are grounded in the resource-oriented civilization and wisdom.

   (2) Extract necessary elemental technologies from the design and find matching technologies from nature that

   First, it is necessary to clarify what is meant by irresistible convenience. There will be some irresistible conveniences that cannot be given up regardless of the environmental constraints. It is important to be clear as to why they are irresistible even if the reasons seem trivial.

   Second, it is necessary to clarify how much impact it would have on the environment if we continue enjoying the irresistible convenience available through the current technologies. (At this point, it would be helpful to critically discuss if they are truly irresistible conveniences.)

   Third, imagine a lifestyle that allows for irresistible conveniences but does not have large environmental impacts. One example is a marketing phrase used by an auto maker, “Cars. The more you drive, the cleaner the air will be.” This phrase was submitted in one of the children’s workshops at the 2005 World EXPO in Aichi, Japan. It is important to write down many imaginative ideas, forgetting about feasibility and common sense. It tends to work better to invite children rather than engineers for discussion at this stage, because the most imaginative ideas seem to appear impossible to engineers who are grounded in the resource-oriented civilization and wisdom.

   (2) Extract necessary elemental technologies from the design and find matching technologies from nature that

   First, it is necessary to clarify what is meant by irresistible convenience. There will be some irresistible conveniences that cannot be given up regardless of the environmental constraints. It is important to be clear as to why they are irresistible even if the reasons seem trivial.

   Second, it is necessary to clarify how much impact it would have on the environment if we continue enjoying the irresistible convenience available through the current technologies. (At this point, it would be helpful to critically discuss if they are truly irresistible conveniences.)

   Third, imagine a lifestyle that allows for irresistible conveniences but does not have large environmental impacts. One example is a marketing phrase used by an auto maker, “Cars. The more you drive, the cleaner the air will be.” This phrase was submitted in one of the children’s workshops at the 2005 World EXPO in Aichi, Japan. It is important to write down many imaginative ideas, forgetting about feasibility and common sense. It tends to work better to invite children rather than engineers for discussion at this stage, because the most imaginative ideas seem to appear impossible to engineers who are grounded in the resource-oriented civilization and wisdom.
have very low environmental impacts and are highly functional.

Engineers are welcome to join at this point because they are the best at creating technologies from imaginative design ideas, putting them into practice, and making a clear outline. There will be cases where it is important to define one function as a collective entity of several technological elements (an irrigation system which can pump water up to 100 m, or adhesive glues which can stick to any kind of surface) or where we can define a function with one technology. Depending on how to interpret the design, we may need to define one function with several technologies. In some cases, it may not be possible to identify a technology at all because the idea is too imaginative.

Next, is by using the outline that has been created to find a matching technology from nature. It is more effective to use a database instead of trying to find these technologies individually. My research team has collected about 500 cases of Nature Technologies, and is now in the process of organizing them. Some of them are disclosed to public (see Showroom for Amazing Nature at http://nature-sr.com). The database is not yet complete and fully functional, as it is currently being adapted for general purposes. Related information is disclosed by The Biomimicry Institute (see http://www.biomimicryinstitute.org). As for paper publications, there is Power of Insects by Manabu Akaike, and Mosquitoes Cure Brain Infarction! Amazing Power of Insects by Takayuki Nagashima, The Gecko’s Foot: Bio-inspiration: Engineering New Materials from Nature by Peter Forbes, and 100’s Best (Diamond magazine, December 2008) by Janine M. Benyus, advocate of biomimicry, and Gunter Pauli, director of Zero Emissions Research and Initiatives (ZERI). We are currently exploring the possibility of an alliance with them.

(3) Redesign a technology from similar elements found in nature instead of simply mimicking the amazing mechanisms of nature.

Once a promising technology is found, check how many of the elements of Nature Technology are met. Will the technology stop material wants? Can it reduce the environmental impacts? Most of all, is it clear and simple enough to support people, allowing them to enjoy it and develop a sense of attachment? This process is repeated to narrow down the field to the best technologies.

Having verified the technology, there may still be a need to redesign it to meet the requirements, and there is a possibility that some technologies will be eliminated at this stage because they cannot meet the low-environmental-impact threshold. It may be suprising to know how perfect the natural system is, but this is the final hurdle all technologies must pass, and if they cannot pass it, it is necessary to narrow them down once again. Only technologies that meet all the requirements can be considered Nature Technology.

Because creating Nature Technology is not simple biomimicry, there are many difficulties and there are still not many success stories. This is why the database must be taken to a higher level, so that anybody can make use of it and link to the third natural selection. It should be noted that since the focus is on convenience for the general public, it would be a helpful solution for issues that are involved with various stakeholders, such as social infrastructure. Despite its weaknesses, the current system continues to be improved in order to make a contribution to solving the most urgent issue: how to increase public awareness about the third selection.

3. SOME PRESENT EXAMPLES [1]

3.1. Living in Earthen Buildings—Air Conditioning Without Electricity

Situated in a monsoon region, Japan experiences a wide variation in temperature and humidity, from exceptionally cold and dry winters to hot, humid summers. The climate in Europe, on the other hand, tends toward low temperatures and high humidity during the winter season. In Europe, simply turning on the heater raises the temperature and lowers humidity, creating an ideal climate for people—16°–26 °C with around 40–70% humidity. In Japan, however, using an air conditioner in the summer actually increases humidity, contributing to an in-house climate that is less than ideal (Figure 2).

Traditionally, the houses in Japan tend to have higher roofs and are constructed with very basic materials: wood, paper, and earth. The main concern behind this building style was to create the ideal climate in the humid summer months. After World War II, however, the Western concept of insulated houses was more widely adopted and, in 1973, the oil shock prompted a national policy that helped highly insulated home construction to gain a foothold. This western style of home manufacturing is effective for temperature control, but humidity control is extremely difficult.

This new style of home building in a climate like Japan’s, where humidity control is of higher importance, created a need for electric fans, humidifiers, and dehumidifiers and similar appliances that were introduced one after another. While these are generally considered to increase the overall quality of life, the truth is that the use of volatile organic compounds (VOCs) contained in many of these new devices led to a noticeable increase in Sick House Syndrome and cases of allergy. This then led to a need for even more appliances for air purification. This was just the start of a vicious feedback cycle that resulted in a 229% increase in-household energy consumption since
1973. While it should be noted that business energy usage remained the same, this increase marked a doubling in consumer consumption compared with the time of the oil shock.

No one can dispute the fact that there are many new appliances that are more efficient than older models, but given our current lifestyle, along with the increasing number of households, overall energy usage is increasing.

3.1.1. Learning from Termites

Let us look at this situation from a Nature Technology perspective. We first assume that humankind will not willingly give up modern-day comforts now or in the year 2030, and that making a transition from modern insulated homes to the more traditional high-roof style houses by 2030 is impossible. Even taking into consideration the environmental, energy, and resource state of the world in 2030, it would be difficult to do away with air conditioners and other convenience appliances. This means that we must come up with alternatives to provide the same functionality (Figure 3).

Getting back to the basics, we find that the floor, the walls, and the ceiling could monitor and even control a room’s temperature and humidity. Ideally, without electricity, production of the materials would require as little energy as possible.

This begs the question: are there no examples to be found in nature? Actually, if we take a quick peek through nature’s door, we find the termite’s nest. Termites living in Savanna regions spend decades building enormous nests that can reach heights of up to 6 or 7 m. While temperatures

FIGURE 2 World climate. For color version of this figure, the reader is referred to the online version of this book.

FIGURE 3 Comfortable living without environmental load; control of indoor environment by floor, wall, and ceiling. For color version of this figure, the reader is referred to the online version of this book.
outside the nest reach daytime highs of 50 °C and nighttime lows of 0 °C, the inside temperatures remain at a constant 30 °C.

We do not yet fully understand exactly how this is possible, but we do know that the nest itself is built and functions like a chimney to facilitate ventilation. The main effect is undoubtedly related to the use of moist soil taken from deep underground (Figure 4).

Taking a closer look at the humidity regulation in a well-insulated, energy efficient home, we find that the ideal humidity for human comfort is generally considered to be between 40% and 70%. Excess humidity is absorbed, and dry air is humidified.

A process called capillary condensation makes it possible to open the microscopic holes in a material, and thus maintain an ideal humidity level. If the holes are not within 4–8 nm in size (1 nm is 1 billionth of a meter), the process breaks down, but with enough holes of the right size, it should be possible to very effectively regulate indoor humidity.

The question then, is how to make such a material with so many microscopic holes? There are solutions based on using petroleum resources, but these require enormous amounts of energy. Soil, however, has, by nature, minute holes, cracks, and crevasses naturally. Earthen walls have long been utilized for this amazing feature that also assists the natural fermentation process required to make miso, soy sauce, and sake.

3.1.2. The Greatness of Soil

Soil has existed on earth for 400 million years. Before that, there was no soil, and without soil, no greenery, and no perfect natural cycle that we observe today. It is a shame that the role of soil and dirt is so often overlooked. Even as we readily declare our love for green, very few say the same of dirt. Not only that, but we use such negatively charged phrases as “I got dirt on me,” or “covered in mud.”

In fact, soil is an indispensable part of our lives and lifestyles. A normal sheet of western paper is 30% soil, and glossy paper used for artistic purposes has up to 40% soil and clay content. (This is one reason that traditional Japanese paper is relatively lighter than western paper.) It is not only paper either: a 6H pencil contains 55% soil, average makeup and lipstick is 15% soil, and foundation alone is between 40% and 70% earth and clay. Earth is required in everything from medicines to the plastic used in car bumpers.

It makes sense then, to take advantage of soil’s unique humidity-regulating properties. Unfortunately, it cannot be simply incorporated into modern airtight, highly insulated, temperature-controlled houses. We must also take into consideration the fact that, in order to be useful, the earth must be compacted and hardened, reducing the risk of dust-induced allergies. Using the soil as a raw material for kilned ceramics, however, causes the tiny holes to close, but if we knock on nature’s door for advice, we find unique rock formations in hot-spring areas that are formed by a reaction occurring in water at temperatures <200 °C.

Taking advantage of the natural properties of the soil, it is possible to add a small amount of lime to the earthen mixture and bake it for 3–4 h at 150 °C (a process called hydrothermal solidification, carried out using an autoclave). This solidifies the soil, processed at such a low temperature, energy consumption is reduced to only 1/6 of that used to kiln ceramic tiles. Ceramic’s value as a low-impact material is clear when we consider that plastics require 3 times the energy used for ceramics. Iron requires 10 times, and aluminum 17 times the energy used to produce ceramic tiles. What’s more, any kind of soil contains these microscopic holes and, being freely available, there are very few limitations to its use.
This material is the first of its kind to be successfully developed through hydrothermal solidification. It is easy to make, and it requires minimal energy. Why, then, has this not been discovered until now? Perhaps it is because we were looking in the wrong places. Nature Technology, however, has the benefit of using nature as a guide.

3.1.3. Bringing Soil into Daily Life

What would it be like to actually use such materials in our daily lives? This photograph illustrates a real-life application of materials that were developed for the floor and walls of my insulated apartment in Handa City, Aichi Prefecture. Compared with the previous carpeting, there is a noticeable difference in the temperature and humidity, and household yearly energy usage has dropped to a dramatic 20% (Figure 5).

There were also some unexpected results. The small holes are ideal for absorbing odors, making it possible to cook fish not only in the kitchen but in the living room as well, so much so that even boiling fish leaves no odors the next day. Quality time with my family has increased as well, as we enjoy experimenting with drying various foods.

We expected the ceramic walls to provide up to 80% of our air-conditioning needs and anticipated using conventional air conditioning to provide the remaining 20%. In reality, however, the soil walls gave us the opportunity to brainstorm as a family, to develop novel ways to cool the room with less energy, such as sprinkling water on the walls and floor.

Remember, this material is all natural and takes environmental impacts into account through its entire life cycle, from production to usage. Making this change has led to more time as a family and deeper conversations about how we can live a more sustainable, higher quality life.

Using clay and earthen building materials with a sprinkling of water is a concept simple enough for anyone to grasp and promotes human interaction—not only between immediate family members but houseguests have also begun staying longer. This fostering of relationships is precisely the point of Nature Technology.

Clay and earthen materials, such as Inax’s Eco-Karat and Soil ceramics, are already available to consumers, as are other useful materials that utilize the same basic concept. Clay has a cohesive structure and contains many nanometer-sized holes that, when used to their full potential, regulate not only humidity, but also VOCs and odors.

3.1.4. Seeing the Obvious

I am lucky enough to own a small piece of jungle in Okinoerabu Island, in the Amamisho Islands. Since my first visit in 1996, I have made it a point to stop by several times a year, and through these regular visits, I have developed a deep interest in the island. Houses have names in many rural Japanese areas, and the name of our house mentioned above was “Sui-An I,” which means “Drinking Oasis #1.”

FIGURE 5  Hydrothermally solidified soil used for the floor in the kitchen and living room of the building caused suppression of changes in the temperature and humidity, leading to the reduction of 17% of the energy consumption. For color version of this figure, the reader is referred to the online version of this book.
I once thought that what drew me to this island was its natural beauty and brown-sugar liquor, but as it turns out, there was more to it than that (Figure 6).

The people of these islands know how to enjoy life, they know how much is enough to be satisfied, and they don’t accept defeat. I have found that I am just as captivated by this aspect of the island, and this prompted me to acquire a small piece of jungle in 2004, where I built a small house and named it “Sui-An II.” Sadly, the majority of new homes built on the island follow the modern, fully insulated Western model. This is despite the fact that, even with the direct sunlight on the south side of the island, the sea breeze blowing through the shade of the trees keeps the temperature below 31°C; on the other hand, we find it difficult not to make use of the noon breeze blowing in from the sea and evening breeze coming down from the mountains when the weather is good. These breezes can be funneled through the house, through any room with stucco walls and where both the floors and the verandah have been created with the hydrothermally solidified products. I have no objective statistics, but even without air conditioning, we are in complete comfort. This leads me to believe that perhaps another benefit of Nature Technology is that it helps us to recognize the obvious.

3.2. Learning From a Snail—A Natural Nonstick Surface

One virtue valued by Japanese people is cleanliness. Cleanliness has come to be seen as a convenience that we cannot be without, and the only way to achieve cleanliness is by cleaning. Unfortunately, even cleaning can have a significant environmental impact. While there are environmentally friendly cleaners available for such things as building exteriors, the sad truth is that strong detergents are used more often than not. Reducing this impact would require considerable amounts of water.

Household cleaning products raise similar dilemmas. Soap has little effect on aquatic life, but when looked at in terms of resource usage, soaps require much more fat than that of synthetic soap. Synthetic soap poses a problem for the environment because the pollutants that are slightly soluble in water are readily decomposed. Even vegetable-based soaps are largely made from palm oil, the production of which contributes to deforestation.

Might there be examples found in nature of surfaces that do not collect dirt or surfaces that can be easily cleaned with water alone?

3.2.1. The Nonstick Snail Shell

Once again, let us knock on nature’s door. One of the more obvious examples of perpetually clean surfaces found in nature is a cockroach. Cockroaches, stag beetles, and even dung beetles are always shiny and clean. Unfortunately, they keep themselves clean with secretions from their own body, which is not ideal for our purposes. The next example we come across are the eggs and shells of snails. Snails do not excrete any special cleaning agents, and yet, they always have a shiny clean surface. This is because the surface of a snail’s shell wicks away dirt, grime, and even rain. The shell, made of calcium carbonate, can be compared with calcite rocks, which are also composed entirely of calcium carbonate. In one experiment, we dipped both the calcite and the snail shells in water and applied a drop of oil. The oil stuck to the calcite but not to the snail shell. In fact, when in the water, the oil droplet rolled down and around the shell of the snail. We naturally wondered why (Figure 7).

Upon further inspection of the shell with an electron microscope, we found what can best be described as...
microscopic mountains ranging from a millimeter to as small as a few nanometers. These miniature mountains are the reason that oil and grime do not adhere to the shell. The natural conclusion is to simply reproduce the texture of the snail shells, but this would require tremendous amounts of energy and resources, which is why the idea for a redesign of the technologies is requested from the point of the environment.

3.2.2. Perfect Cleaning with Water Alone

What is the relation between the texture of a snail’s shell and cleanliness? The main connection is related to the microscopic mountains and valleys on the surface of the shell that actually converted the surface energy. If the difference of surface energy of dirt and material is greater than the surface energy of material and water, the shell would become dirty in air. When water is applied, however, the water penetrates between the surface and the dirt, pulling the dirt away. If that surface energy can be controlled, it would be possible to achieve a surface that is resistant to dirt and grime, and thus easy to clean. The biggest challenge facing us is to achieve this with the smallest environmental impact possible.

This kind of nonstick technologies is already being used in products ranging from kitchen sinks to building exteriors. Surfaces of buildings and others of the city are mostly dirtied by oil and carbon emissions from cars and trucks, and if we recognize this “kinds of dirtiness,” it will open our minds to innovate new materials for these surfaces using redesigned snail shell surfaces. When we apply artificial dirt to ceramic tiles that have been remade with this new nonstick technology in mind, we find that the mixture of carbon and oil can be simply rinsed away with water. The surface energy of ceramics is very close to that of a snail’s shell meaning that with a slight adjustment and minimal impact to the earth, a new nonstick ceramic can be created. If we apply the same knowledge in other areas, we could develop kitchen sinks and building exteriors that stay as clean as a snail’s shell (Figure 8).

The two examples of Nature Technology mentioned here are already being evaluated in the marketplace. Nature Technology is about a shift in our value system from a focus on things to a focus on functionality. This shift of focus can lead to technologies that make use of solar power and the gifts of nature, providing an exciting lifestyle of abundance with minimal impact on the environment.

4. CREATION OF THE FUNCTIONAL MATERIALS BY HYDROTHERMAL REACTION WHICH SUPPORT THE EARTH CIRCULATION

Based on these concepts, we are now challenging to create new materials.

The amount of CO2 discharge has increased during the last decade, leading to global warming. For the construction of sustainable society, it is necessary to reduce the environmental load. The main proposition is to reduce both input, such as the amount of resource and energy, and output, such as waste in the materials manufacturing. Recently, recycling and reuse processes for the wastes in the view of output reduction have attracted the attention of many researchers as a way to protect the environment [2,3]. In nature, there are many resources not utilized as starting materials due to their relatively low purity. The preparation of functional materials using these materials would cause reduction of the input in the materials manufacturing, leading to contribution of reducing the environmental load.
Some mesoporous materials such as Wakkanai diatomaceous shale have been found to adsorb water vapor by the principle of capillary condensation and are used as ingredients in commercially available, humidity-controlling building materials [4]. Humidity-controlling materials can reduce energy consumption by 5–30%, resulting in the reduction of the energy in use [5]. In general, the relative humidity range in which people feel comfortable and water vapor does not condensate is 40–90%. Pores with diameters of 3.2 nm and 22.1 nm would yield a relative humidity of 40% and 90%, respectively, by prediction of the calculations using zero of a contact angle of a material based on the Kelvin equation of capillary condensation and Arai’s correction equation. That is, the introduction of mesopores with appropriate diameters and a broad pore size distribution into the materials enables humidity control.

The energy required to synthesize ceramics by hydrothermal treatment is even lower, about 2.7 GJ/m³, which is only one-sixth that of the energy needed to synthesize ceramics by sintering [6]. That is, a hydrothermal treatment as the material process is suitable for reduction of environmental load. Our idea for material design with low energy consumption is to prepare mesoporous materials using the natural resource in a sustainable society is to prepare mesoporous materials using the natural resource in a hydrothermal condition. Research has been conducted on the hydrothermal reactions and solidification behavior of typical calcium silicate hydrate materials, for example, tobermorite and C–S–H gel, in a CaO–SiO₂–H₂O system [7–9]. The precipitate of typical calcium silicate hydrates such as C–S–H gel and tobermorite in the hydrothermally solidified materials has been suggested to cause the formation of fine pores of a diameter of 10 nm, mesopores in a CaO–SiO₂–H₂O system [10,11].

We had paid attention to zeolite and allophane as the natural resources. These materials have a unique microporous structure, and hence, their potential for use in environmental applications. Zeolite was reported to have a higher ammonia gas absorption capacity than alumina and silica gel [12]. Allophane has been studied on the basis of its ion exchange, ion retention [13,14]. However, the microporous property of the materials causes intracrystalline diffusion limitation. Both micropores and mesopores are proposed to be newly formed in the materials to improve mass transport properties of the materials [15]. The solidification of zeolite/allophane with mesopores/macropores has played an important role in broadening the application of zeolite/allophane in various fields. However, the solidification of the materials sintered at high temperatures using zeolite and allophane due to phase transformation and low sintering property [16,17] has not been reported.

Our strategy for zeolite/allophane solidification in addition to mesopores is to maintain the microporous structure of zeolite/allophane and to form mesopores originated from the formation of newly deposits during hydrothermal reaction. Hydrogarnet has been reported to be formed as the major phase when a CaO–SiO₂–Al₂O₃–H₂O system with an Al/(Al + Si) ratio >0.12 is subjected to hydrothermal reaction. The formation of hydrogarnet decreases the strength of the hydrothermally treated materials [18,19]. To overcome this problem is important for the preparation of zeolite/allophane solidified hydrothermally with mesopores.

4.1. Solidification of Zeolite with Micropores/Mesopores by Hydrothermal Reaction

To prevent the formation of hydrogarnet in a hydrothermal reaction using zeolite, we considered mordenite, which is known as zeolite, with a high silica content. Natural zeolite minerals containing a large amount of mordenite with traces of quartz and feldspar were used as the mordenite material. The materials, mordenite and slaked lime were used as starting materials, were mixed, and the mixture was then added to 10 mass% of distilled water for molding. The mass ratios of mordenite/slaked lime in the mixture were 90:10. The mixture was then uni-axially pressed at 15 MPa in a stainless-steel die. The powder compacts were hydrothermally treated under saturated steam pressure at 180 °C for 12 h. The time of the hydrothermal reaction was determined by an optimization process based on a trial-and-error approach to achieve satisfactory results in terms of adequate strength of the samples for solidification.

XRD result indicated the dissolution of slaked lime as the starting material. On the other hand, no peaks corresponding to the newly formed deposits were seen in the XRD pattern of the sample. Numerous plate- or needle-like fine deposits, which are filled in the space between the particles of the starting materials, were observed in the scanning electron microscopy (SEM) photograph of the specimen as shown in Figure 9.

![SEM micrograph of the fracture surface of the specimen.](image)
It is well known that crystalline calcium silicate hydrates such as tobermorite form via C₅S₆H gel as a precursor in the hydrothermal reaction [20]. The deposits formed on the sample are a precursor gel such as C₅S₆H gel. Sample with 41% of the porosity had the bending strength 6.5/6.0.8 MPa. Filling the macropores in the space between the particles in the powder compacts because of the formation of calcium silicate hydrates played an important role in improving the strength of the samples after the hydrothermal process [21]. The strength of the sample seems to depend on their porosity and the filling in their macropores because of the formation of calcium silicate hydrates.

To investigate the mesopore structure of the sample after the hydrothermal reaction, the mesopore size distribution, as shown in Figure 10, was determined by the Barrett–Joyner–Halenda (BJH) method from the nitrogen gas desorption isotherms.

The pore size distribution curve of mordenite shows a gradual increase in the range of 5–20 nm. It is evident that the sample exhibits a broad mesopore size distribution of >3 nm and has a greater volume of each pore compared with that of mordenite. As shown in the SEM micrograph (Figure 9), the structure intertwining with each calcium silicate hydrate particle formed between the particles of the starting materials leads to the formation of mesopores. The porous properties of the sample after the reaction are shown in Table 1. The sample has a greater mesopore volume than mordenite. On the other hand, the specific surface area of sample A is lower than that of mordenite. After the reaction, the micropore volume of the samples decreased drastically compared with that of mordenite. The hydrothermal reaction leads to the dissolution at the surface of mordenite, resulting in partial collapse of the micropore channel network of mordenite.

The sample has higher water-vapor adsorption/desorption abilities because of the formation of mesopores as shown in Figure 11, as compared to mordenite, which shows type I behavior in the Brunauer–Deming–Deming–Teller classification. The specific surface area of the samples has little influence on the water-vapor adsorption/desorption abilities.

The amount of adsorbed water vapor and hysteresis loop depended on the mesopore volume of the sample. The amount of water vapor adsorbed on materials, <0.1 relative water-vapor pressure, depends on the micropore volume of the materials calculated from the Kelvin equation. That is, the decrease in micropore volume because of the

---

**TABLE 1** Porous Properties of Sample and Mordenite. The Micropore Volume was Calculated by the t-plot Method.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Specific surface area (m²/g)</th>
<th>Mesopore volume &lt; 20 nm (cm³/g)</th>
<th>Micropore volume (cm³/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>18</td>
<td>5.5 × 10⁻²</td>
<td>5.2 × 10⁻³</td>
</tr>
<tr>
<td>Mordenite</td>
<td>36</td>
<td>4.1 × 10⁻²</td>
<td>10.5 × 10⁻³</td>
</tr>
</tbody>
</table>
dissolution of mordenite causes reduction of the adsorbed amount of water vapor at relatively low water-vapor pressure. To examine the water-vapor adsorption—desorption abilities of the materials, the weight change of the materials was monitored. The materials were kept at 53% humidity and 25 °C in a 7-L of airtight chamber. The materials, excluding their top surfaces, were covered with aluminum tape so they would adsorb and desorb water vapor only through the top surface. After the weight stabilized at a constant value, the humidity was changed to 75%. This humidity was kept for 24 h, and then, the humidity was changed to 53% again for 24 h. The 53% and 75% relative humidity levels were controlled with Mg(NO₃)₂ 6H₂O—H₂O mixtures and NaCl—H₂O mixtures, respectively. Figure 12 shows the relationship between the weight change of the samples and measurement period at 25 °C at 53%—75% relative humidity.

In the case of the sample, the moisture weight increased logarithmically when the level of humidity was changed to 75%. On the other hand, the moisture weight rapidly decreased after 12 h and then stabilized when the humidity decreased. The moisture weight of the powder compact of mordenite dramatically increased after 8 h of humidity change, and then remained almost constant. The amount of moisture adsorbed and desorbed after 24 h of the humidity change depended on the mesopore volume. The moisture weight adsorbed on the samples was higher than that desorbed on the samples because of the hysteresis loop. Mordenite has a higher moisture adsorption/desorption rate during the initial period than the sample. The macropore volume in the range of 0.05—10 μm of the powder compact of mordenite and the sample was 0.32 and 0.22 cm³/g. The moisture adsorption/desorption rate on the samples during the initial period at 53% and 75% humidity had an influence on macropore volume. An increase in the macropore volume of the samples is proposed to accelerate the diffusion of moisture, leading to an increase in the adsorption/desorption rate during the initial period of the humidity change.

Figure 13 shows the ammonia gas concentration adsorbed on the sample, mordenite, and the commercially available tobermorite (surface area: 48 m²/g) for 4 h at 25 °C. Tobermorite showed the lowest gas adsorption capacity under this experimental condition.

In our preliminary experiments, tobermorite had much higher ammonia gas adsorption ability than C—S—H gel. The ammonia gas adsorption capacity of the sample was greater than that of mordenite, in spite of less micropore volume. The sample has higher harmful gas sorption abilities, compared with the prediction based on the theory of composite material. It implies that the interaction derived from mordenite combined with calcium silicate hydrates may influence the gas sorption property. Alkali treatment of zeolite using NaOH was reported to be a method to extract a siliceous species from the framework of the zeolite to obtain a lower SiO₂/Al₂O₃ molar ratio [22]. SiO₂/Al₂O₃ molar ratio of zeolite influences the strength of Brønsted acid and gas adsorption capacity. The sample was prepared under alkali conditions because of dissolution of slaked lime in a hydrothermal reaction. The SiO₂/Al₂O₃ molar ratio of mordenite in the samples would change during the reaction. Thus, the SiO₂/Al₂O₃ molar ratio of mordenite in the samples also influences the harmful gas adsorption capacities. The material design for the combination of
mordenite with calcium silicate hydrates in hydrothermally solidified materials is a great candidate for environmental materials with controlling properties of the indoor humidity utilizing the mesopores and removing the harmful gas.

4.2. Solidification of Allophane with Micropores/Mesopores by Hydrothermal Reaction

Allophane behaves as an amphoteric ion exchange material because of the presence of both H\(^+\) and OH\(^-\) ions on the surface of the particles [23]. It is supposed that allophane particles have an extremely high surface activity. The deposits formed between particles in powder compacts enhance the strength of hydrothermally solidified materials of a CaO–SiO\(_2\)–Al\(_2\)O\(_3\)–H\(_2\)O system [24]. The space associated with each fine-sized deposit was suggested to cause the formation of mesopores [25]. The formation of new deposits on its surface during hydrothermal reaction would play an important role in the solidification of allophane. The amount and the shape of the deposits in the solidified material are considered to influence the mesoporous structure of the materials. The temperature of the hydrothermal reaction influences the solubility of the materials. The hydrothermal reaction was carried out at various temperatures in a CaO–SiO\(_2\)–Al\(_2\)O\(_3\)–H\(_2\)O system to solidify materials derived from allophane.

Allophane isolated from the pumice bed in Kanuma, obtained from the Otsuka Collection (specific surface area: 746 m\(^2\)/g), and slaked lime were used as starting materials. The materials were mixed, and the mixture was then added to 40 mass% of distilled water. The Al/(Al + Si) and Ca/(Al + Si) molar ratios were 0.54 and 0.7, respectively. Six hours after the addition to water, the mixture was pressed uniaxially at 30 MPa in a stainless-steel die. The powder compacts were set in a Teflon-lined stainless-steel apparatus and then subjected to hydrothermal reaction under saturated steam pressure at 80 and 180 °C for 24 h. The materials prepared by hydrothermal solidification at 80 and 180 °C are named AL80 and AL180, respectively.

XRD results showed the formation of hydrogarnet after the hydrothermal reaction in the samples. Hydrogarnet forms due to the dissolution of allophane, which is the source of aluminum and silicon, and slaked lime, which is the source of calcium. The bending strength of the powder compacts was 1.8 ± 1.0 MPa before the hydrothermal reaction. AL80 (6.5 ± 0.1 MPa) showed a much higher bending strength than AL180 (1.1 ± 0.1 MPa). The strength of the materials prepared by hydrothermal solidification at 80 °C increased despite the formation of hydrogarnet.

Numerous nanoparticles with a diameter of 300 nm, which were considered as allophane aggregates, can be seen in the SEM micrograph before the reaction (Figure 14). After the hydrothermal reaction, no allophane aggregates are seen in any of the micrographs. Some fine deposits were observed in the SEM micrograph of AL80. The fracture surface of AL180 was completely covered with spherical deposits with a diameter of approximately 0.8 μm. It appears that the temperature of the hydrothermal reaction influences the morphology and amount of deposits on the sample.

Figure 15 shows the Fourier Transform Infrared (FT-IR) spectra of the samples and allophane. Allophane shows two absorption bands at around 1000 and 580 cm\(^{-1}\), which were due to the Si–O–Al and Al–O bonds in allophane, respectively [26]. The spectrum of AL80 was almost unchanged after the hydrothermal reaction at 80 °C, indicating that allophane remains in AL80 even after the reaction. In the spectrum of AL180, the absorption band at around 1000 cm\(^{-1}\) shifts toward a lower wavenumber. The band at around 1000 cm\(^{-1}\) was reported to shift toward a higher wavenumber with an increase in the Si/Al ratio [27]. Under alkaline conditions, the dissolution rate of Al is higher than that of Si [28]. The Si/Al molar ratio of the sample before the hydrothermal reaction was determined to be 2.12 by XRF. The Si/Al molar ratios of the samples were >2.20. It is proposed that the solubility of allophane increases during the hydrothermal reaction at high temperatures, thereby leading to an increase in the Si/Al ratio.

To investigate the allophane aggregates after the hydrothermal reaction, the mesoporous properties of the
samples were determined from nitrogen gas adsorption–desorption isotherms (Figure 16).

The pore size distribution curve of the samples before hydrothermal process showed a maximum at a diameter of around 4 nm. It has been reported that allophane is an aggregate of hollow spherical particles with diameters in the range of 3.5–5 nm [29]. The diameter of 4 nm is attributed to the space between the allophane aggregates in the powder compacts. It was clear that the peak shifts toward larger diameters after the hydrothermal reaction at 80 °C. The space between the allophane aggregates increases after hydrothermal reaction at 80 °C because of the dissolution of some of the allophane particles. New pores with diameters >10 nm were found in AL80. Numerous fine deposits were formed in the spaces between the starting materials in the powder compacts after hydrothermal reaction at 80 °C, as shown in the SEM micrograph (Figure 14). The space tangled with each fine deposit, which forms on the surface of allophane aggregates, caused the formation of mesopores. Almost all mesopores are destroyed in AL180. It is proposed that during hydrothermal process at a high temperature, the reactivity of allophane increases, because the aggregates are destroyed. Solidified materials hydrothermally with micropores due to the allophane structure and mesopores due to the aggregation structure and the spaces associated with each fine-sized hydrogarnet can be applicable for use in environmental applications.

REFERENCES


Recent Advances in HIP Technology and Atmosphere Control in HIP Treatment

Takao Fujikawa* and Yasuo Manabe†

*Metal Technology Co., Ltd. (Kinzoku-Giken, Co. Ltd.), 27th floor, Harmony-Tower, 1-32-2 Hon-cho, Nakano, Tokyo 164-8721, Japan,
†Machinery Division, Kobe Steel, Ltd., 2-3-1 Shinbama, Arai-cho, Takasago, Hyogo 676-8670, Japan

Chapter Outline

1. Introduction 1203
2. HIP Technology 1203
3. Conclusions and Future Perspectives 1210
References 1211

1. INTRODUCTION

Since the invention of HIP (hot isostatic pressing) technology in the US during the mid-1950s, it has made great progress as one of the important processes to produce pore-free materials [1]. The main application areas at that time were consolidation of encapsulated gas-atomized metal powders and the elimination of casting defects found in the investment cast super-alloy and titanium alloys. In the 1970s in Japan HIP was introduced to improve the yield of cemented carbide tools; however, it drew attention of ceramists as well. Because in the area of ceramics, which are usually manufactured by the sintering process from powder material, the residual pores were always a problem in order to fabricate dense and reliable components. Since then, HIP treatment where high pressure is applied at high temperatures has been recognized as an ideal process to realize full densification of sintered materials, such as ceramics. In this article, the history of HIP process in the last 15 years is reviewed focusing on equipment and one of the unique techniques of high-pressure gas atmosphere explained in detail.

2. HIP TECHNOLOGY

HIP treatment is carried out in a special furnace contained in a pressure vessel where a high-pressure gas charging unit is connected. Inside the pressure vessel, usually an inert gas such as argon of 100–200 MPa is filled and temperature is raised by the use of electric heaters. High-pressure gas has unique characteristics of high density and low viscosity; so a strong natural convection takes place inside the furnace and higher-temperature gas tends to flow up and stay around the top of the furnace or circulates if an inappropriate flow route exists. At the early period of the invention of HIP process, how to suppress this convection to realize temperature uniformity was a serious matter from the viewpoint of the HIP furnace design.

The configuration of the standard HIP unit is shown in Figure 1. Under high pressures, heated gas tends to gather in the top region of the furnace; thus, temperature distribution with high temperature at the top and low temperature at the low region is easily generated. In order to prevent such temperature distribution to occur, usually heaters are designed to control such temperature distribution by dividing heaters into several vertical zones to put heating energy to keep the furnace temperature uniform. Now the furnace design technology has been well established to obtain practically sufficient temperature uniformity and further progressed to facilitate the natural convection and even forced convection to improve cooling rate after temperature holding at the processing condition.

The issues of technological development for HIP treatment in the last 15 years were as follows: 1) the rapid cooling technology to shorten cycle time, 2) preheating technology to improve the productivity, and 3) atmosphere control to optimize the resultant mechanical, electronic properties and colors of the products. The rapid cooling technology and the preheating technology are discussed briefly in the following sections and the atmosphere control
treatment is explained in the next section with some of the experimental results.

2.1. Rapid Cooling Technology [2]

One of the disadvantages of HIP process has been its very long cycle time. Usually, HIP treatment takes from ten to thirty hours to complete a single cycle, depending on the size of the HIP unit used. From the time analysis of a single cycle, the cooling time from the processing temperature down to the near-room temperature to recover the work piece was recognized as too long and redundant. Thus, intensive efforts were made to shorten the cooling time and one of the methods to realize this was found to utilize the high-pressure gas convection inside the HIP furnace. In the early stage of the development, the utilization of natural convection by forming a circulation flow between workload space and the clearance between the thermal insulation cover just within the inner surface of the pressure vessel was explored.

The concept of this natural convection-cooling furnace is shown in Figure 2. At the holding stage, the valve on the top of the furnace is closed for the gas not to flow inside the furnace. After the holding at the temperature and pressure for a predetermined time, the valve for gas circulation is opened and the gas inside the furnace flows out from this valve to the outer region of the thermal insulation. This gas is cooled down by flowing along the inside surface of the pressure vessel and flows into the furnace from the bottom. This workload space and the furnace are cooled down faster than in the usual HIP equipment shown in Figure 1. The results were excellent, from the viewpoint of the cooling rate; however, it was found that the temperature distribution inside the workload space from the top to the bottom was too large, especially during the cooling procedure.

In order to overcome this problem, the utilization of forced convection by using electrically driven fan was introduced. The configuration of the rapid cooling HIP unit with the forced convection is schematically shown in Figure 3. The forced convection is generated by the fan below the pedestal and the heater to cause a strong upward gas flow inside the furnace. This fan is always driven even at the holding stage to realize the temperature uniformity. With this forced convection-type HIP furnace, short cooling time and good temperature uniformity during the cooling were
achieved at the same time. The configuration becomes a little complex and the space for the fan and a motor is necessary, as shown in Figure 3, so that this concept has been applied to considerably large HIP units with diameters larger than 600 mm, because the cooling time and total cycle time become longer when the HIP unit becomes larger. Now it is becoming much more common to use the forced convection with the HIP units for commercial production.

2.2. Preheating HIP Equipment [2]

As mentioned in the previous section, in the 1970s one of the main areas of HIP application was the consolidation of gas-atomized powder packed in metal cans. In this case, the processed material is not directly exposed to the air, so the heating to the HIP treatment temperature could be performed out of the HIP unit to reduce the occupation time of the expensive pressure vessel. This is the basic idea of the preheating HIP treatment to improve the productivity. Naturally, the workload is limited to the canned material such as high-speed tool steel billets or castings made of oxidation-resistant materials. Preheating temperature is optional to the material to be processed. Usually, bottom-loading-type HIP unit is used for this application, because in the case of the bottom-loading HIP unit, the heat inside the HIP furnace does not easily flow out and the temperature inside the furnace is controlled at the level of preheating. The appearance when a preheated can is inserted from the bottom into the HIP furnace is shown in Figure 4.

2.3. Atmosphere Control HIP Treatment

Usually, in the HIP process, high-pressure gas is used as a medium to compress the materials or a can which contains processing material. Here, the high-gas pressure works as a force to compress the material to cause densification, so inert gases such as argon or helium has been
used. However, this inert atmosphere is not always good for materials to be processed because some oxides, for example, are reduced even in the complete inert atmosphere without partial pressure of oxygen. The most well-known example is manganese zinc ferrite materials. This material exhibits soft magnetic property, which is useful for the magnetic recording heads and usually has the spinel crystal structure. This spinel structure is stable only in the narrow region of temperature—oxygen content of the atmosphere (partial pressure) region shown in Figure 5 [3]. As can be seen in this figure, the wustite is formed by the reduction, if there is not enough oxygen partial pressure, especially at high temperatures. So, sometimes reduction was observed during sintering to fabricate soft ferrite blocks before cutting out pieces for magnetic core. In the actual production of magnetic recording heads, the powder embedding method explained later was adopted, but a trial directly to control the oxygen partial pressure was carried out by the partial pressure method. Further, silicon nitride ceramics which drew attention in the 1980s for the high-temperature gas turbine components was found not to be sintered to the full density at ambient pressure and the use of HIP method seemed very effective. However, HIPing of silicon nitride at high temperature under argon was not successful because of the decomposition of silicon nitride. Actually, the decomposition of silicon nitride occurs according to the nitrogen pressure—temperature diagram for the decomposition of silicon nitride shown in Figure 6 [4]. To improve the densification, the use of nitrogen in place of argon was tried and found very effective to produce high-strength silicon nitride.

As explained above, there are a couple of ways to control the atmosphere surrounding the work pieces. Those were termed as 1) the full-pressure method, 2) the partial-pressure method, and 3) the powder embedding method.

2.3.1. Full-Pressure Method

The term full pressure is used here in contrast to the partial-pressure method explained in the next section. In this method, an appropriate gas for the processing material is chosen as pressure transmitting gas. In the case of silicon nitride ceramics which drew attention in the 1980s for the high-temperature gas turbine components.
nitride, nitrogen gas is used as high-pressure gas in place of argon to suppress the decomposition of silicon nitride. This example is already briefly explained above.

In addition to the above, the full-pressure method is very useful for the preparation of nitrides from pure metals. Metal powders or the surface of metal sheet can be easily nitrided by the full-pressure method of nitrogen. Especially in the case of molybdenum nitride, it is drastic, because it is common knowledge that molybdenum cannot be nitrided by nitrogen at ambient pressure. However, H. Jehn and P. Ettmayer made an investigation on the preparation of Mo–N system phase diagram under high pressures and postulated that dissociation pressure of Mo$_2$N could be represented by the equation

$$\log(p_{N2}) = \frac{5.63}{T} - 5990$$

for \( p_{N2} \) in atm and \( T \) in K.

The authors have attempted the formation of molybdenum nitride under pressures up to 180 MPa [5]. Some of the results are shown in Figure 7 (a) (b). In this experiment, molybdenum powder with the particle size of ~44 microns was used. The retention time at the maximum temperature was kept at 1 h based on a preliminary experiment for 0.5-mm thickness molybdenum wires to evaluate the nitrogen diffusion rate in the molybdenum. The crystalline structure identified by XRD (X-ray Diffraction) was Mo$_{16}$N$_7$ according to the ASTM data. It is very interesting to note that the amount of nitride varied with pressure at a fixed temperature, because for other materials it was not greatly affected by the pressure as shown in Figure 7 (b) for NbN. At higher temperatures, MoN which has been drawing attention as a high-Tc, B1 type superconductor may be synthesized.
2.3.2. Partial-Pressure Method

This method is sometimes called as gas mixture method. A gas component to create the targeted atmosphere is mixed with the pressurizing gas argon. The partial pressure, that is, the amount of the gas added is calculated from the equilibrium diagram of the processing material. Usually, the pressure needed to densify the presintered ceramics and the pressure to suppress their decomposition are not identical. In such a case, the gas component targeted is added up to the pressure to suppress the decomposition and argon is added to reach the total pressure needed to realize the full densification. The most typical example of this method is oxygen HIP treatment of oxide ceramics. In this case, from the viewpoint of safety, the amount of oxygen mixed into argon is limited to 20 vol% which is the same level with the air.

In the case of Mn–Zn ferrite, the oxygen partial pressure has to be maintained in the spinel region shown in the zone shown in Figure 5 [6] as explained earlier. One of the problems in this case is that oxygen partial pressure needed is very low and has to be controlled with the temperature very carefully and was not adopted in the commercial production because of the complexity of the control.

Another example of the application of oxygen HIP is HIP treatment of alumina and zirconia ceramics [7]. Conventionally, HIP treatment of presintered alumina was carried out in argon atmosphere with molybdenum or graphite heater and it was known that the strength after HIP treatment is improved by 10–20%; however, the color after HIP treatment is a little different form the one sintered under normal air conditions. The effect of HIP treatment of alumina ceramics under various conditions is summarized in Figure 8. In the case of zirconia, HIP treatment of presintered parts was HIP treated in argon atmosphere with a graphite heater and about 50% of improvement in the bending strength was experimentally confirmed, although the color changes from white to black. However, this HIPed black zirconia got good reputation with very high strength. The influence of the atmosphere in the HIP treatment of zirconia is shown in Figure 9.

One of the examples applied to commercial production is the high-pressure sintering of Bi–Sr-super-conductive ceramics. In 2003, a large high-pressure gas-sintering unit for the production of Bi–Sr- system super-conductive cables was installed in one of the Japanese electric cable manufacturing companies. The specifications of this unit are 1.2 m in diameter and 1.2 m in height, 950 °C, maximum temperature, and 30 MPa maximum operation pressure. Roughly wound cable materials were sintered in an atmosphere of argon with several hundred ppm of oxygen. The cables manufactured using this unit are already used for ground testing in a power generation system built in Albany, NY. The configuration of this high-pressure gas-sintering unit is shown in Figure 10 [2].

2.3.3. Powder Embedding Method

This method has been used for the production of soft ferrites, such as Mn–Zn ferrite and Ni–Zn ferrite, and lead titanate ceramics. Presintered blocks are packed with powders of the same composition or other chemically stable materials in a container made of chemically stable
material with a lid. This container is charged into an ordinary HIP unit and processed. During heating, some oxygen will be emitted from the packing powders to form an equilibrium partial pressure of oxygen at the corresponding temperature. To avoid the reduction of the processed blocks, the space inside the container has to be densely packed with powder or the processed material itself and, at the same time, the convection of high-pressure gas near the processed material must be prevented. The greatest advantages of this method are that: 1) this method utilizes a natural phenomenon and appropriate atmosphere can be spontaneously created, and 2) under high pressures, the partial pressure of the desired constituent can be higher than 1 atm which is the absolute limit at ambient pressure.

A very good example of the application of this method is the sintering and gas-pressure densification of the coral sand ceramics [5]. Coral sand is a natural material composed of calcium carbonate with several percent of magnesium carbonate. When coral sand material is heated
over 600 °C under atmospheric pressure, it decomposes and forms calcium oxide, lime, and cannot be obtained as sintered coral sand material. However, the application of this method, by using coral sand or calcium power as embedding material, carbon dioxide gas is liberated in the crucible and forms CO₂ gas-rich atmosphere which suppresses the decomposition of calcium carbonate. The experimental results of this method are summarized in Table 1.

### 3. CONCLUSIONS AND FUTURE PERSPECTIVES

In the HIP treatment, temperature, pressure, and holding time have been the parameters which decide the

<table>
<thead>
<tr>
<th>No.</th>
<th>Temperature (°C)</th>
<th>Pressure (MPa)</th>
<th>Time (h)</th>
<th>CaO/CaCO₃ (%)</th>
<th>Density (g/cm³)</th>
<th>CaO/CaCO₃ (%)</th>
<th>Density (g/cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>2.0</td>
<td>1</td>
<td>0</td>
<td>2.21</td>
<td>0</td>
<td>2.18</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>10.0</td>
<td>1</td>
<td>3.0</td>
<td>2.20</td>
<td>0</td>
<td>2.30</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>100.0</td>
<td>1</td>
<td>20.0</td>
<td>2.36</td>
<td>0</td>
<td>2.33</td>
</tr>
<tr>
<td>4</td>
<td>1000</td>
<td>200.0</td>
<td>1</td>
<td>6.0</td>
<td>2.39</td>
<td>0</td>
<td>2.30</td>
</tr>
<tr>
<td>5</td>
<td>1100</td>
<td>2.0</td>
<td>1</td>
<td>90.7</td>
<td>2.14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>1100</td>
<td>10.0</td>
<td>1</td>
<td>3.2</td>
<td>2.34</td>
<td>1.8</td>
<td>2.33</td>
</tr>
<tr>
<td>7</td>
<td>1200</td>
<td>20.0</td>
<td>1</td>
<td>12.1</td>
<td>2.03</td>
<td>2.5</td>
<td>2.33</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>10.0</td>
<td>5</td>
<td>115.0</td>
<td>2.18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>1100</td>
<td>+1000</td>
<td>1</td>
<td>0</td>
<td>2.60</td>
<td>0</td>
<td>2.18</td>
</tr>
</tbody>
</table>

**TABLE 1** Experimental Conditions and Resultant Densities and CaO/CaCO₃ Ration by XRD for Coral Sand and Reagent Class Calcium Carbonate

**Figure 10** Oxygen partial pressure control HIP unit for Bi–Sr system super-conductive cables.
properties such as density, mechanical strength of the products after processing, and, now in the case of ceramics, atmosphere is taken as the third important parameter which decides functional properties or colors. Thus, the importance of the atmosphere control during HIP treatment is being recognized as important as temperature or pressure.

In this article, recent advances in HIP technology is summarized placing a stress on the high-productivity HIP equipment and the atmosphere control HIP treatment technology which is attracting attention, especially in the area of the manufacture of dense ceramic materials. Now special atmospheres such as high partial pressure of zinc or even some other metals can be generated with a special HIP unit, which contains a gas-tight casing inside the furnace. These new technologies will make it possible to form special atmospheres and will open a new future of the HIP process for ceramics.

REFERENCES
Handbook of Advanced Ceramics
Materials, Applications, Processing, and Properties
Second Edition

Shigeyuki Somiya, Editor-in-Chief
Professor Emeritus of Tokyo Institute of Technology, Japan
Masanori Abe (Ch. 7.3), Professor Emeritus of Tokyo Institute of Technology, Japan

Joerg Adler (Ch. 8.1), Fraunhofer-Institut fuer Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Tadafumi Adschiri (Ch. 11.1.5), World Premier International Research Center-Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

Kazuki Akamatsu (Ch. 4.1), Department of Environmental and Energy Chemistry, Faculty of Engineering, Kogakuin University, 2665-1 Nakanomachi, Hachioji, Tokyo 192-0015, Japan

Jun Akedo (Ch. 10.5), National Institute of Advanced Industrial Science and Technology, 1-2-1 Namiki, Tsukuba, Ibaraki 305-8564, Japan

Fritz Aldinger (Ch. 11.1.10), Professor Emeritus of University of Stuttgart (Universitaet Stuttgart), Germany, Former Affiliate: Max-Planck-Institute for Metals Research (Max-Planck-Institut fuer Metallforschung) (at present Max-Planck-Institute for Intelligent Systems; Max-Planck-Institut fuer Intelligente Systeme)

Rebecca Alway-Cooper (Ch. 2.8), Center for Advanced Engineering Fibers and Films, and Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina 29634, United States

Nobuaki Aoki (Ch. 11.1.5), World Premier International Research Center-Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

Toshihiko Arita (Ch. 11.1.5), Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

Alexander Badev (Ch. 6.5), SPCTS (Science des Procédés Céramiques et de Traitements de Surface; Science of Ceramic Processing and Surface Treatments) - UMR 7315 CNRS (Centre National de la Recherche Scientifique; National Center for Scientific Research), European Ceramic Centre, 12, rue Atlantis, 87068 Limoges, France

George H. Beall (Ch. 5.1), Corning Inc., SP-FR-05-1, Corning, New York 14831, United States

Raul Bermejo (Ch. 9.7), Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Franz-Josef-Strasse 18, 8700 Leoben, Austria

Diego Cazorla-Amorós (Ch. 2.9), Dpto. de Química Inorgánica e Instituto de Materiales (Department of Inorganic Chemistry and Institute of Materials), Universidad de Alicante (University of Alicante), Aptdo 99, 03080 Alicante, Spain

Thierry Chartier (Ch. 6.5), SPCTS (Science des Procédés Céramiques et de Traitements de Surface; Science of Ceramic Processing and Surface Treatments) - UMR 7315 CNRS (Centre National de la Recherche Scientifique; National Center for Scientific Research), European Ceramic Centre, 12, rue Atlantis, 87068 Limoges, France

Toyohiro Chikyo (Ch. 11.1.11), Comet Inc., 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan

Robert Danzer (Ch. 9.1), Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Peter-Tunner-Strasse 5, 8700 Leoben, Austria

Marco Deluca (Ch. 9.7), Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Franz-Josef-Strasse 18, 8700 Leoben, Austria

Mildred S. Dresselhaus (Ch. 2.2), Department of Physics and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States

M. Endo (Ch. 2.2), Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan
Naoya Enomoto (Ch. 11.1.8), Department of Applied Chemistry, Faculty of Engineering, Kyushu University, 744 Motooka, Nishi, Fukuoka 819-0395, Japan

Theo Fett (Ch. 9.5), Institut fuer Angewandte Materialien-Keramik im Maschinenbau (IAM-KM) (Institute for Ceramics in Mechanical Engineering), KIT (Karlsruher Institut fuer Technologie; Karlsruhe Institute of Technology), Haid- und Neu-Strasse 7, 76131 Karlsruhe, Germany

Tobias Fey (Ch. 7.1), Department of Materials Science and Engineering (Glass and Ceramics), University of Erlangen-Nuernberg, Martensstrasse 5, 91058 Erlangen, Germany

Jeffrey T. Fong (Ch. 9.4), Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Stephen W. Freiman (Chs. 9.4, 9.6, 9.8), Freiman Consulting, 8212 Inverness, Hollow Terrace, Potomac, Maryland 20854, United States

Takao Fujikawa (Ch. 11.2.6), Metal Technology Co., Ltd. (Kinzoku-Giken, Co., Ltd.), 27th floor, Harmony-Tower, 1-32-2 Hon-cho, Nakano, Tokyo 164-8721, Japan

Koji Fujita (Ch. 5.2), Department of Material Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo, Kyoto, Kyoto 615-8510, Japan

Ryuzo Furukawa (Ch. 11.2.5), Graduate School of Environmental Studies, Tohoku University, 6-6-20 Aoba, Aramaki, Aoba, Sendai, Miyagi 980-8579, Japan

Dipankar Ghosh (Ch. 3.3), Materials Science and Engineering, University of Florida, Gainesville, Florida 32607, United States

Takashi Goto (Ch. 10.4), Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

E. Gracia-Espin (Ch. 2.2), Department of Chemistry, and Department of Physics, Umeå University, SE-901 87 Umeå, Sweden, Former Affiliate: Advanced Materials Department, IPICYT

Peter Greil (Ch. 7.1), Department of Materials Science and Engineering (Glass and Ceramics), University of Erlangen-Nuernberg, Martensstrasse 5, 91058 Erlangen

Hiroshi Handa (Ch. 7.3), Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Yoshiyasu Harada (Ch. 1.1.1), JEOL Ltd., 3-1-2 Musashino, Akishima, Tokyo 196-8558, Japan

Mamoru Hatakeyama (Ch. 7.3), Solutions Research Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

T. Hayashi (Ch. 2.2), Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

Mathias Herrmann (Ch. 3.4), Fraunhofer-Institut fuer Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Kiyoshi Hirao (Ch. 3.2), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Daisuke Hojo (Ch. 11.1.5), World Premier International Research Center-Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, 2-1-1 Katahira, Aoba, Sendai, Miyagi 980-8577, Japan

Wolfram Höland (Ch. 5.1), Research and Development, Ivoclar Vivadent AG, Bendererstrasse 2, 9494 Schaan, Liechtenstein

Takuya Hoshina (Ch. 6.1), Graduate School of Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

Hideo Hosono (Ch. 6.4), Frontier Collaborative Research Center and Materials and Structures Laboratory, Tokyo Institute of Technology, Mail Box R3-1, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Mikinori Hotta (Ch. 3.2), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Hideki Hyuga (Ch. 3.2), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Toshiyuki Ikoma (Ch. 7.2), Department of Metallurgy and Ceramics Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

Yuichi Ikuhara (Ch. 1.1.1), Institute of Engineering Innovation, School of Engineering, University of Tokyo, 2-11-16 Yayoi, Bunkyo, Tokyo 113-8656, Japan

Michio Inagaki (Ch. 2.1), Professor Emeritus of Hokkaido University, Japan
Masashi Inoue (Ch. 11.1.4), Professor Emeritus of Kyoto University, Japan

Emanuel Ionescu (Ch. 11.1.10), Technische Universitaet Darmstadt (Darmstadt University of Technology) - Institut fuer Materialwissenschaft (Institute of Materials Science), Petersenstrasse 23, 64287 Darmstadt, Germany

Emile H. Ishida (Ch. 11.2.5), Graduate School of Environmental Studies, Tohoku University, 6-6-20 Aoba, Aramaki, Aoba, Sendai, Miyagi 980-8579, Japan

Hiroaki Itami (Ch. 2.7), Product Planning Department, Toyo Tanso Co., Ltd., 2181-2 Nakahime, Ohnohara-cho, Kanonji, Kagawa 769-1612, Japan

Yoshiyasu Ito (Ch. 10.2), Tocalo Co. Ltd., 4-13-4 Fukaekita-machi, Higashinada, Kobe, Hyogo 658-0013, Japan, Former Affiliate: R&D Center, Power and Industrial Systems, Toshiba Corp.

Toshio Itoh (Ch. 11.1.7), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Noriya Izu (Ch. 11.1.7), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Moriyama, Nagoya, Aichi 463-8560, Japan

Young-Pyo Jeon (Ch. 2.8), Center for Advanced Engineering Fibers and Films, and Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina 29634, United States

Se-Young Jeong (Ch. 11.1.11), Department of Cogno-Mechatronics Engineering, Pusan National University, Miryang, Gyeongnam 627-706, Korea

Kiyoshi Kanamura (Ch. 4.2), Department of Applied Chemistry, Graduate School of Urban Environmental Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

Kiyofumi Katagiri (Ch. 11.1.9), Department of Applied Chemistry, Graduate School of Engineering, Nagoya University

Yuuuzou Kawahara (Ch. 10.3), Dai-Ichi High Frequency Co., Ltd., 1-45 Mizue-cho, Kawasaki, Kawasaki, Kanagawa 210-0866, Japan, Former Affiliate: Yokohama R&D Center, Mitsubishi Heavy Industries, Co., Ltd.

Hitotoki Kawaji (Ch. 6.3), Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

Y.A. Kim (Ch. 2.2), Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

Hideki Kita (Chs. 3.2, 11.2.1), Department of Molecular Design and Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8603, Japan, Former Affiliate: National Institute of Advanced Industrial Science and Technology

Hideo Koinuma (Ch. 11.1.11), Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan

Sridhar Komarneni (Ch. 11.1.1), Materials Research Institute, Materials Research Laboratory building, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

Naoki Kondo (Ch. 3.2), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8550, Japan

Masashi Kotobuki (Ch. 4.2), Department of Applied Chemistry, Graduate School of Urban Environmental Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

Kunihito Koumoto (Ch. 11.1.9), Department of Applied Chemistry, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8603, Japan

Ji-Guang Li (Ch. 11.1.6), National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

Angel Linares-Solano (Ch. 2.9), Dpto. de Química Inorgánica e Instituto de Materiales (Department of Inorganic Chemistry and Institute of Materials), Universidad de Alicante (University of Alicante), Aptdo 99, 03080 Alicante, Spain

Mikk Lippmaa (Ch. 11.1.11), Institute for Solid State Physics, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8581, Japan

F. López-Urías (Ch. 2.2), Advanced Materials Department, IPICYT (Instituto Potosino de Investigación Científica y Tecnológica, A.C.), Camino a la Presa San José 2055, Colonia Lomas 4a sección, 78216, San Luis Potosí, Mexico

Tanja Lube (Ch. 9.1), Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Peter-Tunner-Strasse 5, 8700 Leoben, Austria

Hirotaka Maeda (Ch. 11.2.5), Center for Fostering Young and Innovative Researchers, Nagoya Institute of
Technology, Gokiso-cho, Showa, Nagoya, Aichi 460-8555, Japan

Yasuo Manabe (Ch. 11.2.6), Machinery Division, Kobe Steel, Ltd., 2-3-1 Shinhama, Arai-cho, Takasago, Hyogo 676-8670, Japan

Lalit M. Manocha (Ch. 2.10), Department of Materials Science and Centre for Advanced Studies in Carbon and Nanomaterials, Sardar Patel University, Vallabh Vidyanagar, Gujarat-388120, India

Branko Matović (Chs. 3.1, 4.3), Vinca Institute of Nuclear Sciences, University of Belgrade, 11001 Belgrade, Serbia, Former Affiliate: Research Laboratory of Nuclear Reactor, Tokyo Institute of Technology

Ichiro Matsubara (Ch. 11.1.7), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Yuji Matsumoto (Ch. 11.1.11), Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori, Yokohama, Kanagawa 226-8503, Japan

David McKinney (Ch. 11.1.3), MM Virtuoso, Inc., Gainesville, Florida 32606, United States

Kimitaka Minami (Ch. 11.1.5), New Industry Creation Hatchery Center, Tohoku University, 6-6-10 Aoba, Aramaki, Aoba, Sendai, Miyagi 780-8577, Japan

Marlon Morales (Ch. 2.8), Center for Advanced Engineering Fibers and Films, and Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina 29634, United States

R. Morrell (Ch. 9.1), Materials Division, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

Hirokazu Munakata (Ch. 4.2), Department of Applied Chemistry, Graduate School of Urban Environmental Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

Dietrich Munz (Ch. 9.5), Professor Emeritus of Karlsruhe Institute of Technology (Karlsruher Institut fuer Technologie), Germany

Shunsuke Murai (Ch. 5.2), Department of Material Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo, Kyoto, Kyoto 615-8510, Japan

H. Muramatsu (Ch. 2.2), Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

Takashi Nakagawa (Ch. 7.3), Management of Industry Technology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Shin-ichi Nakao (Ch. 4.1), Department of Environmental and Energy Chemistry, Faculty of Engineering, Kagokai University, 2665-1 Nakano-machi, Hachioji, Tokyo 192-0015, Japan

Amod A. Ogale (Ch. 2.8), Center for Advanced Engineering Fibers and Films, and Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina 29634, United States

Tatsuki Ohji (Chs. 9.2, 9.3, 11.2.2), National Institute of Advanced Industrial Science and Technology, 2266-98 Anagahora, Shimoshidami, Moriyama, Nagoya, Aichi 463-8560, Japan

Osamu Okada (Ch. 2.6), Product Planning Department, Toyo Tanso Co., Ltd., 2181-2 Nakahime, Ohnohara-cho, Kanonji, Kagawa 769-1612, Japan

Eiji Osawa (Ch. 2.3), NanoCarbon Research Institute, Ltd, AREC (Asama Research Extension Center), Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda, Nagano 386-8567, Japan

Jun-ichi Ozaki (Ch. 2.4), Graduate School of Engineering, Gunma University, 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan

Uwe Petasch (Ch. 8.1), Fraunhofer-Institut fuer Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Ralf Riedel (Ch. 11.1.10), Technische Universitaet Darmstadt (Darmstadt University of Technology) - Institut fuer Materialwissenschaft (Institute of Materials Science), Petersenstrasse 23, 64287 Darmstadt, Germany

Rustum Roy (Ch. 11.1.1), Materials Research Institute, Materials Research Laboratory building, The Pennsylvania State University, United States; Passed away on August 26, 2010

Sumio Sakka (Ch. 11.1.2), Professor Emeritus of Kyoto University, Japan

Yoshio Sakka (Ch. 11.1.6), Advanced Materials Processing Unit, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

Cross J. Scott (Ch. 7.2), Department of Metallurgy and Ceramics Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan

Taiju Shibata (Ch. 2.5), Japan Atomic Energy Agency, 4002 Narita-cho, Oarai-machi, Higashiibaraki, Ibaraki 311-1393, Japan
Robert Danzer, Institut fuer Struktur- und Funktionskeramik (Institute for Structural and Functional Ceramics), Montanuniversitaet Leoben (University of Leoben), Peter-Tunner-Strasse 5, 8700 Leoben, Austria

Stephen W. Freiman, Freiman Consulting, United States, Former Position: Division Head, United States National Institute of Standards and Technology, United States

Peter Greil, Professor of University of Erlangen-Nuernberg (Department of Materials Science and Engineering, (Glass and Ceramics)), Germany

Mathias Herrmann, Fraunhofer-Institut fuer Keramische Technologien und Systeme (Fraunhofer IKTS; Fraunhofer Institute for Ceramic Technologies and Systems), Winterbergstrasse 28, 01277 Dresden, Germany

Masayuki Kaneno, Former Position: Manager, Nagoya Industrial Science Research Institute, Japan

Katsutoshi Komeya, Professor Emeritus of Yokohama National University, Japan

Tatsuki Ohji, Prime Senior Research Scientist, AIST Chubu, National Institute of Advanced Industrial Science and Technology, Japan

Shigeyuki Somiya (Editor-in-Chief, the First and the Second Edition), Professor Emeritus of Tokyo Institute of Technology, Japan and Teikyo University of Science and Technology, Japan

Eichi Yasuda, Professor Emeritus of Tokyo Institute of Technology, Japan
To Mentors

Prof. Fritz Aldinger
Prof. Nils Claussen
Prof. Robert (Bob) F. Davis
Prof. Shikao Ikehara
Prof. Yunoshin Imai
Prof. Chihiro Kawashima
Dr. Robert Alfred Laudise (Bob)
Prof. Sanichiro Mizushima
Prof. Taro Moriya
Prof. Arnulf Muan
Prof. Tokiti Noda
Prof. Elburt Franklin Osborn
Prof. Joseph Adam Pask (Joe)
Prof. Günter Petzow (Petz)
Prof. Albrecht Rabenau
Prof. Rustum Roy
Prof. Shinroku Saito
Professor Ikutaro Sawai
Prof. Richard M. Spriggs
Prof. Hiroshige Suzuki

Prof. Keiichiro Takezawa
Prof. Megumi Tashiro
Prof. Toshiyoshi Yamauchi

To Old Friends

Jean Francois Baumard
Richard C. Bradt (Dick)
Robert Pang Heng Chang (Bob)
Robert C. DeVries (Bob)
Fred P. Glasser (Fred)
Jack B. MacChesney (Jack)
Bert Phillips (Bert)
Helmut Schubert

To Parents

Father in Law: Uichi Hashimoto
Father: Takayuki Somiya

Shigeyuki Somiya, Editor-in-Chief
Acknowledgments to the First Edition

First and foremost, I would like to thank all the co-editors, Fritz Aldinger, Nils Claussen, Richard M. Spriggs, Kenji Uchino, Kunihito Koumoto and Masayuki Kaneno for their valuable suggestions with regard to the chapters, authors, and their editorial assistance. Without their help, this book would not have been possible. Especially, Fritz Aldinger, Nils Claussen, and Richard M. Spriggs gave me good advice for this book. Mr Kaneno offered secretarial assistance.

Second, I wish to thank all the publishers who gave permissions to authors to reproduce and use the materials from their original published papers.

Third, many authors wrote their chapters within a short time in spite of their tight schedule. I would like to express my gratitude to all these authors.

I would also like to extend my appreciation to Ms Amanda Weaver and her group at Elsevier, Oxford for their contribution to the publishing works.

Finally, I was able to study abroad in the USA under the Fulbright Exchange Program and in Germany under the scholarship program by Max Planck Institut fuer Metallforschung, Pulvermetalligisches Laboratorium. Without these experiences, I would not have made it as Editor-in-Chief of this book. I thank all my professors and friends around the world who have made this possible.

Shigeyuki Somiya,
Editor-in-Chief
First, I would like to thank

Dr. Paul F. Becher, Prof. Arthur H. Heuer, and Prof. Günter Petzow because they recommended coeditors of this book. Without their efforts, it was difficult to find excellent members.

Second, I would like to thank all the coeditors:

Prof. Robert Danzer, Dr. Steve Freiman, Prof. Peter Greil,
Dr. Mathias Herrmann, Mr. Masayuki Kaneno, Prof. Katsutoshi Komeya,
Dr. Tatsuki Ohji, and Prof. Eiichi Yasuda.

Coeditors made their valuable suggestions and advice with regards to editorial work, such as the organization of chapters and sessions, and the recommendation of authors. Without them, this book would have not been created. For example, Editor Masayuki Kaneno offered excellent secretarial assistance.

Third, I would like to express my appreciation to all authors for this book and people who recommended authors.

Fourth, I would like to extend my appreciation to

Dr. Louisa Hutchins, Mr. Poulouse Joseph, Dr. Steve Merken,
Dr. Graham Nisbet, and people at Elsevier for their contribution to the publication of this book.

Fifth, I would like to thank all organizations and publishers who have given permissions to authors to reproduce and use materials from original papers.

Sixth, I appreciate my father-in-law Uichi Hashimoto and my father Takayuki Somiya. They kept encouraging my study. Without them, it was difficult to find mentors and friends, especially friends in the world.

I studied at the Department of Geochemistry, The Pennsylvania State University in the U.S.A. under the Fulbright Exchange Student Program, and at the Pulvermetallurgisches Laboratorium in the Max Planck Institute for Metals Research (at present the Max Planck Institute for Intelligent Systems) in the F. R. G. (West Germany; at present Germany) under the scholarship program supported by the Pulvermetallurgisches Laboratorium. Without these experiences, I would not have accomplished this publication as editor in chief.

Last, I would like to thank mentors, colleagues, students, friends, and my family who have made this publication possible.

Shigeyuki Somiya,
Editor-in-Chief
Preface to the First Edition

In 1989 Shigeyuki Somiya, the Editor-in-Chief of this book, published Advanced Technical Ceramics (Academic Press, Inc.; original publication in Japanese, 1984). Well over a decade has passed without the appearance of an authoritative new title on the ever-changing subject of Advanced Ceramics. The purpose of this book is to provide an up-to-date account of the present status of Advanced Ceramics, from fundamental science and processing to application.

The Handbook of Advanced Ceramics has an internationally renowned group of contributing editors. They are well known throughout the world in their fields of study. These editors discussed the contents and chose the authors of each of the book’s chapters very carefully. The chapters consist of review and overview papers written by experts in the field.

Up until about 50 years ago, ‘ceramics’ were considered to be porcelains, bottle glass, sheet glass, refractory bricks, enamels, cements, lime, gypsum and abrasives. In recent years the field of ceramics has broadened and expanded. Ceramics are now used in new fields of research as well as in the old fields. This handbook describes these developments and the new processes and applications.

The handbook will enable the reader to understand the present status of Ceramics and will also act as an introduction, which may encourage further study, as well as an estimation of the role advanced ceramics may have in the future.

The handbook is a two-volume set. Part I deals with Materials Science and Part II with Processing and Applications.

Part I serves as an introduction to the basic science, raw materials, forming, drying, sintering, innovative processing, single crystal growth, machining, joining, coating, fracture mechanics, testing, evaluation, etc. Part I is intended to provide the reader with a good understanding of the new techniques in advanced ceramics, such as thin films, colloidal processing, active and passive filler, pyrolyses process and precursor derived ceramics, as well as providing a template for the deposition of ceramics from aqueous solutions.

Part II deals with more recent processes and applications and functional and engineering ceramics. The engineering ceramics covered in this book were developed within the last decade. The functional ceramics covered include electroceramics, optoelectro-ceramics, superconductive ceramics, etc. as well as the more recent development of piezoelectric ceramics and dielectric ceramics.

The use of ‘Engineering’ Ceramics, introduces entirely new fields to be considered. These include mechanical properties, decorative ceramics, environmental uses, energy applications, bioceramics, composites, functionally graded materials, intelligent ceramics and so on.

The term Advanced Ceramics is opposite in meaning to ‘Traditional’ or ‘Classical’ Ceramics. In the past, Advanced Ceramics were often confused with New or Newer Ceramics, Modern Ceramics, Special Ceramics and so on. Furthermore, Fine Ceramics, at least in the USA and Europe, is synonymous with Fine Grain Ceramic Products and/or Fine Grain Porcelain; Fine Ceramics in Japan is similar to what we understand as Advanced Ceramics. So for this edition, the term Advanced Ceramics was chosen as the most suitable title for a book providing an in-depth survey of the current state of Ceramics Science and its applications.

It is the editors’ wish that this book will provide the reader with a detailed understanding of the many applications of Advanced Ceramics in both today’s world and in that of the future.

The editors wish to thank all those who participated in the preparation of this book such as authors, publishers and copyright owners in Europe, USA, Asia and the rest of the world.

Fritz Aldinger
Nils Claussen
Masayuki Kaneno
Kunihito Koumoto
Shigeyuki Somiya, Editor-in-Chief
Richard M. Spriggs
Kenji Uchino
The first edition of the *Handbook of Advanced Ceramics* from Elsevier was published in 2003 and almost 10 years has passed. Science and technology in the field of ceramics, however, have developed day by day. This is one of reasons why the second edition of the *Handbook of Advanced Ceramics* has been published.

First ceramic products would be low-firing earthenware appeared some 15,000 years ago. This is a starting point subsequent to developments of ceramics. Until about 200 years ago, the word ceramics meant pottery, and ceramic products were limited to tableware, roofing tiles, bricks, and clay pipes. These ceramics or ceramic products are called traditional ceramics or classic ceramics.

In the nineteenth century, products of the iron and steel industry required new refractory bricks containing silica, alumina, chrome-magnesia, and magnesia, which were different from traditional fireclay bricks. Since then ceramics also have made essential contributions in many fields of the society.

There are a few words to express ceramics against classic ceramics or traditional ceramics: advanced ceramics, new ceramics, modern ceramics, special ceramics, and technical ceramics. New and modern ceramics mean ceramics against classic or traditional ceramics. Moreover, special ceramics also has a similar meaning as new and modern ceramics, and it originated from special refractories or special porcelains. As for technical ceramics, it includes all industrial ceramics except products related to traditional ceramics, such as pottery, porcelain, and tableware. The meaning of advanced ceramics is similar to the meaning of technical ceramics. However, advanced ceramics particularly emphasizes advanced features in ceramics. Therefore, advanced ceramics is used in this handbook to express “current ceramics.”


Applications of advanced ceramics are based on and resistant to various properties, such as biological, chemical, electrical, mechanical, optical, physical, structural, and thermal properties. In addition, there are various applications: For example, biocompatible parts, catalysts, sensors, electrical conductors, electrical insulators, semiconductors, superconductors, positive and negative temperature coefficient resistors, magnetic parts, dielectrics, ferroelectrics, piezoelectrics, pyroelectrics, abrasives, carbon fibers, hard parts, lubricants, high-temperature structural parts, automobile engine parts, cutting tools, optical fibers, laser oscillators, optically transparent parts, thermal conductors, thermal insulators, heaters, low-thermal expansion coefficient products, and jewelry.

Materials used for advanced ceramics are borides (e.g. boron nitride (BN)), carbides (e.g. silicon carbide (SiC)), carbons (e.g. diamond (C)), hydroxides (e.g. hydroxyapatite (Ca₅(PO₄)₃(OH))), nitrides (e.g. boron nitride (BN), Si₃N₄, Al₂O₃), oxides (e.g. alumina (Al₂O₃), barium titanate (BaTiO₃), mullite (3Al₂O₃·2SiO₂), silicon oxide (SiO₂), titanium oxide (TiO₂), zirconia (ZrO₂)), and so on (References to applications and materials of advanced ceramics: (1) Banno, H. (1984). New Ceramics. Tokyo, Japan: Pawaa-Sha (in Japanese); (2) The Japan Fine Ceramics Association (Ed). (1989). FC Annual Report for Overseas Readers: Fine Ceramics for Future Creation, p. 2; (3) Miyajima, S. (1994). New Ceramics. Tokyo, Japan: Chijin Shokan (in Japanese).)

In recent years, fields of ceramics have become broad. This handbook describes developments, processes, properties, characterization, and applications in fields of advanced ceramics: For instance, powder preparation, coating, mechanical properties, fracture mechanics, and measurement methods. In addition, it covers ceramics materials, such as carbides, carbons, and oxides.

Authors are experts in their fields and well-known scholars and engineers. The purpose of this handbook is same as the first edition of the handbook and to provide an up-to-date account of the present status of several fields of advanced ceramics from fundamental science and technology, to processing, to applications.

This handbook will enable readers, graduate students, professors, and engineers to introduce and understand the status of advanced ceramics.
The picture “Fine ceramic tree” shows relationships between advanced ceramic materials and properties. The Japan Fine Ceramics Association provided it. In Japan, the expression “Fine Ceramics” is similar to advanced ceramics and used rather than advanced ceramics. The Japan Fine Ceramics Association is acknowledged.
Activated carbon fibers (ACFs), 155–169
 activated multiwalled carbon nanotubes (AMWCNTs), 160–161
 Aerosol deposition (AD) method, 847, 849
 Aerogels, 898–900
 Aeroengine components, carbon–carbon composites as, 195
 Aerosol deposition (AD) method, 847, 850–851, 857
 AD-yttrium oxide layer, 854–855
 alumina, morphology of, 932
 AlN, 775
 Alumina
 x-alumina, glycothermal synthesis of, 930–932
 effects of seeds, 931–932
 foreign ions, occlusion of, 931
 gas–liquid equilibrium, 932
 glycothermal x-alumina, morphology of, 932
 microcrystalline gibbsite, reaction of, 930
 thermodynamical consideration, 930–931
 for high-temperature use, properties of, 248f
 radiation damage of, 360–361
 Aluminum nitride, 775
 radiation damage of, 361–362
 Aluminum titanate (tialite) wall flow filters, 598
 ammonolysis of chlorosilanes, 1033
 of monochlorosilanes, 1033
 Sr–B–C–N precursors from, 1042–1044
 using lift-off process with photoresist, 853f
 using mask deposition method, 852f
 yttrium oxide films formed by, 854f
 Aerospace industry, carbon–carbon composite materials for, 195
 Air-cooled structure, of furnace wall, 815
 Al-high-Si alloy materials, super-plasticity and strength improvement of, 1167–1169
 Alkoxides, hydrolysis and condensation of, 887
 Alkoxide group in silicon alkoxide, 886
 Alkylalkoxysilanes, types and effect of, 886
 use of, 888
 Alkyl halide elimination, 935
 Alkytrimethoxysilane, radical polymerization of, 888
 Allophane with micropores/mesopores, properties and applications of, 899
 Al2O3 blasting nozzle, 3-dimensional
 complex near-net-shape forming of, 1167–1169
 Al2O3/Yb2O3, SPS-consolidated, 1160
 Amorphous silica membranes, hydrogen-production technologies using, 331–341
 Amorphous solids, magnetic properties of, 389
 Amphiphilic organoalkoxysilane self-assembly, ceramization via, 1011–1013, 1012f
 Analytical electron microscope (AEM), 4f
 to ceramics, application of, 12–20
 convergent beam electron diffraction, 14–15, 15f
 high-resolution, 13–14, 13f–14f, 15f
 composition of, 4
 electron beam and specimen, interaction between, 3–4, 4f
 electron guns characteristics, 6t
 electron guns performance, 4
 objective lens, 4–6, 6t
 overview of, 3–7
 spherical aberration corrector, 6–7, 7f
 Angiosperms, 529
 Anisotropic porous structure with fibrous grain alignment, 668–669
 Anisotropic texture, 768
 Annular bright field STEM (ABF-STEM), 10, 10f
 to ceramics, application of, 16–17, 17f
 Antifogging film, 896
 Antireflective coating, 894
 Antireflective film, 894
 Apatite-containing glass-ceramics, 376–377
 Applied moment loading, 720
 Armchair graphene nanoribbons, electronic properties of, 62
 width and length dependence of, 62–63
 ASCII STL file, 517
 Ash problem, 590
 ASTM International Standards
 creep fracture measurements, 648
 fatigue/slow crack growth measurements, 643
 fracture strength measurements, 637
 Atmosphere control HIP treatment, 1205–1210
 full-pressure method, 1206–1207
 partial-pressure method, 1208
 powder embedding method, 1208–1210
 Attractive surface forces, 912
 Autoclaves, 872–874
 Automobile windows, coating films for, 894–896
 Antifogging film, 896
 Antireflective film, 894

Note: Page numbers with "/" denote figures; "t" tables.
Automobile windows, coating films for (Continued)
colored film, 895
hydrophilic coating, 896
IR-shielding film, 895
reflective film, 894–895
UV shielding film, 895
water repellent film, 896

B
Bamboos, 1180
Barium titanate, 874
Barrett–Joyner–Halenda (BJH) method, 1198
BaTiO3-based ceramics, 430
- ceramics, size effect of, 421–423
Bearings, Si3N4 applications in, 265
Biatomic tests, 640
(Bi1/2K1/2)TiO3-based system, 440–443
Binary STL file, 517
(Bi1/2Na1/2)TiO3-based ceramics, 433–435
(Bi1/2Na1/2)TiO3-(Bi1/2K1/2)TiO3-BaTiO3 system, 435–436
(Bi1/2Na1/2)TiO3-(Bi1/2Li1/2)TiO3-(Bi1/2K1/2)TiO3 system, 437–440
Binders, 919
Bioactive hydroxyapatite (HAp) coating for dental implants, 843–845, 844f–845f
Bio-based precursors, 151–152
Bioceramics and cells, interface between, 557–559
importance of, 558–559
in situ monitoring of, 559–570
challenges to, 557–558
current status of, 557–558
defined, 557
Biomorphous carbon, 531–535
Biomorphous ceramics, from lignocellulosic perovskites, 527–556
applications of, 548–551
mechanical properties of, 541–548
creep behavior, 548
erosion, 547
hardness, 547
models of, 542–543
strength, 543–545
stress and strain distribution, 527–556
FE-calculations of, 545–547
microstructure of, 541–548
Birefringence glass, 396–397
Bis(methylidichlorosilyl-ethyl)chlororoborane, 1043
B/N-doped carbon alloys, 107–108
Bombs, 872–874
Borazine-based Si–B–C–N precursors, 1038–1042
synthesis of, 1039f, 1041f
Boron-modified chlorosilanes, aminolysis reactions of, 1042–1044
Boron-modified hydridosilanes, Si–B–C–N precursors from, 1044–1045

Brake discs, carbon–carbon composites as, 195
Bridging stresses from crack profiles, 701–702
determination of, 700–703
from post-fracture tensile tests, 702–703
from R-curve, 696–697, 700–701
Brittle materials, mechanical reliability of, 675–680
crack growth parameters, 677
initial strength distribution, 676–677
time lifetime prediction expressions, 676
reference strength, 677
stresses in component, 677
uncertainty calculations, 677–679

C
C12A7:H+, light-induced doping for, 457–459
Calcination, 870
Calorimetry, 1003
Canasite, 376
Cancer metastasis, sentinel lymph node mapping for monitoring, 579–580
12CaO 7Al2O3, 457–468
crystal structure of, 457
electron-doped C12A7, thin film fabrication of, 463–466
crystal chemical, 465–466
ion implantation, 464–465
doping, 457–463
by chemical reduction, 459–463
via melt processes, 463
electronic structure of, 457
low work function of, 466–467
superconducting transition in, 467–468
Capacitance aging, in multi-layered ceramic capacitors, 418–421
Carbides
covalent, 226, 233–235
intermediate transition, 226
interstitial, 226
natural, 226–227, 227f, 241
salt-like, 225–226
silicon, 228
stellar sources of, 227–228
synthetic, 241
transition metal, 228–233
Carbon aerogels, 51
alloys, 105–109
B/N-doped, 107–108
nonoshell carbons, 105–106
atomic radius and electronegativity of, 229
blacks, 37, 37f–38f
carbon nanotubes, 40
nano-coating, 218–221
Carbonate apatite (CAP), 768
Carbon–carbon composites (C/Cs), 171–198
applications of, 194–196
fibrous performs, 172
frictional properties of, 192–193
market trends of, 195f
oxidation protection of, 193–194, 193t
processing of, 172–184, 175f
chemical vapor infiltration/densification technique, 173–176
liquid infiltration/pyrolysis technique, 176–177
nano reinforced C/Cs, 177
properties of, 172, 184–193
mechanical, 184–188
thermal properties of, 188–190
structural aspects of, 178–184
macroscopic features, 179–184
micro/macromolecular features, 178t
thermal expansion of, 190
thermal shock resistance of, 190–192
types of, 173f
Carbon fiber reinforced plastics (CFRPs), 54
Carbon fibers, 44, 45f, 143–154
bio-based precursors, 151–152
carbonization, 31
classification of, 44t
PAN-based carbon fibers, 144–147, 144f
precursors, 144–147
processing routes of, 146–147
properties of, 146t
pitch-based carbon fibers, 147–151
precursors, 147–151
properties of, 146t
rayon-based precursors, 151–152
Carbon-free polysilazanes, 1032, 1032f
Carbon haeckelite nanoribbons, 67
Carbon fiber, 33, 32f
gas-phase, 33t, 36–41
liquid-phase, 33t, 45–48
novel, 48–51
Carbon molecular sieves (CMS), 67
Carbutilation reduction, 238–239
extension from organic hydrocarbons to, 26f
materials, 25, 36–51
microtextures in, 53–56, 54f–55f
nanotexture in, 33–35, 34f, 35f, 51–53, 52f, 53f
matrix
characteristic features of, 180t
properties of, 180t
nanofibers, 39–40, 41f
nanotubes, 40
graphene/graphitic nanoribbons synthesis from, 70–71
ORR activity and structure of, 109
stellar sources of, 227–228
xerogels, with conductive polymer nano-coating, 218–221

Index
Colloidal silica, 213t
Colloid templating, organic—inorganic hybrid hollow capsules preparation by, 1016—1021
Colorants, 506
Colored film, 895
Combinatorial laser molecular beam epitaxy (CLMBE), 1111—1112, 1112f
Combinatorial material business, 1122—1123
Combinatorial technology in ceramics research, 1107—1115
Composition spread libraries, 1107—1109, 1108f—1109f
single-crystal film growth, combinatorial screening of flux materials for, 1114—1115
superlattice libraries, 1111—1113
temperature-gradient epitaxy, 1113—1114
ternary phase diagrams, 1109—1111, 1110f—1111f
discoveries of, 1115—1121
co-doped TiO2, room-temperature transparent ferromagnetism in, 1115—1117
photocatalytic TiO2, quantum size effect in, 1117—1121
single-crystal oxide films growth, by flux-mediated epitaxy, 1117—1121
instrumentation of, 1105—1107
in materials research, 1103—1107, 1104f
COMET Inc., 1122—1123
Composite tubing, 817
Composition spread libraries, 1107—1109, 1108f—1109f
Compound electroplating, NSCD as strengthening additive in, 99—100
Compressive strength of ceramics, 613—614
Compressive tests, 639
creep, 651
for fracture strength, 639
Constant volume techniques, 923—925
extrusion, 923—924
gel casting, 923
injection molding, 924
pressing, 924—925
Continuous regeneration trap (CRT), 589
Convergent beam electron diffraction (CBED) to ceramics, application of, 14—15, 15f
Cooled furnace wall, 812—815
copper-cooled structure, 815
ceramic tile system, 815
water cooled lining, 812—815
Cordierite
electrical and magnetic properties of, 378—379
wall flow filters, 597—598
Core—shell-type ceria/polymer nanoparticles, 992—994
characteristics of, 992—994
sols, characterization of, 994
Corrosion of ceramic sealing/bearing materials, 314—324
resistance of silicon nitride ceramics, 316—321, 320f
stability of silicon carbide, 321—324, 321t
Covalent carbides, 226, 233—235
Cracking (branching, 699
growth in crystals, 759—761
molecular models of, 758—759
rates, determination of, 721, 723—724
healing due to oxidation, 724
introducing and guiding, 720—721
profiles, bridging stresses from, 701—702
surface, in flexure, 725—726
tip toughness, determination of, 695—696
from load versus displacement curves, 695—696
from near-tip crack opening displacements, 695
R-curves extrapolation to Δa→0, 695
Cr-doped millfuse glass ceramics, spectral properties of, 374—375
Creep
behavior of biomorphous ceramics, 548
fracture, testing and evaluation of, 650—654
compressive creep test, 651—652
flexural creep test, 651
silicon nitride, 652—654
standards for, 651—652
tensile creep test, 651
Cr2O3/NiCr cermet coating, 822—825
Crystalline diamond coatings, 313f
Crystallochemistry, of transition metal carbides, 230—233
Crystals, crack growth in, 759—761
Cutting tool tips, Si3N4 applications in, 761—763
DC pulse generator, 1157
density functional theory within local spin density approximation (DFT-LSDA), 263—265
Densification, by SPS sintering, 1154—1158
deNOx catalysts, 594
DeNOx catalysts, 594—595
Dense SiAlON nanoceramics, fabrication of, 315—320
Dental implants, HAe coating for, 843—845, 844f—845f
Diamagnetic glasses, 396
Diamond, 40—41
family, structure of, 30, 30f
Diamond-like carbon (DLC), 30, 40—41
coatings, 313f
Diaphragm actuators, 857
Dielectric crystals
 topics, 1108f
Die pressing, 924
Diesel particulate filters (DPFs), 585—606, 1131—1132
ceramic, see Ceramic diesel particulate filters
ceramic fiber filters, 600
design and properties of, 590—596
backpressure, 592—593
coating and catalysts, 594—595
filtration efficiency, 593
heat capacity, 595
load capacity, 593
manufacturing costs, 595
mass, 596
material selection, 590—591
mechanical stability, 595
pore size, 591—592
regenerability, 595
shape, 591—592
shaped parts, flexibility of, 596
simulation tools, 596
volume, 591—592, 596
market situation for, 601—602
metal, 601
open-cell ceramic foam filters, 600—601
outlook, 602—603
working principle of, 588—590
ash problem, 590
filter regeneration, 588—590
filtration of, 588
Diesel particulate matter (DPM), 585—586
effects on health, 586
limits of, 586—588
cars and commercial vehicles, emission limits for, 586—587
emissions/pollutions, 586
for nonroad machinery and stationary equipment, 587
workplace limits, 587—588
origin of, 585—586
sources of, 585—586
Differential pressure process, 176
Differential thermal analysis (DTA), 1051
Digital micromirror device (DMD), 514
Dimethoxynaphthalene (DMNPS), 333—334
(DADB), 1042
Dimethyldioxyamine (DMDO), 902
Dimethyl sulfoxide (DMSO), 929
Flexure, surface crack in, 725–726
Flocculation, 914–915
Flory—Huggins parameter, 917
Flow reactor apparatus, for supercritical hydrothermal synthesis, 953, 953f
Fluorescent bioprobes, 97–98
Foreign atoms in structure, acceptance of, 31
Forging techniques, fine grain control via, 661–665
Form study prototype, 491
Fracture
indentation, 726–727
mechanics, see Fracture mechanics
resistance, 663–664
strength, see Fracture strength
toughness, see Fracture toughness
Fracture mechanics, 681–715
background of, 718–719
basic relations of, 681
bridging stresses, determination of from crack profiles, 701–702
determination of, 700–703
from post-fracture tensile tests, 702–703
from Raman spectroscopy, 703
from R-curve, 696–697, 700–701
data interpretation and use of, 729–730
double cleavage drilled compression, 728–729
double torsion test, 721–724
crack growth measurements, 723–724
toughness, determination of, 723
based on flexural loading, 724–726
chevron notch bend tests, 724–725
flexure, surface crack in, 725–726
single-edge-notch-bend test, 724
single-edge-precracked-beam method, 725
fracture toughness
determination of, 684–688
determination of, 684–688
edge-cracked bending bar, 684–685
mixed-mode, 688–690
determination of, 688–689
effective stress intensity factors, 689–690
mode-Ⅱ, 688–690
determination of, 688–689
effective stress intensity factors, 689–690
specimen with chevron notches, 686
specimen with Knoop indentation, 687–688
Fracture toughness
apparent, 742–743
of carbon—carbon composites, 187
determination of, 684–688, 721, 723
effective stress intensity factors, 689–690
mode-Ⅱ, 688–690
determination of, 688–689
effective stress intensity factors, 689–690
effective stress intensity factors, 689–690
specimen with chevron notches, 686
specimen with Knoop indentation, 687–688
of Zr(Hf)B2 based UHTCs, 286–287
of ZrO2-based ceramics, 766–767
Frictional properties, of carbon—carbon composites, 192–193
Fuel cells, 103–105, 904
types and features of, 104t
Fullerene—Nafion composites, 206
Fullerenes, 41
-based proton conductor, 205
-based MEA without Nafion, 208
Fullerene
treatment method, 1206
Functional ceramics, heat capacity study of, 447–454
dielectric crystals
giant particle-size effect on phase transition in, thermodynamic studies of, 451–452
relaxors
ferroelectric nanoregion formation and growth, thermodynamic properties of, 448–451
Functionally graded materials (FGMs), 1179–1188
applications of, 1171–1173, 1184–1186
defined, 1179
fabrication of, 1181–1184
future of, 1186
history of, 1179–1180
large-size FGMs, development of, 1170–1174
fine-WC/Co hard-alloy FGMs, 1170–1173
WC/CoNi FGMs, 1171–1173
ZrO2(3Y)/stainless steel large-size FGMs, 1170
natural materials with graded structures, 1180–1181
synthesized, by SPS sintering, 1165
Functional prototype, 491
Fused deposition modeling (FDM), 501–503
advantages of, 502–503
applications of, 502
history of, 501
limitations of, 503
principle of, 501–502
Fusion reactors, advanced materials for, 366–367

G
Gas permeability, 1141–1142
Gas-phase carbonization, 33t, 36–41
carbon blacks, 37, 37f–38f
carbon nanofibers, 39–40, 41f
carbon nanotubes, 40
diamond, 40–41
diamond-like carbons, 40–41
fullerenes, 41
pyrolytic carbons, 37–39, 38f–39f
vapor-grown carbon fibers, 39, 40f
Gas-phase process, 240
Gas-pressure sintered Si3N4, 247
properties of, 249
Gas-pressure sintering (GPS), 246–247, 263
Gas storage, activated carbon fiber applications in, 163–164
Gas turbine high-temperature reactor 300-cogeneration (GTHTR-300C), 114
Gas turbines, heat-resistant coating technology for, see heat-resistant coating technology for, gas turbines
Gas (vapor) phase, ceramic fine powders from, 866–867
Gel casting, 923
Geometrically necessary dislocations (GNDs), 292–294
Glass-ceramics, 371–382
apatite-containing, 376–377
applications in energy technology, 379
canastite, 376
electrical and magnetic properties of, 378–379
cordierite, 378–379
ferrite, 379
mica, 379
perovskite-type glass-ceramics, 379
enstatite, 375–376
leucite-containing, 377
lithium disilicate, 376–377
mica platelets, curved, 375
mica platelets, straight, 375
with minimal thermal expansion, 372–373
β-quartz solid-solution glass-ceramics, 372–373
β-spodumene solid-solution glass-ceramics, 372–373
with special optical properties, 373–375
Cr-doped multilite, 374–375
lithium metasilicate, 374
oxyfluoride, 374–375
β-quartz solid-solution glass-ceramics, 373–374
β-spodumene solid-solution glass-ceramics, 373–374
Zn-spinel, 374–375
with two crystal phases, 378
ZrO2-containing, 377–378
Glass industry, carbon
Glass morphol
Grain boundary phase control, 663–665
Grain boundary sliding, 766–767
Grain morphology control, 657
Graphitic nanoribbons (GNRs), 61–62
ZrO2-containing, 377–378
with two crystal phases, 378
heat resistance, 664–665
fracture resistance, 663–664
composites for sensor devices, 81
-based field-effect transistors, 80
synthesis of, 68–71
from carbon nanotubes as precursor, 70–71, 70f
chemical routes to, 69–70
by chemical vapor deposition, 68–69, 69f
zigzag, electronic properties of, 62
Graphitization, 35–36
Gymnosperms, 529
H
Hafnium, atomic radius and electronegativity of, 229t
Half-polished fiber coupler, pumping demonstration using, 408–411
Hamaker’s constant, 912
Hardness of ceramics, 609, 613–614
biomorphous, 547
Head-up-display (HUD) system, 894
Heat resistance, 664–665
Heat-resistant coating technology, for gas turbines, 789–806
development of, 792–794
evaluation technology, 799–805
standardization of evaluation test methods, 801–805
trends in development of coating evaluation tests, 799–801
heat-resistant superalloys, development of, 790–791
thermal barrier coatings, 795–796
cross-sectional view of, 795–796
damage modes observed in, 796–799
development of, 794–796
effect of, 794–795
Hexameric silsesquixazane, 1036
Hexamethylsiloxane (HMDS), performance under hydrothermal conditions, 334
Hexamethylphosphoramide (HMPA), 929
High-angle annular dark-field STEM (HAADF-STEM), 9–10, 10f
ceramics, application of, 15–16
rare earth element-doped alumina ceramics, grain boundary of, 15–16, 15f
High-capacitance MLCCs with nickel internal electrodes, 415–417
High Cr–Mo–Ni base alloys, 820
High Cr–Ni–Fe base alloys, 820
High-order Laue zone (HOLZ), 14–15
Highly-oriented pyrolytic graphite (HOPG), 39
High-performance MRI contrast agent, 578
High pressure infiltration and carbonization (HIP) processes, 176–177
High-resolution electron microscopy (HREM), 7–8
to ceramics, application of, 13–14, 13f–14f
High-resolution transmission electron microscopy (HR-TEM), 34, 39
High Si–Cr–Ni Cr–Cr alloy, 820
High-strain-rate superplasticity, 767
High-temperature corrosion-resistant ceramics, 807–836
Cr2O3/Al2O3 cermet coating, 822–825
deterioration mechanisms of, 825–833
ceramic spray coatings, 833
metallic materials forming protective oxides film, 825–833
metal spray coatings, 833
TiO2–625 cermet HVOF coating, 821
waterwall tubes, corrosion-resistant metal coatings for, 817
ZrO2–Ni–base alloy plasma-jet coating, 822–825
High-temperature engineering test reactor (HTTR), block-type fuel element of, 115f
components arrangement in, 114–119
components design criteria for, 116, 117f
graphite structures in, 114–119
irradiation effects on graphite material properties, 116–119
dimensional change, 117–118
longitudinal elastic modulus, 118
steady-state irradiation creep coefficient, 119
strength, 119
thermal conductivity, 118
thermal expansion, mean coefficient of, 118
High-temperature fasteners, carbon–carbon composite applications in, 196
High temperature gas-cooled reactors (HTGRs), 113–119
advanced materials for, 362–365
development of, 113–114
High-temperature reactor-steam-cooled reactor (HTGR), 113–119
Advanced materials for, 362–365
development of, 113–114
High-temperature reactor-steam-cooled reactor (HT-R-M), 114
High thermal conductivity silicon nitride ceramics, 245–248
structure of, 246f
High-velocity oxygen fuel (HVOF), 793
Homogeneous precipitation method, 869
Honeycomb-type 3D battery, 343–346
Hot isostatic pressing (HIP), 854–855, 1203–1212
atmosphere control treatment, 1205–1210
full-pressure method, 1206–1207
partial-pressure method, 1208
powder embedding method, 1208–1210
equipment, configuration of, 1203, 1204f
future perspectives of, 1210–1211
preheating equipment, 1205
rapid cooling technology, 1204–1205, 1205f
technological development in, 1203–1204
Hot-pressing (HP) sintering, 248
characteristics of, 1153
isostatic, see Hot isostatic pressing
Zr(HF)B2 based UHTCs, 273
conditions of, 274
HPGL (Hewlett-Packard Graphics Language) format, 519
Human bone, 1181
Hybrid nanomaterials, supercritical hydrothermal synthesis of, 972–973
bimolecules and, 973
needs and breakthrough point, 972
superhybrid materials, 972–973
Hydridopolysilazane (HPZ), 1039
Hydrogen economy, 103
Hydrogen-production technologies, using amorphous silica membranes, 331–342
membrane performance, under hydrothermal conditions, 334
methane steam reforming, 339–340
organic hydrides, dehydrogenating, 335–337
silica membranes, pore size control of, 332–334
Hydrogen storage, by high-pressure adsorption, 164
Hydrogen sulfide, decomposing, 337
Hydrolysis method, 869
Hydrophilic coating, 896
Hydrolytic coating, 896
Hydrophobic silica aerogel, 899
Hydrothermal powders
characteristics of, 877–879
syntheses of, 871–879
autoclaves or bombs, 872–874
batch system and continuous system, 874
heating methods, 875
hydrothermal reactions, 875–877
starting materials, 875
Hydrothermal reaction, functional materials creation by, 1196–1201
allophane with micropores/mesopores, solidification of, 1200–1201
zeolite with micropores/mesopores, solidification of, 1197–1200
nonstick snail shell, 1195–1196
cleaning with water, 1196
system, 1190–1191
Intermediate transition (IM) carbides, 226
Intermolecular Inc., 1122
Interstitial carbides (IC), 226
Investment casting, 520
Ion implantation process, 132
Iron, atomic radius and electronegativity of, 229
Iron-pnictide superconductors, 477–483
advances in materials, 480–482
current status, 482
discovery of, 477–480
perspective, 482–483
IR-shielding film, 895
ISO (International Organization for Standardization) Standards
creep fracture measurements, 652
fatigue/slow crack growth measurements, 646–647
fracture strength measurements, 640–641
Iso-electric point (IEP), 912
Isothermal technique, 176
Isotropic graphite, for electric discharge machining, 135–142
Isotropic porous structure, 666–668
Isovite, 227
Iwatani–Ishibashi method, 869–870
J
Jedwabite, 227
JENDL, 354
JISC (Japanese Industrial Standards Committee) Standards
creep fracture measurements, 652
fatigue/slow crack growth measurements, 647
fracture strength measurements, 642
K
K-2, 249
K-2I, 249
K-2IO3, 249
K2O, 249
K3S2O8, 249
K2SO4, 249
KBr, 249
Knob indentation, specimens with, 687–688
Kramers–Kronig relation, 396–397
Kumada rearrangement, 1029, 1029f, 1052
L
LaFePtO, 478–480
Lamb waves, 856–857
Laminates optimal with external compressive stresses, 744–745
with internal compressive stresses, 745–746
with strong interfaces, 741–742
with weak interfaces, 740–741
Large-size metal and ceramic compact, fabrication of, 1165–1166
Laser flush method, 618
Lasers, 893–894
LaTiO, electromagnetic properties of, 478
Layer-by-layer (LbL) assembly cerasomes hierarchical assembly on substrates via, 1013–1015
nanoactive coatings via water-soluble precursor, 1015–1016
Layered ceramics, 733–752
design guidelines to strength/toughness optimization, 742–746
fracture mechanics approach, 742–744
new design concepts search, 746
optimal laminates with external compressive stresses, 744–745
optimal laminates with internal compressive stresses, 745–746
mechanical behavior of, 739–742
laminates with strong interfaces, 741–742
laminates with weak interfaces, 740–741
outlook, 746–747
residual stresses in, 735–739
analytical solution, 735–736
design limitations, 736–737
experimental determination of, 737–739
LCD panels, driving backplane of, 475–476
Lead-free piezoelectric ceramics, 429–445
BaTiO3-based ceramics, 430
(Bi1/2K1/2)TiO3-based system, 440–443
(Bi1/2Na1/2)TiO3-based ceramics, 433–435
(Bi1/2Na1/2)TiO3–(Bi1/2K1/2)TiO3–BaTiO3 system, 435–436
(Bi1/2Na1/2)TiO3–(Bi1/2La1/2)TiO3–
(Bi1/2K1/2)TiO3 system, 437–440
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KNbO₃-based system, 430–433</strong></td>
</tr>
<tr>
<td><strong>KNbO₃–NaNbO₂–LiNbO₃ system, 430</strong></td>
</tr>
<tr>
<td><strong>LEAF file, 519</strong></td>
</tr>
<tr>
<td><strong>Lean NOₓ trap (LNT) technology, 594–595</strong></td>
</tr>
<tr>
<td><strong>LENS, 520–521</strong></td>
</tr>
<tr>
<td><strong>Leucite-containing glass-ceramics, 377</strong></td>
</tr>
<tr>
<td><strong>Lifetime prediction expressions, of brittle materials, 676</strong></td>
</tr>
<tr>
<td><strong>LIGA technique, 494</strong></td>
</tr>
<tr>
<td><strong>Lignocellulosic cellular performs, 528–529</strong></td>
</tr>
<tr>
<td><strong>biomorphous ceramics from, 527–531</strong></td>
</tr>
<tr>
<td><strong>hierarchical organization of, 529</strong></td>
</tr>
<tr>
<td><strong>biomorphous carbon, 531–535</strong></td>
</tr>
<tr>
<td><strong>oxide ceramics (cell wall intercalation), 539–541</strong></td>
</tr>
<tr>
<td><strong>reaction infiltration into biocarbon templates, 535–539</strong></td>
</tr>
<tr>
<td><strong>wood performs, 529</strong></td>
</tr>
<tr>
<td><strong>Linear-elastic fracture mechanics, 681–684</strong></td>
</tr>
<tr>
<td><strong>Liquid-based rapid prototyping, 491–492</strong></td>
</tr>
<tr>
<td><strong>Liquid infiltration/pyrolysis (LIP) technique, 176–177</strong></td>
</tr>
<tr>
<td><strong>Liquid phase, ceramic fine powders from, 867–870</strong></td>
</tr>
<tr>
<td><strong>Liquid-phase carbonization, 33t, 45–48</strong></td>
</tr>
<tr>
<td><strong>cokes, 32f, 34f, 45–46</strong></td>
</tr>
<tr>
<td><strong>mesocarbon microbeads, 46, 47f–48f</strong></td>
</tr>
<tr>
<td><strong>polycrystalline graphite blocks, 47, 48, 49f</strong></td>
</tr>
<tr>
<td><strong>Liquid phase-sintered silicon carbide (LPS-SiC) wall flow filters, 599–600</strong></td>
</tr>
<tr>
<td><strong>Lithium disilicate glass-ceramics, 376–377</strong></td>
</tr>
<tr>
<td><strong>Lithium metasilicate glass ceramics, spectral properties of, 374</strong></td>
</tr>
<tr>
<td><strong>Localized surface plasmon resonance (LSPR), 396–397</strong></td>
</tr>
<tr>
<td><strong>Low-carbon economy, 103</strong></td>
</tr>
<tr>
<td><strong>Low-pressure injection molding (LPIM), 494–495</strong></td>
</tr>
<tr>
<td><strong>Low-temperature co-fired ceramics (LTCCs), 733–734, 775</strong></td>
</tr>
<tr>
<td><strong>Lubricants, 921</strong></td>
</tr>
<tr>
<td><strong>Lubricating oil, 94</strong></td>
</tr>
</tbody>
</table>

**M**

| Machinable glass-ceramics, 375                                   |
| mica platelets, curved, 375                                      |
| mica platelets, straight, 375                                    |
| Magnetic drug delivery, 577                                      |
| Magnetic hyperthermia, 577                                        |
| Magneto-optical effect, 392–394                                  |
| Magneto-optical glass, 389–396                                   |
| amorphous ionic solids, magnetism of, 389–392                    |
| amorphous solids, magnetic properties of, 389                    |
| diamagnetic glasses, 396                                        |
| with large rare-earth ions, 394–395                             |
| with large transition metal ions, 395–396                       |
| magneto-optical effect, 392–394                                  |
| Manganese, atomic radius and electronegativity of, 229f           |
| Materials Informatics, 1104–1105                                 |
| Mathematical modeling, for environmentally enhanced crack growth, 761 |
| Maximum packing fraction, 916                                   |
| MCM-48, 212–213, 213t                                           |
| gravimetric capacitance of, 215t                                |
| Mechanical machining versus electric discharge machining, 136–137 |
| Mechanical properties of ceramics, 609–632                       |
| compressive strength, 613–614                                   |
| elastic constant, 610                                           |
| elastic properties, 610                                         |
| fracture toughness, 609                                         |
| hardness, 609, 613–614                                          |
| plasticity, 613–614                                             |
| specific fracture energy, 609                                    |
| strength, 622–625                                               |
| degradation with time, 625–628                                   |
| tensile strength, 610                                           |
| thermal diffusion, 610, 617–618                                  |
| thermal strains, 614–616                                        |
| thermal stresses, 616–617                                       |
| toughness, 619–622                                              |
| yield strength, 613–614                                         |
| Mechanical seals, 309–314                                       |
| properties of, 310f                                              |
| Melting gels, 902                                               |
| Membrane performance, under hydrothermal conditions, 334         |
| Membrane reactors, performance prediction of, 335                |
| MEMS devices, application to, 855–857                           |
| Menhaden fish oil, 918–919                                      |
| Mesocarbon microbeads (MCMBs), 34–35, 46, 47f–48f               |
| Metal/ceramic joints, 775–788                                    |
| interface chemistries, 776–778                                   |
| joining process, 782–787                                        |
| active metal brazing and Mo–Mn method, 782–783                   |
| eutectic liquid bonding, 783–784                                |
| field-assisted bonding, 786–787                                  |
| room-temperature bonding, 784                                   |
| soldering, 785–786                                              |
| SQ process, 784–785                                            |
| physical contact at interface, 779                              |
| surface roughness and damage of bond face of ceramics, 780       |
| thermal stress, 780–782                                         |
| Metal bearings, properties of, 304t                             |
| Metal-casting tools, Si₃N₄ applications in, 266                   |
| Metal diesel particulate filters, 601                           |
| Metal/metal brazing systems, 776                                 |
| Metal oxide nanoparticle synthesis, 956                          |
| Metal separators, carbon coating for, 201                       |
| Metal spray coatings, 833                                       |
| 3-Methacryloxypropyltrimethoxysilane (MOPS), 406                |
| Methane steam reforming, 339–340                                |
| Methanides, 226                                                  |
| Methylchlorosilylaminodiacloroborane (MADB), 1042                |
| (Methylchlorosilyl-ethyl)chloroborane, 1043                      |
| Methyltrichlorosilane, 1035                                      |
| Methyltriethoxysilane (MTES), 894, 900, 902                      |
| Methyltrimethoxysilane (MTMS), 900, 902–903                      |
| Mica                                                               |
| electrical and magnetic properties of, 379                      |
| platelets curved, 375                                           |
| straight, 375                                                   |
| Microcapsule method, 869                                        |
| Microcrack zone, 699                                            |
| Microcrystalline gibbsite, reaction of, 930                     |
| Microfabrication (MFCX) processes, 924                           |
| Microfibrill dominance theory, 531–532                          |
| Micrograins, 765                                                |
| Micromechanical cracking, in carbon–carbon composites, 178       |
| Microreactor, for supercritical hydrothermal synthesis, 953–956, 954f |
| Microscrotoeleolithography (µSLA), 513–516                       |
| advantages of, 516                                              |
| applications of, 514–516                                        |
| history of, 513                                                 |
| limitations of, 516                                             |
| materials for, 514                                              |
| Microstructural control, 657–674                                 |
| fibrous grain alignment, 659–663                                |
| coarse grain control, via seeding and tape-casting techniques, 659–661 |
| fine grain control via forging techniques, 661–663               |
| grain boundary phase control, 663–665                           |
| fracture resistance, 663–664                                     |
| heat resistance, 664–665                                         |
| grain morphology control, 657–659                               |
| porous structure control, 666–671                               |
| anisotropic porous structure with fibrous grain alignment, 668–669 |
| isotropic porous structure, 666–668                             |
| thermal shock resistances, 669–671                               |
| Microtextures, in carbon materials, 53–56, 54f–55f               |
| Micro-total analysis systems (µ-TASs), 855–856                    |
| Minimachine analysis, in carbon–carbon composites, 178           |
| Minor actinide transmutation, advanced materials for, 365–366    |
| MJM technique, 496                                              |
| MJS technique, 494–495                                          |
| ModelMaker (MM), 505                                            |
| Molded carbon resin separators, 200                              |
| Molds defined, 135                                               |
| kinds of, 136t                                                  |
| Molten carbonate fuel cell (MCFC), types and features of, 104t   |
| Molten salt solidification of, 1007                             |
| Molybdenum, atomic radius and electronegativity of, 229f         |
| Mo–Mn method, 782–783                                           |
| Mono-component adsorption, of hydroxyapatite, 563–565            |
Monocotyledons, 529
Monocrystalline silicon, manufacturing process for, 128–129
Monochlorosilanes, ammonolysis of, 1033
Monolithic ceramics, applications of, 1080–1083
Monomers, 506
Mullite (al silicate) wall flow filters, 600
Multicomponent adsorption, of hydroxyapatite, 565–567
Multilayered ceramic capacitors (MLCCs), 415–428
BaTiO3 ceramics, size effect of, 421–423
high-capacitance MLCCs with nickel internal electrodes, 415–417
lifetime in HALT, 423–425
nonlinear dielectricity of, 417–418
reliability of, 423–425
technology, computer simulation and prospective of, 425–426
Multilayered graphene nanoribbons, 63–64

N
Nafion®, 199–200
Nano-Al2O3/Yb2O3, SPS-consolidation of, 1162–1164
Nanodiamond, 89–102
as ceramics, 89–90
Nanohybrid coatings via water-soluble precursor layer-by-layer assembly, 1015–1016
Nanoparticles synthesis, by supercritical hydrothermal synthesis, 956–957
commercialization of, 956–957
metal oxide nanoparticle, 956
Nanopores, formation of, 889–890
noncontrolled pore formation, 889
porous materials with ordered mesopores, controlled formation of, 889–890
Nanopowders, 549–550
Nanoreinforced carbon–carbon composites, 177
Nanoshell (NS), 105–106
carbons, 105, 106f
electrochemical activity of, 106
ORR activities improvements in, 108
Nanosingle-crystalline diamond (NSCD) applications of, 94–100
catalyst carriers, 99
fluorescent bioprobes, 97–98
nanospacer lubrication in water, 94–97
particles as platform for drug delivery system, 98–99
single-crystal diamond film by CVD, seed for, 98
strengthening additive in compound electroplating, 99–100
characteristic values of, 95f
color of, 94
detonation of, 93

discovery of, 90–91
interparticle interactions of, 91–92
particles shape, 93
purification of, 93–94
salvation of, 92–93
surface structure, 93
Nanospace lubrication, in water, 94–97
Nanotextures, 51–53
in carbon materials, 33–35, 34f, 35f, 51–53, 52f, 53f
Near edge X-ray absorption fine structure (NEXAFS)
graphitic nanoribbons edge state detection by, 73–74
Neck formation, by SPS sintering, 1154–1156, 1156f
Neutron irradiation effects with matter, 353–354
New Energy Development Organization (NEDO), 103–104
Nickel, atomic radius and electronegativity of, 229t
Niobium, atomic radius and electronegativity of, 229t
Niobocarbide, 227t
Nitrogen-doped graphene nanoribbons, synthesis of, 73
Nitrogen-enriched carbons, 213–216
gravimetric capacitance of, 215t
Nitroglycol, 929
Nonhydrolytic sol–gel reaction, 889
Nonlinear dielectricity, of multi-layered ceramic capacitors, 417–418
Nonlinear optical glass, 384–388
fibers, second-harmonic generation in, 386–387
fundamentals of, 385
glasses after poling, second-harmonic generation in, 387–388
effect in glass materials, 386
second-order, 385–386
thermal poling, phenomenon related to, 388
Non-Pt cathode catalysts, history of, 105f
Non-sp2 carbon defects, in graphitic nanoribbons, 67
Nonstick snail shell, 1195–1196
cleaning with water, 1196
Novel carbonization processes, 48–51
carbon aerogels, 51
polymer blend method, 49–50, 51f
polymers, defluorination of, 50–51
template methods, 48–49, 50f, 50t, 51f
Nuclear applications, advanced ceramics for, 353–368
fusion reactors, 366–367
high-temperature gas-cooled reactors, 362–365
minor actinide transmutation, 365–366
neutron irradiation effects with matter, 353–354
radiation damage, 354–362
Nuclear graphite, 113–124
Nuclear hydrogen development and demonstration (NHDD) project, 114
Nucleation and growth, in ceramic processing, 1004

O
Objective lens, 4–6
OLED panels, driving backplane of, 475–476
Oligomers, 506
Oligoethylene, 1030, 1030f
Open-cell ceramic foam filters, 600–601
Optical isolator, 393f
Optical memory glass based on spectral hole burning, 398
Ordered–mesostructured organic–inorganic hybrid, 888
Organic compounds, pyrolysis of, 239
Organic hydrides, dehydrogenating, 335–337
Organic–inorganic hybrid, formation of, 887–888
alkylalkoxysilanes, use of, 888
hybrids, synthesis of containing characteristic polymers, 888
containing characteristic silicate anion, 888
ordered–mesostructured organic–inorganic hybrid, 888
polydimethylsilane, use of, 888
polysilazane, use of, 888
radical polymerization of alkyltrimethoxysilane followed by hydrolysis, 888
Organic–inorganic hybrid hollow capsules preparation, by colloid templating, 1016–1021
Organic–inorganic hybrid materials preparation, through supramolecular assembly, 1011–1024
cerasome formation, 1011–1015
via amphiphilic organoalkoxysilane self-assembly, 1011–1013, 1012f
hierarchical assembly on substrates via layer-by-layer assembly, 1013–1015
via sol–gel reaction, 1011–1013
hollow capsules preparation, by colloid templating, 1016–1021
nanohybrid coatings via water-soluble precursor layer-by-layer assembly, 1015–1016
Organic–inorganic hybrid nanoparticles self-assembly of, 965–971
fixing on substrate, 967–968
nanoparticle-organic framework, 968–971
2D self-assembly on substrate, 965–967
supercritical hydrothermal synthesis of, 957–960
bonding and surface nature, 962–964
dispersion in solvents, 964–965
mechanism, 960–962
Orthogonalized linear combinations of atomic orbitals (OLCAO) method, 19, 20f
Osmotic consolidation, 922
Oxidation behavior of Zr(Hf)B₂ based UHTCs, 275–277
in ceramic processing, 1004
protection of carbon-carbon composites, 193–194, 193t, 194f
Oxide ceramics (cell wall intercalation), 539–541
Oxides TETs, history of, 468–469
Oxyfluoride glass ceramics, spectral properties of, 374–375
Oxygen balance (OB), 929
Oxygen reduction reaction (ORR), 104–107
activities improvements in nanoshell carbons, 108
activity and structure of carbons, 109

P
PAN-based carbon fibers, properties of, 172t
Partial-pressure atmosphere control HIP treatment method, 1208
Partial sintering, 1133–1135
Perhydropolysilazane (PHPS), 888
Perovskite-structured piezoelectric ceramics, 986
Perhydropolysilazane (PHPS), 888
Polyacrylonitrile (PAN) as electrocatalyst, 205
as collectors, 205
characteristics of, 199
operating principles, 199–200
as proton conductor, 205–208
as separators, 200–201
stack, construction of, 201f
Polymer impregnation and pyrolysis (PIP) process, 1089–1090
Polymer infiltration and pyrolysis (PIP) method, 366
Polymerization, sonochemical, 1004–1007
Polymer, defluorination of, 50–51
Polymer-to-ceramic transformation, 1050–1065
silicon carbide, thermolysis of precursors to, 1051–1052
silicon carbide/nitride ceramics, thermolysis of precursors to, 1053–1059
polysilazanes, thermolysis of, 1053–1057
polysilylcarbodiimides, thermolysis of, 1057–1059
Polymethylmethacrylate (PMMA), 1136
Polymethylsilsequioxane (PMS), 1063, 1064f–1065f
Poly(methyl-silylenemethylene), 1051
Poly(methylvinylsilazane) (PMVS), 1055–1056, 1055f, 1059
Poly(p-fluorophenylthiophene) (PFPT), 218–221, 218f
Polysilanes, 1027
Polysilazanes
carbon-free, 1032
thermolysis of, 1053–1057
use of, 888
Polysiloxanes
chemical modification of, 1050f
synthesis of, 1049f
Poly(silsequioxane), 1035
Poly(silylcarbodiimide), 1037–1038, 1037f–1038f
Si–B–C–N precursors based on, 1045–1046
synthesis of, 1046f
thermolysis of, 1057–1059, 1058f
Poly(silylene), see Polysilanes
Poly(vinylsilazane) (PVs), crosslinking reactions of, 1034–1035, 1034f
Poly(vinyl silsequioxane) (PNS), 1055–1056, 1056f
Porous ceramic materials, 1131–1148
classification of, 1132–1133, 1132f
direct foaming, 1140–1141
fabrication of, 1132–1133, 1132f
gas permeability, 1141–1142
partial sintering, 1133–1135
replica templates, 1139
sacrificial fugitives, 1135–1139
Porous materials formation, by SPS sintering, 1169–1170
Porous SiAlON ceramics, fabrication of, 263–264
Porous structure control, 666–671
anisotropic porous structure with fibrous grain alignment, 668–669
isotropic porous structure, 666–668
thermal shock resistances, 669–671
Post-fracture tensile tests, bridging stresses from, 702–703
Powder(s)
- based rapid prototyping, 492
  characteristics of, 864–866
  real powders, 864–865
  ideal powders, 865–866
  configuration process, 1183
  embedding method, 1208–1210
  -graded seal, 782
  processing, 911, see also Ceramic powders, for advanced ceramics
  Precipitation, in ceramic processing, 1003–1004
  Precursor-derived ceramics (PDCs), 1025–1102
  applications of, 1080–1092
  ceramic coatings, 1083–1085
  ceramic fibers, 1086–1089
  ceramic membranes, 1085–1086
  comments, 1080
  fiber-reinforced ceramics, 1089–1092
  monolithic ceramics, 1080–1083
  comments, 1026
  high temperature properties of, 1065–1080
  comments, 1065–1066
  Si–C–N ceramics, 1068–1069
  Si–B–C–N ceramics, 1069–1078
  Si–M–O–C ceramics, 1078–1080
  Si–O–C ceramics, 1078–1080
  thermodynamic aspects, 1066–1068
  polymer-to-ceramic transformation, 1050–1065
  Si–B–C–N ceramics, thermolysis of precursors to, 1059–1061
  silicon carbide, thermolysis of precursors to, 1051–1052
  silicon carbide/nitride ceramics, thermolysis of precursors to, 1053–1059
  Si–M–O–C ceramics, thermolysis of precursors to, 1062–1065
  Si–O–C ceramics, thermolysis of precursors to, 1062–1065
  preparation of, 1025–1026
  quaternary Si–B–C–N ceramics, precursors to, 1038–1049
  ammonolysis, 1042–1044
  attempts toward industrial synthesis, 1046–1049
  based on poly(silylcarbodiimides), 1045–1046
  borazine-based Si–B–C–N precursors, 1038–1042
  from boron-modified chlorosilanes, 1042–1044
  from boron-modified hydridosilanes, 1044–1045
  quaternary Si–M–O–C ceramics, precursors to, 1049–1050
  silicon carbide, precursors to, 1026–1031
  silicon nitride, precursors to, 1031–1032
  ternary Si–C–N ceramics, precursors to, 1032–1038
  ternary Si–O–C, precursors to, 1049–1050
  Preheating HIP equipment, 1205
  Pressure, 924–925
  Pressureless sintering (PS) technique
  Zr(Hf)B2-based UHTCs, 270–273
  conditions of, 271t
  surface oxides, removal of, 271–273
  Proof-of-principle prototype, 491
  Protein adlayers, Voigt-based viscoelastic property of, 567–568
  Protein adsorption studies, on hydroxyapatite, 563–568
  mono-component adsorption, 563–565
  multicomponent adsorption, 565–567
  protein adlayers, Voigt-based viscoelastic property of, 567–568
  Proton conductor, polymer electrolyte fuel cells as, 205–208
  Proton-exchange membrane fuel cell (PEMFC), 103–104
  principle of, 104t
  types and features of, 104t
  Proton-exchange membrane fuel cells (PEMFCs), 904
  Prototype, 490–491
  defined, 490
  types of, 490–491
  Pseudocapacitance, 217
  Pulsed laser deposition (PLD), 1105
  Pulse electric current sintering (PECS), 1133
  Pyrolysis, 31–33
  Pyrolytic carbon matrix, mechanical properties of, 185–187
  Pyrolytic carbons, 37–39, 38f–39f

Q
Quartz crystal microbalance with dissipation (QCM-D), 559–568
bioceramics–cells interfacial phenomena, in situ monitoring of, 557–576
fundamental principles of, 559–561
hydroxyapatite sensor surface fabrication using, see Hydroxyapatite
β-Quartz solid-solution glass-ceramics minimal thermal expansion of, 373
spectral properties of, 373–374
Quaternary Si–B–C–N ceramics, precursors to, 1038–1049
ammonolysis, 1042–1044
attempts toward industrial synthesis, 1046–1049
based on poly(silylcarbodiimides), 1045–1046
borazine-based Si–B–C–N precursors, 1038–1042
from boron-modified chlorosilanes, 1042–1044
from boron-modified hydridosilanes, 1044–1045
quaternary Si–M–O–C ceramics, precursors to, 1049–1050
Quenched gas injection, TiO2 particle size control via, 986–988
Quotingite, 227t

R
Radiation damage, of advanced ceramics, 354–362
Raman characterization of graphene/graphitic nanoribbons, 74–77
Raman spectroscopy, bridging stresses from, 703
Rapid cooling technology, 1204–1205, 1205f
Rapid manufacturing, 521
Rapid micro product development (RMPD) technique, 496
Rapid prototyping (RP), 489–524
advantages of, 492–494
applications of, 519–522
future developments of, 521–522
rapid manufacturing, 521
rapid tooling, 520–521
classification of, 491–492
data file formats, 516–519
HPGL format, 519
IGES format, 518–519
LEAF file, 519
SLC file, 519
STL file format, 516–518
tessellated model, 518
versus full production model, 493f
history of, 489–490
problems associates with, 492–494
solid freeform fabrication, 494–498
file format, 496–498
finishing processes of, 498
master model fabrication, 496
object setup, 498
process chain of, 495–496
process of, 494–495
3D additive manufacturing techniques, 498–516
3D printing, 503–504
fused deposition modeling, 501–503
inkjet printing, 505–507
microstereolithography, 513–516
selective laser sintering, 499–501, 499t
stereolithography, 507–513
RapidTool, 520
Rapid tooling, 520–521
direct, 520–521
indirect, 520
Rare earth aluminum garnet (REAG), 940
Rare earth gallium garnets (REGGs), 940
Rayon-based precursors, 151–152
R-curve behavior, 691–704
determination methods for, 693–695
controlled bending tests, 693
stable fracture test under opposite roller loading, 693–694
from strength measurements on Knopp-damaged specimens, 694–695
experimental results, 691
$K_I_{0s}$ determination of, 695–696
S

Sacrificial fugitives, 1135–1139
Salt-like carbides (SCs), 225–226
Sand casting, 520
Sandblasting, 521
Sandwich composites, 187
SBA-15, 212–213, 213t
Gravimetric capacitance of, 215t
Scanning transmission electron microscope (STEM), 3, 8–10, 9f
ABF-STEM, 10, 10f
HAAADF-STEM, 9–10, 10f
to ceramics, application of, 15–16, 15f
Superlattice thin film, observation of, 16, 16f–17f
transmitted electron, detection of, 9
Scanning tunneling microscopy (STM), 28–29
Graphitic nanoribbons edge state detection by, 73–74, 74f
Scanning tunneling spectroscopy (STS), 912
Graphitic nanoribbons edge state detection by, 73–74
Scotch tape method, 61
Secondary ion mass spectrometry (SIMS), 1118
Second-harmonic generation in fibers, 386–387
in glasses after poing, 387–388
Seeding technique, coarse grain control via, 659–661
Selective catalytic NOx reduction (SCR) technology, 594
Selective laser sintering (SLS), 494–495, 499–501, 499t
advantages of, 501
applications of, 500–501
history of, 499
limitations of, 501
materials for, 500
principle of, 499–500
Self-controlled induction heating, mediators for, 578–579
Self-propagating high temperature synthesis (SHS), 1184
Sensor devices, graphene composites for, 81
Sentinel lymph node mapping, for monitoring cancer metastasis, 579–580
Sesquicarbides, 226
α-SiAlON ceramics, development of, 261–262
dense SiAlON nanoceramics, fabrication of, 263
porous SiAlON ceramics, fabrication of, 263–264
spherical hollow powders, synthesis of, 262–263
Si–B–C–N ceramics
high temperature properties of, 1069–1078
quaternary, precursors to, 1038–1049
ammonolysis, 1042–1044
attempts toward industrial synthesis, 1046–1049
based on poly(silylcarbodiimides), 1045–1046
borazine-based Si–B–C–N precursors, 1038–1042
from boron-modified chlorosilanes, 1042–1044
from boron-modified hydridosilanes, 1044–1045
thermolysis of precursors to, 1059–1061, 1060f–1061f
SiC, SPS-consolidated, 1160–1162
Siemens method, 128f
Silica membranes, pore size control of, 332–334
Silicon, atomic radius and electronegativity of, 229f
Silicon alkoxide, alkoxy group in, 886
Silicon carbide, 228, 768
characteristics of, 233
corrosion stability of, 321–324, 321t
environmental applications of, 240–241
for high-temperature use, properties of, 248t
liquid phase sintered, 311–312, 316
polytypes of, 236f, 237f
precursors to, 1026–1031
polyacarosilanes, 1027–1031
polyacarilanes, 1027
radiation damage of, 355–356
Si-infiltrated, 311–312
stoichiometric, precursor synthesis for, 1035f
synthesis of, 235–240
carbothermal reduction, 238–239
gas-phase process, 240
organic compounds, pyrolysis of, 239
sol–gel route, 239–240
solid combustion, 236–238
vapor-phase, 239
thermolysis of precursors to, 1051–1052, 1052f
wall flow filters, 598–600
clay-bonded, 599
liquid phase-sintered, 599–600
mullite, 600
reaction-bonded, 599
recrystallized, 598–599
Silicon carbide/nitride (Si–C–N) ceramics
high temperature properties of, 1068–1069
thermolysis of precursors to, 1053–1059, 1053f
polysilazanes, thermolysis of, 1053–1057, 1054f–1055f, 1056f, 1057f
polysilylcarbodiimides, thermolysis of, 1057–1059, 1058f
Si–B–C–N ceramics, thermolysis of precursors to, 1059–1061, 1051–1052
Si–M–O–C ceramics, thermolysis of precursors to, 1062–1065, 1062f
Si–O–C ceramics, thermolysis of precursors to, 1062–1065, 1062f
Silicon diimide, 1032
Silicon nitride, 767–768
applications of, 264
bearing materials, 302–309
corrosion resistance of, 316–321, 320f
properties of, 304t, 309t
composite containing molybdenum and iron, 252
experimental results and discussion, 252–253
production of, 252
creep of, 652–654
Silicon nitride (Continued)
  fine-grained, with sintering additive, 259–261
gas-pressure sintered, 247
  high-strength, 256–259
  for high-temperature use
development of, 261–262
  properties of, 248
  high thermal conductivity, 245–249
  high-toughness, 256–259
  low-cost fabrication of, 254–256
  low-friction, with carbon fiber, 250–251
  precursors to, 1031–1032
  carbon-free polysilazanes, 1032
  radiation damage of, 356–360
  sintered reaction-bonded, 247
  stoichiometric, precursor synthesis for, 1035f
  superplasticity of, 767–768
  wall flow filters, 600
  Silicon single-crystal ingot, machining process for, 129
  Si–M–O–C ceramics
    high temperature properties of, 1078–1080
    thermolysis of precursors to, 1062–1065
  SIMPLER algorithm, 986–987
  Si3N4–Li2O–SiO2 system, phase diagram of, 249f
  Single-crystal film growth, combinatorial screening of flux materials for,
    1114–1115
  Single-crystal oxide films growth, by flux-mediating epitaxy, 1117–1121
  Single-edge-notch-bend (SENB) test, 724
  Single-edge-precracked-beam (SEPB) method, 621, 724
  Single-edge-V-notched beam (SEVNB) method, 621–622, 661–662, 724
  Single-walled carbon nanotubes (SWCNTs), 212
  super growth method of, 212
  Sintered reaction-bonded Si3N4 (SRBSN), 247
  Si–O–C ceramics
    high temperature properties of, 1078–1080
    ternary, precursors to, 1049–1050
    thermolysis of precursors to, 1062–1065
  Si semiconductor manufacturing, 125–127, 126f, 127f
  carbon materials for, 125–133
  CVD membrane forming process, 130
  dry-etching process, 131–132
  epitaxial wafers, manufacturing process for, 129–130
  ion implantation process, 132
  monocrystalline silicon, manufacturing process for, 128–129
  polycrystalline silicon, manufacturing process for, 128
  silicon single-crystal ingot, machining process for, 129
  thermal diffusion process, 132–133
  SLC file, 519
  Slip avalanches, 290–291
  Slip casting, 922
  Slow coagulation, 915
  Slow crack growth (SCG), testing and evaluation of, 642–650
  Slurry, 911
  SN282, properties of, 249f
  Sodium dodecyl sulfate (SDS), 1004
  Soft sonication, 621, 724
  mediated epitaxy, 1117
  1114
  1115
  photocatalyst film, 896
  transparent conductive film, 897
  comparison of acid catalyst with base function, 900
  catalysis of, 885
  chemical reactions in, 885
  brief history of, 883
  hydrolysis and polycondensation, 885
  gelation in phase-separable solution, 886
  fiber drawing, 890
  elastic aerogel, 901
  nanopores in gel, formation of, 889
  reaction for nonsilica oxides, 887
  thermolysis of precursors to, 1062
  high temperature properties of, 1078
  gas-pressure sintered, 247
  superplasticity of, 767–768
  sintered reaction-bonded, 247
  stoichiometric, precursor synthesis for, 1035f
  Slip casting, 922
  polydimethylsilane, use of, 888
  polysilazane, use of, 888
  radical polymerization of alkyltrimethoxysilane followed by hydrolysis, 888
  melting gels, 902
  migration of components in thin films during processing, 901–902
  reaction for nonsilica oxides, 887
  shapes and microstructures, formation of, 890–892
  shapes of materials produced by, 884f
  sol–gel coating films, 890–892
  adherence, 891
  critical thickness of, 891
  formation of, 890–891
  microstructure formation of, 892
  striations in, 891–892
  thickness of, 891
  sol–gel nanotubes, 903
  sol–gel synthesis in ionic liquid, 902–903
  synthesis in ionic liquid, 902–903
  types of microstructures of materials produced by, 884f
  Solid–gel route, 239–240
  Solid-based rapid prototyping, 492
  Solid combustion synthesis, of carbides, 236–238
  Solid freeform fabrication (SFF), 494–498
  file format, 496–498
  finishing processes of, 498
  master model fabrication, 496
  object setup, 498
  process chain of, 495–496
  process of, 494–495
  Solid oxide fuel cell (SOFC), types and features of, 104f
  Solid phase carbonization, 33t, 41–45
    activated carbons, 42–43, 42f–43f
    carbon fibers, 44, 44t, 45f
    glass-like carbons, 43, 43f
    ceramic fine powders from, 870–871
  Solids loading, 915–916
  Sol(s)
    characterization of, 994
    dispersion, flow properties of, 996
  Solvothermal synthesis of metal oxides, 927–948
  α-alumina, glycothermal synthesis of, 930–932
  foreign ions, occlusion of, 931
  gas–liquid equilibrium, 932
  glycothermal α-alumina, morphology of, 932
  microcrystalline gibbsite, reaction of, 930
  seeds, effects of, 931–932
  thermodynamical consideration, 930–931
  safety consideration, 928–930
  high-pressure experiments, 928
  organic medium, 928–929
  surface acidity of the products, 930
  titanium alkoxide, 932–939
Index

alkyl halide elimination, 935
in carboxylic acids and anhydrides, 938
glycolthermal reaction of, 934—935, 940—942
hydrolysis of, by water, 933—934, 938—939
in inorganic solvent, thermal decomposition of, 932—933
metal nitrates in alcohol, thermal decomposition of, 942
nonhydrolytic sol—gel-like method, 942
precipitated gel, solvothermal crystallization of, 942
rare earth garnets, 939—942
solvothermal crystallization, 939
solvothermal reaction, 939
titanium chloride, reaction of, 935—938

titanium halide in alcohol/acid mixed solvents, 937
Sonochemical darkening, 1004—1005
Sonochemical switching, 1004
Sonochemistry, 1001—1010
apparatus, 1002—1003
sonicator, types of, 1002
dissolution, 1003—1004
molten salt, solidification of, 1007
nucleation and growth, 1004
origin of, 1001—1002
oxidation, 1004
polymerization, 1004—1005
precipitation, 1003—1004
reduction, 1004
sol—gel (alkoxide) process, 1005—1007
sono-electrochemical process, 1007
ultrasonic power, determination of, 1003
Sono-electrochemical process, 1007
Sonogel, 1005—1006
Sonolysis, 1002
Spark plasma sintering (SPS), 259—260, 263, 678, 1149—1178, 1184
Al2O3 blasting nozzle, 3-dimensional complex near-net-shape forming of, 1167—1169
Al-high-Si alloy materials, super-plasticity and strength improvement of, 1167—1169
porous materials, 1169—1170
characteristics of, 1153t
classification of, 1150f
DC pulse generator for, 1157—1158
grain growth, 1162—1164
historical background of, 1149—1151, 1150f—1151f
industrial demands on, 1165—1166, 1174f
large-size metal and ceramic compact, fabrication of, 1165—1166
Si3N4/Al2O3 composite compacts, for homogenizer component, 1166
sputtering target material, 1165—1166
WC (tungsten carbide) aspheric glass lens mold, 1166
large-size FGMs, development of,
1170—1174
fine-WC/Co hard-alloy FGMs, 1170—1173
WC/CoNi FGMs, 1171—1173
ZrO2(3Y)/stainless steel large-size FGMs, 1170
machine layout and installation, 1159, 1160f
materials for, 1152, 1152t
mechanism of, 1152—1154
nano-Al2O3/Yb2O3, consolidation of, 1162—1164
neck formation and densification by, 1154—1156
outlook, 1174—1175
phase transformation, 1162—1164
production systems, 1160
SiC with Al2O3/Yb2O3 consolidated by, 1160—1162
synthesized FGMs by, 1165
system, configuration of, 1152, 1153f
temperature measurement and difference, 1158—1159
titanium halide in alcohol/acid mixed solvents, 937

structural elements of ceramics, 658f
Sub-critical crack growth, 704—711
cyclic fatigue, 710—711
dynamic strength test, 707—709
lifetime evaluation under constant load, 706—707

v—K curves with macroscopic cracks, determination of, 705
Subcritical crack growth (SCCG), 610, 625—629
Sui-An I, 1194—1195
Sui-An II, 1195, 1195f
Superalloys, heat-resistant, development of, 790—791
Super-capacitors
activated carbon fiber applications in,
164—166
carbons for, 211—222
carbon-coated transition-metal carbides, 217—218
carbon xerogels with conductive polymer nano-coating, 218—221
electrode carbons pore structure control, template method for, 212—213
exfoliated carbon fibers, 216—217
nitrogen-enriched carbons, 213—216, 215f
electric double-layer capacitors, 211
Superconducting quantum interference device (SQUID) microscope, 1106f, 1107
Super-critical hydrothermal synthesis, 949—978
apparatus, 953—956
flow reactor apparatus, 953, 953f
micro-reactor, 953—956, 954f
basic principles of, 950—952
of hybrid nanomaterials, 972—973
biomolecules and, 973
needs and breakthrough point, 972
superhybrid materials, 972—973

nanoparticles synthesis by, 956—957
commercialization of, 956—957
metal oxide nanoparticle, 956
of organic—organic hybrid nanoparticles, 957—960
bonding and surface nature, 962—964
dispersion in solvents, 964—965
microscopic mechanism, 960—962
organic—organic hybrid nanoparticles, self-assembly of, 965—971
fixing on substrate, 967—968

nanoparticle–organic framework, 968—971
2D self-assembly on substrate, 965—967
Superheaters
corrosion-resistant alloys for, 818—820
Cr2O3/NiCr cermet coating, 822—825
TiO2—625 cermet HVOF coating, 821
ZrO2/Ni-base alloy plasma-jet coating, 822—825
Superhybrid materials, supercritical hydrothermal synthesis of, 972—973
Superlattice libraries, 1111—1113
Superlattice thin film, STEM observation of,
16, 16f—17f
Superplastic ceramics, 765–772
composites, 767
functional materials, 768
mechanical properties, 765–766
silicon carbide, 768
silicon nitride, 767–768
zirconia, 766–767
Superplastic deformation, 766
Superplasticity, 765
application of, 768–769
defined, 765
mechanism of, 765–766
Superplastic sinter forging, 765
Surface crack-in-flexure (SCF) method,
Tensile tests, 637
Ternary Si
Ternary phase diagrams, 1109
unique features of, 473
as FPD backplane, progress in, 474
mechanical properties, 765
functional materials, 768
composites, 767
silicon nitride, 767

Tape-casting techniques, 923
damage modes observed in, 796–799
development of, 794–796
effect of, 794–795
YSZ on Ni-based superalloy, 840–841, 841f
Thermal conductivity, of carbon–carbon
composites, 188–190
Thermal diffusion of ceramics, 132–133,
617–618
Thermal expansion, of carbon–carbon
composites, 190
Thermal gradient process, 176
Thermal grown oxide (TGO), 797
Thermal mechanical fatigue test (TMF),
800–801
Thermal phase change inkjet, 505
Thermal poling, phenomenon related to, 388
Thermal shock resistances, 669–671
of carbon–carbon composites, 190–192
Thermal strains of ceramics, 614–616
Thermal stresses of ceramics, 616–617,
780–782
Thermogravimetric analysis (TGA), 1051
Thin films
components migration in, during processing,
901–902
Three-dimensionally ordered macroporous
(3D) structure, all-solid-state Li battery with,
346–350
3D additive manufacturing techniques,
498–516
fused deposition modeling, 501–503
inkjet printing, 505–507
microstereolithography, 513–516
selective laser sintering, 499–501, 499t
stereolithography, 507–513
3D printing, see 3D printing
3D printing, 503–504
advantages of, 504
applications of, 504
history of, 503
limitations of, 504
materials for, 504
principle of, 503–504
Threshold strength, 734–735
Threshold stress, 766–767
Thrower–Stone–Wales defects, in graphitic
nanoribbons, 65–67, 66f
Titanium, atomic radius and electronegativity
of, 229t

Tantalcarbide, 227t
Tantalum, atomic radius and electronegativity
of, 229t
TAOS-TFTs
as FPD backplane, progress in, 474–476
novel display structure, 474–475
OLED and LCD panels, 475–476
unique features of, 473–474
Tape-casting techniques, 923
carbothermal reduction, 238–239
gas-phase process, 240
organic compounds, pyrolysis of, 239
sol–gel route, 239–240
solid combustion, 236–238
vapor-phase, 239

Tensile tests, 637–638
Termites, 1192–1193
Ternary phase diagrams, 1109–1111,
1110f–1111f
Ternary Si–C–N ceramics, precursors to,
1032–1038
carbon-containing polysilazanes, 1032–1037
poly(silylcarbodiimide)s, 1037–1038
Ternary Si–O–C, precursors to, 1049–1050
Terrace-microspheres, for laser emission,
406–408
Tessellated model, 518
Tetraethoxysilane (TEOS), 894
Tetrahydrofuran (THF), 929
Tetrachemethoxysilane (TMOS), 333–334, 406,
902–903
performance under hydrothermal conditions,
334

Thermal barrier coating
cross-sectional view of, 795–796
damage modes observed in, 796–799
development of, 794–796
effect of, 794–795
YSZ on Ni-based superalloy, 840–841, 841f
Thermal conductivity, of carbon–carbon
composites, 188–190
Thermal diffusion of ceramics, 132–133,
617–618
Thermal expansion, of carbon–carbon
composites, 190
Thermal gradient process, 176
Thermal grown oxide (TGO), 797
Thermal mechanical fatigue test (TMF),
800–801
Thermal phase change inkjet, 505
Thermal poling, phenomenon related to, 388
Thermal shock resistances, 669–671
of carbon–carbon composites, 190–192
Thermal strains of ceramics, 614–616
Thermal stresses of ceramics, 616–617,
780–782
Thermogravimetric analysis (TGA), 1051
Thin films
components migration in, during processing,
901–902
Three-dimensionally ordered macroporous
(3D) structure, all-solid-state Li battery with,
346–350
3D additive manufacturing techniques,
498–516
fused deposition modeling, 501–503
inkjet printing, 505–507
microstereolithography, 513–516
selective laser sintering, 499–501, 499t
stereolithography, 507–513
3D printing, see 3D printing
3D printing, 503–504
advantages of, 504
applications of, 504
history of, 503
limitations of, 504
materials for, 504
principle of, 503–504
Threshold strength, 734–735
Threshold stress, 766–767
Thrower–Stone–Wales defects, in graphitic
nanoribbons, 65–67, 66f
Titanium, atomic radius and electronegativity
of, 229t

Titanium dioxide (TiO₂)
based ultrafine particles, controlled
radiofrequency thermal plasma
processing of, 980–988
particle synthesis with solid TiN particles,
as precursor, 980–983
size distribution control via quench gas
injection, 986–988
with liquid precursors, 983–985
codoped, room-temperature transparent
ferromagnetism in, 1115–1117
photocatalytic, quantum size effect in, 1115
Titanium chloride, reaction of, 935–938
Titanium halide
in alcohols, 937
in alcohol/acid mixed solvents, 937
elimination, 935
Titanium isopropoxide (TIP), 932
Titanium oxyacetyletacetonate, 932
Titanium oxychloride, 876
Tongbaite, 227t
toughened ceramics, cyclic fatigue of,
647–650
Toughness of ceramics, 619–622
Transition metal carbides, 228–233
crystallochemistry, 230–233
properties of, 233
structure and properties, parameters
influencing, 228–230
Transmission electron microscope (TEM), 3,
7–8
electron diffraction pattern, 7, 7f
electron microscope image, 7, 7f
high-resolution, 7–8, 34, 39
Transparent amorphous oxide semiconductors
(TAOS), 468–477
electron transport in, 470–473
history of, 468–469
material design concept of, with high
mobility, 470
perspective of, 476–477
TAOS-TFTs, see TAOS-TFTs
Transparent conductive film, 897
Transparent electro-active materials (TEAM),
456
Transparent oxides
12CaO 7Al₂O₃, 457–468
crystal structure of, 457
electron-doped C12A7, thin film
fabrication of, 463–466
electron doping, 457–463
electronic structure of, 457
characteristics of, 456f
function cultivation in, 455–488
future challenges to, 484
Trichloroisilyl-amino-dichloroborane
(TADB), 1042
Trimethylchlorosilane (TMCS), 899
Tris(hydridosilylethyl)boranes,
dehydrocoupling reactions of, 1044f
TTDS, synthesis of, 1048–1049, 1049f

Index
Tungsten, atomic radius and electronegativity of, 229t
Turbine components, carbon—carbon composites as, 195
Turbine rotor blades, 791
Turbocharger rotors, Si3N4 applications in, 264–265
Tuttle—Roy bomb, 872f, 874

U
Ultrasound
Ultrasonic metallurgy, 1007
Ultrasonic cavitation, 1001–1002
Ultrasonic casting equipment, 1008f
Ultrasonic metallurgy, 1007
Ultrasound defined, 1001
history of, 1001
Unshaped refractory linings, 815–816
castable refractories, 815–816
plastic refractories, 816
Upconversion glass, 397–398
User experience prototype, 491
UV-curable inkjet, 505
UV shielding film, 895

V
Vacuum casting, 520
Vacuum exhausting system, 1158
Vanadium, atomic radius and electronegativity of, 229t
van der Waals forces, 912
Vapor-grown carbon fibers (VGCFs), 39, 40f
Vapor—liquid—solid (VLS) method, 1114
Vapor-phase axial deposition (VAD), 404
Vapor-phase deposition, of carbides, 239
Very high-temperature reactor (VHTR), 1114
developments, R&Ds for, 120–122
graphite components of, standard for, 121
Vinylidichlorosilanes, 1033, 1043
Vinylsilanes, 1030–1031, 1030f
Vinyltrichlorosilane, 1043
Viscosity, 915–916
Visible-light photocatalyst, 897
Visual prototype, 491

W
Wall flow filters
aluminum titanate, 598
cordierite, 597–598
design and manufacturing of, 596–600
with membranes, 597
silicon carbide, 598–600
silicon carbide, 600
Waste-to-energy boilers, 807–836, 810f
advanced refractory materials, application of, 810–816
cooled furnace wall, 812–815
firebrick linings, 810–811
unshaped refractory linings, 815–816
coating environments of, 808–810, 809f
Water
cooled lining, 812–815
nanospacer lubrication in, 94–97
repellent film, 896
-soluble precursor layer-by-layer assembly, nanohybrid coatings via, 1015–1016
Waterwall tubes, corrosion-resistant metal coatings for, 817
WC/Co/Ni FGMs, industrial applications of, 1171–1173
WC (tungsten carbide) aspheric glass lens mold, 1166
Weibull theory, 610
Weld overlay, 817, 819t
Wetting agents, 921
Wildcat, 1122
Wood performs, 529
characteristics of, 530t
Work of fracture (WOF) test, 724
Wurtz—Fittig type dehalogenation reaction, 1027

X
X-ray diffraction (XRD), 1106f, 1107, 1109, 1112–1113, 1120–1121
X-ray photoelectron spectroscopy (XPS), 107, 214

Y
Y2O3-stabilized tetragonal ZrO2 polycrystals (Y-TZP)
deformation of, 767
superelasticity of, 767
Yarlongite, 227t
Yield strength of ceramics, 613–614
Young’s modulus, 609
Yttria-stabilized zirconia (YSZ)
thermal barrier coating on Ni-based superalloy, 840–841, 841f
Yttrium aluminum garnet (YAG), 940
Yttrium-doped zirconia powders manufactured by hydrothermal homogenous precipitation, 877

Z
Zelite, 213t
with micropores/mesopores, solidification of, 1197–1200
Zigzag graphitic nanoribbons, electronic properties of, 62
half-metallicity on, 64, 64f
magnetism on, 64, 64f
width and length dependence of, 62–63
Zinc oxide, 997–998
Zirconia, 765–767
atomic radius and electronegativity of, 229t
phase transformation effects in, 697–698
powders manufactured by hydrothermal homogenous precipitation, 877
Zn-spinel glass ceramics, spectral properties of, 374–375
Zr(Hf)B2 based ultrahigh temperature ceramics, 267–300
deformation behavior of, 287–294
mechanical properties of, 277–287
elastic modulus and hardness, 277–279, 278t
fracture strength, 279–286, 280t
fracture toughness, 286–287
oxidation behavior of, 275–277
pressureless sintering technique, 270–273, 271t
diboride ceramics, unconventional processing of, 273–275, 276t
hot-pressing sintering, 273, 274t
processing of, 270–275
ZrO2
-bonded ceramics, fracture toughness of, 766–767
-containing glass-ceramics, 377–378 monoclinic, 876
ZrO2(3Y)/stainless steel large-size FGMs, 1170
ZrO2/ Ni-base alloy plasma-jet coating, 822–825