CHAPTER W21

Synthesis and Processing of Materials

W21.1 Synthesis and Processing Procedures

The various procedures used in the synthesis and processing of materials can be grouped into a few general classes. Specific examples of many of these procedures are given in Chapter 21 of the textbook[†] and in this chapter. Important classes of *synthesis* include those that produce materials in bulk form or in forms with reduced dimensionality (e.g., powders, fibers, and thin films or layers and surface coatings). Bulk materials and larger powders often require further processing to produce materials with the final desired shape or form. *Processing* that changes only the form and not the microstructure of a material is not stressed here. Smaller powders, fibers, and thin films are more often prepared in essentially their final form but may still require further processing to achieve the desired microstructure.

Important classes of materials synthesis and processing procedures are listed in Table W21.1. Specific examples discussed here and in the textbook are also indicated.

A wide range of energy sources are used in the synthesis and processing of materials, depending on the specific procedure involved and the products desired. Some important examples are listed in Table W21.2.

W21.2 Heteroepitaxial Growth

Consider the case where atoms of type A, with lattice constant a in the solid state, are deposited on a flat substrate consisting of atoms of type B, with lattice constant b, where $b > a$. Assume that the symmetries of the two crystals are the same. At first the A atoms may form a monolayer in registry with the substrate. As additional layers are deposited, however, the bulk strain energy in A builds up since there is a lattice-mismatch strain given by $(b - a)/a$ [see Eq. (W20.8)]. The strain may be relieved by having misfit dislocations form at the interface or, alternatively, by having the surface of the A crystal warp. These possibilities are illustrated in Fig. W21.1. Misfit dislocations are discussed in Section W20.2.

If the surface warps, an undulating pattern appears that may be observed using such high-resolution instruments as the transmission electron microscope or the atomic force microscope. The condition for warping is that the additional surface energy needed to curve the surface be less than the bulk strain energy relieved by allowing the adsorbate

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

TABLE W21.1 Important Classes of Materials Synthesis and Processing Procedures

- Synthesis of bulk samples
  - Synthesis from the liquid phase
    - Czochralski method for growth of single-crystal Si (Section 21.6)
    - Liquid-phase epitaxy (LPE): GaAs
    - Bridgman method
    - Sol-gel synthesis (Section 21.12)
    - Rapid solidification (Section W21.12)
    - Flux growth of ceramics using oxide fluxes
    - Arc melting of metallic alloys
    - Hydrothermal growth: crystalline quartz, TGS, ADP, KDP
  - Synthesis from solid powders or bulk material
    - Sintering of powders (Section 21.11)
    - Catalysis (Section 21.14)
    - Polymers (Section 21.13 to 21.15 and W21.21 to W21.25)
    - High pressure–high temperature synthesis of diamond crystals
  - Synthesis from the vapor phase
    - Modified Lely process (SiC platelets): PVD (Section W21.17)
- Synthesis of fine particles or powders
  - Grinding (Section 21.11)
  - Plasma spraying
  - Gas condensation: carbon nanotubes (Section 21.15)
  - Nucleation from a saturated liquid phase
- Synthesis of fibers
  - Drawing from the melt: silica fibers
- Synthesis of thin films and surface coatings
  - Synthesis from the vapor phase
  - Chemical vapor deposition (CVD) (Section W21.5)
    - Molecular beam epitaxy (MBE) (Section W21.6)
    - Metal–organic CVD (MOCVD), also known as metal–organic vapor-phase epitaxy (MOVPE)
    - Plasma-enhanced CVD (PECVD) (Section W21.7)
  - Physical vapor deposition (PVD)
    - Sputter deposition (reactive versus nonreactive) (Section W21.3)
    - Ion beam deposition
    - Thermal evaporation (electron beam or hot filament)
    - Thermal spraying
  - Synthesis from the liquid phase
    - Chemical deposition (surface plating via immersion)
    - Electrochemical deposition or electroplating (surface plating via passage of a current through a solution)
  - Synthesis via chemical reactions
    - Reaction between a vapor or a liquid and the surface
      - Thermal oxidation: $Si(s) + O_2(g) \rightarrow SiO_2(s)$ (Section 21.7)
- Processing
  - Annealing
    - Rapid thermal annealing
  - Oxidation
    - a-SiO_2 via thermal oxidation or SIMOX (Section 21.7)
  - Doping
    - Via diffusion or ion implantation

TABLE W21.1 (*Continued*)

Ion implantation (Section W21.3)
For surface modification (e.g., carburizing, nitriding, etc.) (Section W21.13)
Etching (Section W21.8)
Plasma treatments (Section W21.8)
Float-zone purification (Section W21.4)
Lithography (Section W21.8)
Mechanical processing (Section W21.10)
Work hardening

TABLE W21.2 Sources of Energy Used in Synthesis and Processing

Thermal (heating due to contact with hot gases and/or thermal radiation)
Annealing
Rapid thermal processing
Pressure and temperature
Sintering
Shock compression
Plasma (heating due to energy absorbed from accelerated electrons and ions, emitted light, also the direct effects of Joule heating)
Electromagnetic radiation
Laser beams
Electric fields and the kinetic energy of accelerated ions
Sputtering

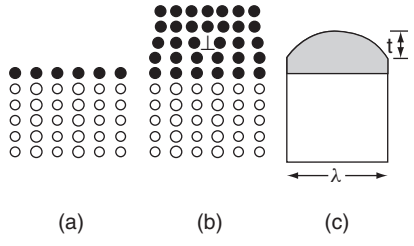

Figure W21.1. Epitaxial growth: (*a*) monolayer of atoms in registry with the substrate; (*b*) formation of a misfit dislocation; (*c*) warping of an adsorbed thick layer of atoms.

to relax its strain. The condition for this may be estimated by assuming a parabolic profile for the warp $y = 4tx(\lambda - x)/\lambda^2$, where t is the height of the warp and λ is the periodicity. If $t \ll \lambda$, the change in surface area is $\Delta A = 8wt^2/3\lambda$ and the volume of the warp is $\Delta V = 4tw\lambda/6$, where w is the surface dimension transverse to the warp. The strain energy relieved is approximately $E\varepsilon^2\Delta V/2$, where the mismatch strain is given by $\varepsilon = (b/a) - 1$, and E is the Young's modulus of the adsorbate. The increase in surface energy is $\sigma\,\Delta A$, where σ is the energy per unit area at the vacuum interface. This leads to the condition

$$\frac{t}{\lambda^2} < \frac{E}{8\sigma}\left(1 - \frac{b}{a}\right)^2 \tag{W21.1}$$

for the development of an undulating surface pattern rather than misfit dislocations.

Recently, a lattice-engineered compliant substrate has been invented which does not cause the adsorbate to develop misfit locations or to warp.[†] This is important, because it permits epitaxial growth of badly mismatched materials without sacrificing crystal quality.

The compliant substrate is a bilayer substrate that is created by having an adsorbed layer bonded to a substrate of the same material but at a twisted angle, as illustrated in Fig. W21.2. The two layers interact, go into partial registry in a domainwise fashion, and form domain walls consisting of screw dislocations, as is shown in Fig. W21.3. This embeds an intrinsic strain into the bilayer substrate. Since the interatomic forces are anharmonic, with the spring constants becoming substantially weaker as the bonds are stretched, the effective spring constants for the substrate are less stiff than they would be for a fully periodic substrate. The compliant substrate is therefore able to deform readily to accommodate an adsorbate with a different lattice constant.

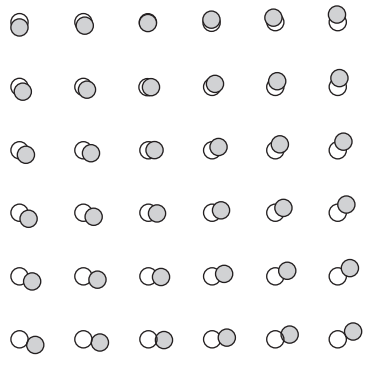

Figure W21.2. Bilayer substrate consisting of a base layer bonded to a twisted overlayer.

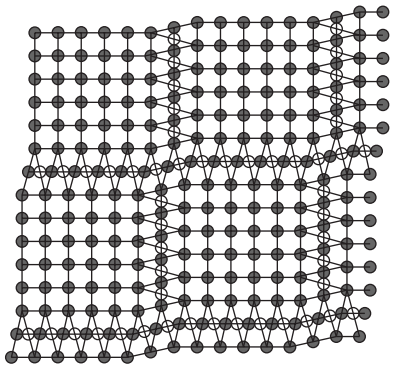

Figure W21.3. Accommodation of the bilayer by the formation of registered domains with domain walls formed by screw dislocations. [Adapted from F. E. Ejeckam et al., *Appl. Phys. Lett.*, **70**, 1685 (1997).]

[†] F. E. Ejeckam et al., *Appl. Phys. Lett.*, **70**, 1685 (1997).

Thin-Film Growth Modes. The *nucleation* and *growth* of thin films on solid surfaces can involve a variety of atomic processes, including adsorption, surface diffusion, and the formation of chemical bonds between adatoms and also between adatoms and atoms of the surface at specific surface sites. These surface processes are discussed in detail in Chapters 19 and W19. Three main modes of thin-film crystal growth are believed to occur at surfaces, at least in those cases in which interdiffusion or chemical reaction between the adsorbing species and the substrate does not lead to the formation of an alloy, chemical compound, or intermetallic compound and in which surface defects such as steps or dislocations do not play a dominant role in the nucleation stage of film growth. Other important modes of thin-film growth include, for example, processes such as the reaction of O_2 with the surface of Si at high temperatures leading to the growth of an amorphous SiO_2 layer or the formation of silicides when metals such as Cu, Au, Ni, Pd, and Pt are deposited on Si.

The three thin-film growth modes to be described here are the *island growth mode*, also known as the *Volmer–Weber mode*, the *layer growth mode*, also known as the *Frank–van der Merwe mode*, and the *layer-plus-island growth mode*, also known as the *Stranski–Krastanov mode*. These growth modes are illustrated schematically in Fig. W21.4. To aid in their description, use will be made of the *surface free energies* σ_A and σ_B of the growing film and the substrate, respectively, as well as the *free energy* σ_{AB} of the A–B *interface*. Examples of thin films growing in each growth mode will also be given. It is, of course, doubtful that concepts such as surface energies can be applied to thin films which nucleate on surfaces as single atoms. In such cases, an atomistic point of view that focuses on individual atomic processes and the potential energies of interaction of adsorbate atoms with the substrate and with each other must be employed. The nucleation of the new phase, whether it be in the form of a cluster or a monolayer, is often a rate-determining step in thin-film growth and, in general, must be understood as resulting from atomic interactions.

Useful reviews of the processes involved in the nucleation and growth of thin films and also of the three growth modes discussed here can be found in Venables et al. (1984) and Venables (1994). Another approach that describes the deposition of thin films from thermal beams and focuses on four different types of atom/molecule-surface interactions has been given by Voorhoeve (1976). A variety of techniques are used to monitor thin-film growth, either in situ or ex situ. These include transmission and

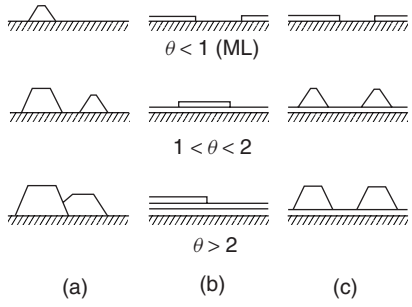

Figure W21.4. Three main thin-film growth modes (ML = monolayer): (*a*) island growth mode, also known as the Volmer–Weber mode; (*b*) layer growth mode, also known as the Frank–van der Merwe mode; (*c*) layer-plus-island growth mode, also known as the Stranski–Krastanov mode.

scanning electron microscopies (TEM and SEM, respectively), reflection high-energy electron diffraction (RHEED), Auger electron spectroscopy (AES), and, more recently, various forms of scanning tunneling microscopy (STM).

Island Growth Mode (Volmer–Weber). In this growth mode, small clusters of adsorbing atoms (or molecules) nucleate on the substrate surface and, if they are stable, continue growing as islands until they coalesce. The islands grow by incorporating atoms that reach the island directly from the vapor phase or by diffusing across the surface. This growth mode is believed to occur when the atoms or molecules of the growing film are more strongly bonded to each other than to the substrate or, in terms of the surface and interface free energies, when $\sigma_A + \sigma_{AB} > \sigma_B$. This inequality is only qualitatively correct since it does not take into account the free energy of A atoms within the bulk of the film when the deposited islands are more than one monolayer thick. Island growth is also expected when the lattice parameters of the film and substrate are very different and when the two lattices cannot be brought into some form of epitaxial alignment by rotation.

Examples of this growth mode include metal films deposited on insulating substrates such as the alkali halides (e.g., NaCl), on the basal plane of graphite and other layered materials, such as MoS_2 and mica, and on insulators such as MgO. By measuring the densities and sizes of stable Au or Ag clusters on the (100) surfaces of alkali halides and comparing with existing theoretical models, researchers have been able to determine that the size of a *stable nucleus* is usually just one metal atom. In addition, values for the exponential prefactors and activation energies associated with desorption and surface diffusion have been determined. Effects associated with cluster mobility at high temperatures can play important roles in this mode of thin-film growth and are therefore often included in the growth models.

Surface reconstructions are common on semiconductor surfaces and can complicate thin-film growth due to the resulting surface anisotropy and possibly to steps with different heights on the same surface. The presence of surface impurities such as carbon or oxygen or of defects such as dislocations can lead to island growth and defective films. In the case of heteroepitaxy [e.g., Si on SiO_2 or on Al_2O_3 (sapphire)], island growth is typically observed, with critical nucleus sizes in the range of one to four atoms.

Layer Growth Mode (Frank–van der Merwe). In this growth mode the adsorbing atoms form a monolayer on the substrate, and additional nucleation and layer growth can occur simultaneously on the substrate and also on the previously deposited layers. The growth in this mode can appear complex, for kinetic reasons (Fig. W21.5), when the thickness of the region in which growth is occurring corresponds to several monolayers. The actual structure of this growth zone or interface transition region will depend

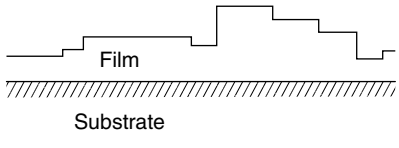

Figure W21.5. Layer growth mode showing nucleation occurring within a multilayer growth zone.

on the relative rates of the nucleation and growth processes. When the nucleation rate is high and monolayer growth is slow, the growth zone will be wider than when nucleation is slow and layer growth is fast. When the growth rate is high enough, deposition will occur monolayer by monolayer (i.e., each monolayer will be essentially completed before nucleation of a new monolayer occurs).

Monolayer-by-monolayer growth can readily be monitored via RHEED, in which case regular oscillations of the RHEED intensity occur with the same period as the monolayer growth. These oscillations are observed when nucleation of each new monolayer occurs on the terraces of existing monolayers but not when growth occurs by step flow (i.e., by the addition of adatoms to existing steps on an off-axis substrate). Decay of the RHEED oscillations can provide evidence for the development of surface roughness due to widening of the growth zone from a single monolayer to several monolayers.

Nucleation will be enhanced at high supersaturations (i.e., high incoming fluxes of growth species) while growth will be enhanced at high temperatures, as long as the temperature is not so high that the growth species tend to be desorbed from the surface before they are incorporated into the growing monolayer. In the limits of very high supersaturation and low temperature, the growing film can be quite disordered and may even be amorphous.

This layer growth mode is believed to occur when the atoms or molecules in each monolayer are more tightly bonded to the substrate than to each other or, in terms of surface and interface free energies, when $\sigma_A + \sigma_{AB} < \sigma_B$. This condition is analogous to that presented in Section W20.1 for the wetting of liquids on surfaces. In some cases the second monolayer to be formed in this growth mode may be less tightly bonded to the first monolayer than the first monolayer is to the substrate.

Examples of this growth mode include inert gases on graphite, some alkali halides, and metal-on-metal [e.g., Ni on Cu(100) or Cu(111) and Ag on W(110)] and semiconductor-on-semiconductor growth systems. Interesting examples include FCC Fe on Ni, Cu, and Au, where the normal BCC crystal structure of α-Fe (ferrite) is not stable due to the strain imposed by the substrate. Misfit dislocations often appear at finite thicknesses in the case of the *heteroepitaxial* growth of metals on metals due to strain in the growing film.

The epitaxial growth of the semiconductors Si, Ge, GaAs, $Ga_{1-x}Al_xAs$, and other compound and alloy semiconductors has been studied widely. In the case of *homoepitaxy* [e.g., Si on Si(100)] the layer growth mode is observed under the ideal conditions of clean substrate surfaces and the high temperatures required for the adatom surface mobility that is necessary to allow crystalline films to be formed. Growth is often carried out on vicinal surfaces that are slightly off-axis ($\approx 1^\circ$ to 4°), in order to have available regular arrays of surface steps at which growth can occur via the layer mode. In this way the difficult initial step involving nucleation of growth on perfectly flat terraces can be avoided.

Layer-Plus-Island Growth Mode (Stranski–Krastanov). As the name suggests, this growth mode is intermediate between the island and the layer growth modes just described in that a strained monolayer (or several monolayers) of growth occurs first, with additional growth occurring in the form of islands nucleating on the growing film. As a result, there is a transition from two- to three-dimensional growth. This growth mode can apparently occur for a variety of reasons: for example, the first monolayer of

the growing film assumes the surface structure of the substrate, which is different from that of the bulk film. This is called *pseudomorphic growth*. In this case layer growth occurs initially when

$$E_{sA'}d + \sigma_{A'} + \sigma_{A'B} < \sigma_B, \tag{W21.2}$$

where A′ refers to the growing film, which is strained when it takes on the structure of the substrate. The term $E_{sA'}d$ represents the *elastic energy* per unit area associated with the strain in the growing film, with $E_{sA'}$ the *strain energy* per unit volume and d the film thickness. As d increases, the left-hand side of Eq. (W21.2) will eventually exceed the right-hand side at a certain *critical thickness*. When this occurs, either misfit dislocations will appear in the film to relieve the strain, as discussed in Section W20.2, or the island growth mode will take over. When island growth that is essentially unstrained takes over, it follows that $\sigma_A + \sigma_{AA'} > \sigma_{A'}$.

The critical nucleus size, ≈ 10 to 100 atoms, for the second, or island, phase of the Stranski–Krastanov growth mode is much larger than in the case of island (Volmer–Weber) growth, where typically a single atom is the critical nucleus. The need for a larger critical nucleus in the Stranski–Krastanov growth mode is likely due to the rather small preference for island growth over layer growth.

Examples of this growth mode include the growth of some metals on metals and on semiconductors [e.g., the Pb/W(110), Au/Mo(110), Ag/W(110), Ag/Si(111), and Ag/Ge(111) systems, among others]. The growth of Ge on Si(100) and Si(111) can also occur via this mode, with a uniformly strained Ge film initially growing to about three monolayers. This is followed by a transition to the growth of three-dimensional Ge nanocrystals on top of the initial strained Ge film, which is often called a *wetting layer*.

W21.3 Processing Using Ion Beams

Ions provide a versatile means for processing solids. They provide a directed source of energy that couples to the ions of a solid via collisions or via excitation of the electrons. Ions play a triple role in the processing of materials. First, an ion beam may be used to sputter material off the surface, thereby cleaning or etching it. Second, ion beams are used to implant ions into surfaces, such as dopants into semiconductors. Third, ion beams may be used to deposit material from another target onto the surface, a process known as *sputter deposition*.

In cleaning or etching via sputtering one generally employs relatively low-energy ions (1 to 10 keV) of an inert gas, such as Ar^+, to deposit energy in the surface region. A collision cascade results in which the ion energy is shared among many atoms, much as when a cue ball strikes an array of billiard balls. When the kinetic energy of an excited surface atom exceeds its binding energy, it will leave the solid. Atomic layers of the solid are thereby removed. The sputtering yield Y is the number of sputtered atoms per incident ion. This number is typically between 0.01 and 10 and depends on the energy of the beam and the material being sputtered.

In the ion-implantation process, a low-flux energetic ion beam (10 to 500 keV) penetrates the solid to a depth of ≈ 10 nm to ≈ 10 μm. For example, 200-keV As^+ ions penetrate 20 μm in Si before coming to rest. Some ions are able to penetrate much deeper if the direction of the beam is nearly parallel to a crystal axis through a process called *channeling*. Boron is used almost exclusively as an acceptor. Donor ions include Sb, As, and P. The ions slow down due to collisions with the nuclei and

the electrons and eventually come to rest some distance below the surface. There are a range of penetration depths that occur, with the net result that the solid is doped by the ions. Essentially, any element may be injected and the absolute concentration as well as the concentration profile may be controlled precisely. Since the technique is not thermodynamic in nature, it permits one to build up high concentrations of dopants, beyond the limits imposed by solubility constraints. By subsequent annealing, much of the radiation damage may be removed and the result can be a supersaturated solid solution of the dopant atoms in the host crystal. Precipitation or segregation may also occur. As the incident ion slows down by nuclear collisions, it leaves a trail of radiation damage in its wake. This consists largely of interstitial ions and vacancies. The concentration of displaced ions, N_d, is proportional to the fluence, ψ (the number of incident ions per square meter), and is given approximately by the formula

$$N_d = \frac{4000\psi F_d}{E_d}, \tag{W21.3}$$

where F_d is the energy deposition per unit length of penetration and E_d is the energy needed to displace an ion (10 to 25 eV). In some circumstances the radiation damage may be annealed out by elevating the temperature. In other cases it may be used to create amorphous material. Typical values of ψ are in the range 10^{16} to 10^{19} ions/m^2.

In the ion sputtering process, ion beams are directed at various target materials with different chemical compositions to create a vapor of varying chemical composition. Atoms or molecules from the vapor strike the substrate of interest and stick to it. For example, ion-beam deposition of highly tetrahedral amorphous C is produced with C ions of energy 10 to 100 eV. Layers as thin as a monolayer may be deposited on a substrate. Ion deposition is frequently used for metallization or for coating disks with magnetic material. In some cases the ion beam can assist in the deposition of a chemical vapor directly on the surface by activating the vapor of the material to be deposited.

The path of an incident ion as it penetrates the solid is a directed random walk. In characterizing the penetration of the ion beam, various moments of the distribution of final resting places are employed. Assuming the beam to be directed in the z direction, there is the mean projectile range or penetration depth

$$R_z = \langle z \rangle = \frac{1}{N}\sum_{n=1}^{N} z_n, \tag{W21.4}$$

where N is the number of ions striking the sample and z_n is the penetration depth of the nth ion. The mean radial displacement is given by

$$R_r = \langle \sqrt{x^2 + y^2} \rangle = \frac{1}{N}\sum_{n=1}^{N} \sqrt{x_n^2 + y_n^2}. \tag{W21.5}$$

Higher moments include the straggling distance,

$$\sigma_z = \sqrt{\langle (z - R_z)^2 \rangle} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (z_n - R_z)^2}, \tag{W21.6}$$

the radial straggling distance,

$$\sigma_r = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n^2 + y_n^2) - R_r^2}, \tag{W21.7}$$

and still higher statistical moments of the distribution, such as the skewness (asymmetry) and kurtosis (sharpness of falloff in the wings). Calculations of the spatial distribution, as well as the statistical moments, may be performed by resorting to numerical simulations in which a large number of trajectories is analyzed.

The physical parameters controlling the ion processes are the atomic numbers and masses of the projectile and target, Z_1, Z_2 and M_1, M_2, respectively, the Thomas–Fermi screening constant of the solid, k_{TF} (which curtails the long-range nature of the Coulomb interaction), the incident current, I_1, the beam area, A, and the kinetic energy of the projectile, E. Two energy loss processes are of importance, nuclear stopping and electronic stopping. In the nuclear-stopping process the projectile and target nuclei make a close collision, interacting via the screened Coulomb interaction. Energy and momenta are shared between the two nuclei. In the electronic-stopping process the electric field pulse of a passing projectile ion excites the electrons in the conduction band or upper valence band of the solid. Both interband and intraband excitations may occur. The gain of energy of the electrons is offset by the loss of energy of the projectile, so that energy is always conserved. By the energy-time uncertainty principle, the shorter the duration of the pulse, the wider is the spread of excitation energies. Thus $\Delta E \Delta t \approx h$ with $\Delta t \approx b/v$, where b is the impact parameter (perpendicular distance between the line of approach of the incident ion and the target nucleus) and v is the projectile speed. Hence electronic stopping is expected to dominate at high energies, where a wider range of excitation energy is available due to the shortness of the pulse.

In the nuclear-stopping process the incident ion is deflected from a target ion through an angle ϕ and therefore transfers an amount of energy T to the recoiling target nucleus, where

$$T = \frac{4M_1M_2E}{(M_1 + M_2)^2}\sin^2\frac{\phi}{2}. \tag{W21.8}$$

Maximum energy transfer for a given M_1 and M_2 occurs during backscattering, when $\phi = \pi$. Furthermore, when $M_1 = M_2$ there will be a maximum energy transfer for a given ϕ.

In discussing the energy-loss processes it is convenient to introduce a dimensionless energy, ϵ, defined as the ratio of an effective Bohr radius to the distance of closest approach in a head-on Coulomb collision. The effective Bohr radius is given empirically by $a \approx 0.8854a_1(Z_1^{2/3} + Z_2^{2/3})^{-1/2}$, where a_1 is the Bohr radius, 0.0529 nm. The distance of closest approach is $r_0 = e^2Z_1Z_2(M_1 + M_2)/4\pi\epsilon_0EM_2$. The dimensionless energy is

$$\epsilon = E(\mathrm{keV}) \times \frac{32.53M_2}{(M_1 + M_2)Z_1Z_2\sqrt{Z_1^{2/3} + Z_2^{2/3}}}. \tag{W21.9}$$

A comparison of the nuclear and electronic-stopping powers, $d\epsilon/d\rho$, is given in Fig. W21.6. The scaled penetration distance ρ is the distance in units of a, the effective Bohr radius. The nuclear and electronic stopping powers become equal at some energy.

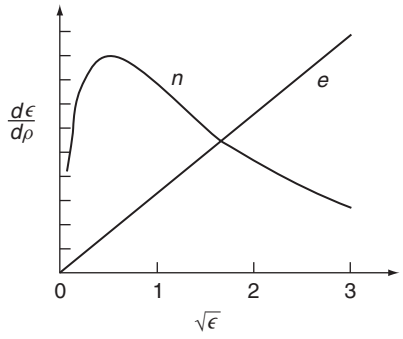

Figure W21.6. Stopping power for nuclear (n) and electronic (e) processes as a function of the parameter ϵ. In Si $\epsilon = 1$ corresponds to $E = 9$ keV for ^{11}Be ions or $E = 1.5$ MeV for Bi ions. (Adapted from J. A. Davies, *Mater. Res. Soc. Bull.*, **17**(6), 26 (1992).

For As, B, and P in Si, this energy is 700, 10, and 130 keV, respectively. Sputtering processes generally occur in the realm $\epsilon < 10$. For $Z_1 > Z_2$ the electronic stopping power is given approximately by the formula $(d\epsilon/d\rho)_e = 0.15\sqrt{\epsilon}$.

The mean projectile range is given by

$$R_z = a \int_0^{\epsilon_{\text{in}}} \frac{1}{(d\epsilon/d\rho)_e + (d\epsilon/d\rho)_n} d\epsilon, \tag{W21.10}$$

where ϵ_{in} corresponds to the incident energy E. An approximate formula for the mean range is

$$R_z(\text{nm}) = E(\text{keV}) \times 13{,}000 \frac{1 + M_2/M_1}{\rho_s Z_1^{1/3}}, \tag{W21.11}$$

with ρ_s being the mass density of the solid (in kg/m^3). The straggling in average total path length R is given approximately for small ϵ by

$$\frac{\Delta R}{R} = 0.7 \frac{\sqrt{M_1 M_2}}{M_1 + M_2}. \tag{W21.12}$$

In reactive-ion etching (RIE) the surface of a solid is exposed to a chemical etchant in the presence of an ion beam. The ion beam serves to excite the reactants, thereby enhancing the chemical reaction rate. The system behaves as if its temperature were elevated. Examples include the etching of Si by F_2, Cl_2, or Br_2 in the presence of an Ar^+ beam. The ion beam also serves to create steps on the surface with dangling bonds available for chemical reaction.

Recently, it has been shown that ion implantation, combined with annealing and recrystallization, can be used to fabricate semiconductor nanocrystals.[†] Alumina substrates were bombarded with semiconductor ion doses up to 10^{21} ions/m^2. If the substrate is kept at a high temperature during bombardment, then cooled and annealed at a relatively low temperature, the substrate retains the α-alumina structure and the

[†] J.D. Budal et al., *Nature*, **390**, 384 (1997).

semiconductor nanocrystals that precipitate align themselves relative to the substrate. If the substrate is bombarded at low temperatures with a high dose of ions, the substrate is amorphized. A low-temperature anneal then leads to the substrate forming γ-alumina. This leads to a different orientation of the nanocrystals than above.

Ion implantation may be combined with etching to produce thin slices of crystals in a technique called *ion slicing*. He^{2+} ions, with an energy of ≈ 4 MeV, impinge on a crystal. The implanted ions deposit a high percentage of their energy near the penetration depth (≈ 10 μm), creating a damage layer. This layer may be attacked with an etching solution and the resulting crystal slice may be delaminated from the rest of the crystal. Subsequently, it could be placed on the surface of a different crystal. This circumvents the need for epitaxial growth of thin films and extends the ability to obtain films on substrates to cases where epitaxial growth may not be possible.

W21.4 Float-Zone Purification of Single-Crystal Si

The purest single crystals of Si are currently grown from the liquid phase using a method in which the molten Si is not in contact with any container, thereby eliminating the main source of impurities. This is the *float-zone* (FZ) method, illustrated schematically in Fig. W21.7, and is a type of zone refining. The starting material is a cylindrical rod of pure, polycrystalline Si which is mounted vertically and held at both ends, either under vacuum or in an inert atmosphere. In this method only a short section of the Si rod away from the ends is molten at any given time. The molten section is heated via radio-frequency induction using a coil surrounding the container and is held in place by surface tension forces. To initiate the growth of a single crystal, a small single-crystal Si seed is placed in contact with the molten end of the rod. A necking process similar to that used in the CZ growth method, described in Chapter 21, is then used to remove any dislocations from the growing crystal.

The external heating coil and the molten Si zone are moved slowly along the Si rod several times in the same direction until the desired purity and crystallinity are obtained. Rotation of the cylindrical rod is also used in this method, to promote cylindrical uniformity of the material. Single crystals of FZ Si of up to 15 cm in diameter

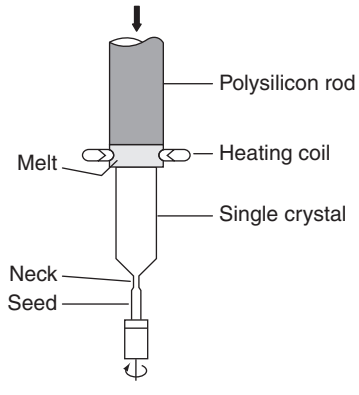

Figure W21.7. Float-zone method used for the growth of extremely pure single crystals of Si and other materials.

can be grown and purified by this technique. The use of FZ Si in Si microelectronic devices is limited due to its low oxygen content, $\approx 10^{22}$ atoms/m^3, a factor of 100 less than in CZ Si. As a result, the beneficial effects of internal gettering and of mechanical strengthening due to oxygen precipitation are not available in FZ Si.

The attainment of extremely high purities in the single-crystal Si rod, corresponding to impurity fractions of $\approx 10^{-10}$ (i.e., 99.99999999% pure Si), results from the much lower solubility of most atoms in solid Si than in liquid Si. This difference in solubility is due to the much more restrictive conditions for the bonding of atoms in solid Si as compared to liquid Si and is expressed in terms of the *equilibrium distribution* or *segregation coefficient* K_{A} for a given atom A. The coefficient K_{A} is the ratio of the equilibrium concentrations of atom A in the two phases:

$$K_{\mathrm{A}} = \frac{c_{\mathrm{A}}(\text{solid})}{c_{\mathrm{A}}(\text{liquid})}. \tag{W21.13}$$

If the fractional concentrations c_{A}(solid) and c_{A}(liquid) are both $\ll 1$, K_{A} is also given by the ratio of the thermodynamic activities of atom A in the two phases. The coefficient K_{A} can be determined experimentally from the equilibrium phase diagram for the Si–A system. If the liquidus and solidus curves are nearly straight lines for low concentrations of A in Si and have negative slopes s_L and s_S, respectively (Fig. W21.8),

$$K_{\mathrm{A}} = \frac{s_L}{s_S} < 1. \tag{W21.14}$$

Solutes that depress the melting temperature of Si have $K_{\mathrm{A}} < 1$, while those that raise T_m have $K_{\mathrm{A}} > 1$.

The distribution coefficient K_{A} for dilute concentrations of A atoms in a solid such as Si can be related to the enthalpy change ΔH_m associated with melting of the solid and to the change of T_m as a function of the A-atom concentration in the solid. The appropriate expression, obtained by equating the chemical potentials of A atoms in the

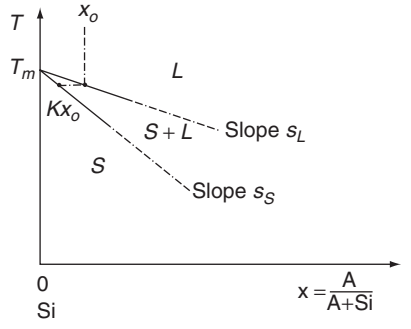

Figure W21.8. Equilibrium phase diagram for the Si–A system. The liquidus and solidus curves are nearly straight lines for low A-atom concentrations and have negative slopes s_L and s_S, respectively.

liquid and solid phases,[†] is

$$K_{\mathrm{A}} = 1 + \frac{\Delta H_{m0}}{RT_{m0}^2} \frac{T_m - T_{m0}}{c_{\mathrm{A}}(\text{liquid})}. \tag{W21.15}$$

Here ΔH_{m0} and T_{m0} correspond to pure Si. For dilute solutions [i.e., c_{A}(liquid) and c_{A}(solid) both $\ll 1$], the ratio $(T_m - T_{m0})/c_{\mathrm{A}}$(liquid) is essentially independent of temperature and so, therefore, is K_{A}. It can be seen from Eq. (W21.15) that, as stated earlier, $K_A < 1$ when $\Delta T_m = T_m - T_{m0}$ is negative, and vice versa.

To illustrate the connection between distribution coefficients and phase diagrams, consider the case of solid-solution Si–Ge alloys whose phase diagram is shown in Fig. W21.9. The distribution coefficients for Ge in Si, K_{Ge}(Si), and for Si in Ge, K_{Si}(Ge), can be obtained from this diagram using the slopes s_L and s_S as the concentrations of Ge and Si tend to zero. The following results are obtained:

$$K_{\mathrm{Ge}}(\mathrm{Si}) \approx 0.3 \quad \text{and} \quad K_{\mathrm{Si}}(\mathrm{Ge}) \approx 5.5. \tag{W21.16}$$

Thus Si atoms have a greater tendency than Ge atoms to enter the solid phase in Si–Ge alloys and actually prefer the solid phase to the liquid phase. The solid phase in equilibrium Si–Ge alloys will therefore always be enriched in Si relative to the liquid phase, as indicated in Fig. W21.9. This follows from the fact that the melting temperature of Si, $T_m = 1414°\mathrm{C}$, is greater than that of Ge, $T_m = 938°\mathrm{C}$. As discussed in Chapter 6, this behavior is also observed for solid-solution Cu–Ni alloys, which are always Ni-rich in the solid phase, Ni having the higher melting point.

Values of c_{A}(solid) obtained experimentally can deviate from those expected from the equilibrium value of K_{A} when the growth process deviates from equilibrium conditions. As an example, K_{A} is observed to depend on the growth rate. It is reasonable to expect that $K_{\mathrm{A}} \to 1$ as the growth rate approaches infinity since A atoms at the growth interface will be trapped in the solid phase due to lack of time to diffuse away.

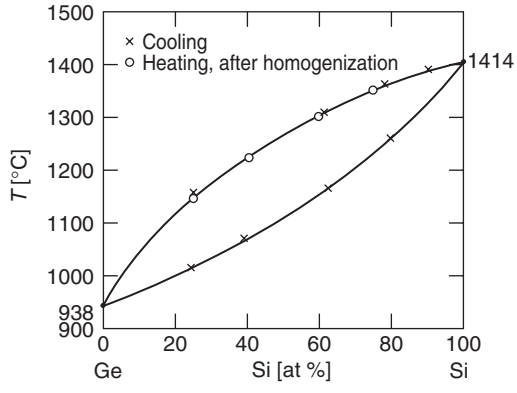

Figure W21.9. Equilibrium phase diagram for solid-solution Si–Ge alloys. (Adapted from M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.)

[†] P. Gordon, *Principles of Phase Diagrams in Material Systems*, McGraw-Hill, New York, 1968, p. 140.

TABLE W21.3 Distribution Coefficients K of Elements in Si Near $T_m = 1414°C$

Column III K		Column IV K		Column V K		Column VI K	
B	0.8	C	0.07	N[a]	$<10^{-7}$	O	0.5
Al	0.002	Si	1	P	0.35		
Ga	0.008	Ge	0.3	As	0.3		
In	0.0004	Sn	0.016	Sb	0.023		

Source: Most values are from F. A. Trumbore, *Bell Syst. Tech. J.*, **39**, 221 (1960).

[a]The value for N is uncertain.

In the FZ method if a given dilute impurity with distribution coefficient $K < 1$ has an initial concentration c_0 in the solid Si rod, the first portion of the Si rod that is melted and then allowed to resolidify slowly will have the lower impurity concentration $Kc_0 < c_0$. The same level of purification will not, however, be achieved in the rest of the Si rod since the concentration of the impurity in the molten zone will slowly increase above c_0. The impurity concentration in the first segment of the Si rod will therefore be reduced by the factor K each time the molten zone is passed slowly through it. Since typically $K \ll 1$ for many unwanted impurities, an extremely low concentration $c \approx K^n c_0$ can in principle be achieved in the first segment of the Si rod after n passes of the molten zone. The opposite end of the Si rod in which the impurities have become concentrated is cut off after the purification process is completed. Since the impurity concentration, while low, will still be nonuniform along the length of the Si rod, homogenizing treatments that involve passing the molten zone repeatedly along the rod in both directions are employed to obtain a uniform impurity concentration.

Values of the equilibrium distribution coefficients for several elements in Si are given in Table W21.3. The only elements with distribution coefficients in solid Si which are greater than 0.05 are from groups III, IV, V, and VI of the periodic table (e.g., B, C, Ge, P, As, and O). The elements B, P, and As are substitutional impurity atoms which are often used for doping Si. Unwanted metallic impurities such as Cu, Au, and Zn have very low values of $K \approx 10^{-7}$ to 10^{-4}. The coefficient K is observed to be temperature dependent, falling rapidly with decreasing T.

In addition to its use for Si, the FZ technique remains the preferred method for obtaining highly purified crystals of a wide variety of semiconducting, metallic, and ceramic materials, including single crystals of the high-T_c superconductor La–Sr–Cu–O.

W21.5 Epitaxial Growth of Single-Crystal Si Layers via CVD

The *homoepitaxial* growth of single-crystal layers (*epilayers*) of Si on Si substrates as carried out via *chemical vapor deposition* (CVD) is the preferred method of growth for the layers used in the fabrication of Si-based electronic devices. The CVD of Si employs a wide variety of deposition systems and conditions and so is a very versatile growth procedure. The CVD process involves the thermal decomposition (pyrolysis) of gaseous precursor molecules, with both vapor-phase (*homogeneous*) and surface (*heterogeneous*) chemical reactions playing important roles. It is desirable, in general, to suppress vapor-phase chemistry to avoid powder formation and the defects that

would result from particle incorporation in the films. The Si epitaxial layers deposited undergo further processing when used in Si-based electronic devices. These additional processing steps are discussed in Section W21.8, where the fabrication of Si-based integrated circuits is described.

The growth of Si from the vapor phase at substrate temperatures in the range $T_s = 500$ to 1150°C has several advantages relative to the Czochralski and float-zone methods, which involve growth from the melt at $T_m = 1414$°C. The advantages include reduced diffusion of both dopant and unwanted impurity atoms and reduced thermal stresses in the film and substrate. Reduced dopant diffusion allows the fabrication of abrupt interfaces between regions of different doping levels, an important factor in the development of smaller and faster devices.

The single-crystal Si wafers used as substrates for the epitaxial growth of Si layers are grown via the Czochralski method and are required to be as defect-free as possible since dislocations and other structural defects present in the substrate can propagate into the growing film. The surface of the substrate must also be smooth and clean (i.e., free from impurities such as carbon and oxygen), to prevent the nucleation of stacking faults and the appearance of other defects, such as dislocations, voids, inclusions, and precipitates in the growing film. There exist well-developed polishing and cleaning procedures, both ex situ and in situ, for the preparation of Si wafers for use as substrates. Ex situ chemical cleaning, which results in an air-stable, oxide-free Si surface, involves an H_2O_2-based chemical cleaning procedure, the *RCA clean*,[†] followed by a 10-s dip in a 10:1 H_2O:HF solution. This treatment generates a hydrophobic Si surface which is chemically stabilized by a surface layer of strong Si–H bonds. In situ cleaning methods include high-temperature treatments, often in H_2, to remove any SiO_2 present on the surface as volatile SiO molecules and also to remove C from the surface via its diffusion into the bulk or by the evaporation of the surface layer of Si.

A typical cold-wall Si CVD system is shown in Fig. W21.10. It consists of a water-cooled fused-quartz tube surrounded by radio-frequency heating coils into which the Si wafer substrates are placed in a susceptor made of graphite, SiC-coated graphite, or quartz. The deposition can be carried out at atmospheric pressure (APCVD) or at reduced pressures (RPCVD), $P \approx 0.01$ to 0.1 atm. The current standard epitaxial growth method is RPCVD, which has the advantage of minimizing autodoping (i.e., the doping of the growing Si layer by dopant atoms originating from the Si substrate).

Film growth from the vapor phase is a very general method of materials synthesis and typically involves the following steps, each of which may in fact represent a complicated sequence of more elementary steps:

1. Transport of gaseous species from the source to the substrate
2. Adsorption onto the substrate surface
3. Nucleation and growth of the film
4. Removal from the surface of unwanted species that might interfere with film growth

The nucleation and growth steps are described in Section W21.2. The thermal decomposition of the gaseous species can occur either in the vapor phase or on the

[†] W. Kern and D. A. Puotinen, *RCA Rev.*, **31**, 187 (1970).

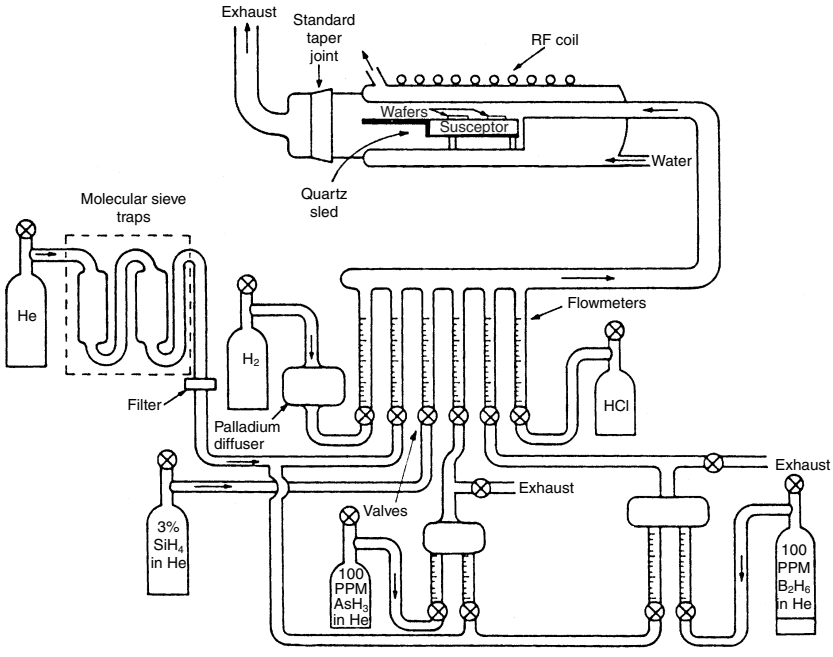

Figure W21.10. Typical cold-wall Si CVD system. (From D Richman et al., RCA Review, **31**, 613 (1970).)

heated substrate surface. The hydrodynamics of the flowing gases in the CVD system can have a significant influence on the growth process.

In the case of Si CVD, there are many possible choices for the molecular precursors, including SiH_4 and $SiHCl_3$. The important growth species present on the surface are then the highly reactive radicals silylene, SiH_2, and $SiCl_2$. These radicals are the products of the thermal decomposition of the feedstock gases and will undergo further reactions on the surface of the growing film. Carrier gases such as H_2 and He are often used to aid in the transport of vapor species to the substrate. The concentrations of atoms, radicals, and molecules adsorbed on the growing surface are controlled by their incident fluxes (i.e., by their partial pressures in the vapor phase) and by the substrate temperature T_s which controls their desorption rates.

Typical net chemical reactions resulting in the growth of the Si epilayer include the following:

$$\begin{aligned} SiH_4(g) &\longleftrightarrow Si(s) + 2H_2(g), \\ 2SiHCl_3 &\longleftrightarrow 2Si(s) + 3Cl_2(g) + H_2(g). \end{aligned} \tag{W21.17}$$

These reactions actually represent a series of elementary steps taking place in the vapor phase and on the substrate surface. Growth rates are ≈ 1 μm/ min at $T_s \approx 1100^\circ$C and decrease rapidly as T_s is lowered (see Fig. 21.3). Homogeneous vapor-phase reactions leading to the formation of disilane Si_2H_6 are

$$\begin{aligned} SiH_4(g) &\longrightarrow SiH_2(g) + H_2(g), \\ SiH_4(g) + SiH_2(g) &\longrightarrow Si_2H_6(g). \end{aligned} \tag{W21.18}$$

These reactions can ultimately lead to the formation of undesirable polymeric silicon hydride powder, $(SiH_2)_n$.

The partial pressures of the vapor species involved in growth must exceed their equilibrium vapor pressures with respect to the Si surface at T_s in order for the net deposition of a film to occur. The growth species must therefore be supersaturated in the vapor phase, with the *supersaturation ratio* SSR for the case of Si(g) atoms defined by

$$\mathrm{SSR}(\mathrm{Si}(g), T_s) = \frac{P(\mathrm{Si}(g))}{P_{\mathrm{eq}}(\mathrm{Si}(g), T_s)}, \tag{W21.19}$$

where $P(\mathrm{Si}(g))$ is the actual vapor pressure of Si(g) just above the substrate surface and $P_{\mathrm{eq}}(\mathrm{Si}(g), T_s)$ is the equilibrium vapor pressure of Si(g) with respect to pure Si(s).

A wide variety of investigations have allowed the following conclusions to be reached concerning the growth of Si epilayers via CVD:

1. The rate-controlling step for the growth of Si is either the removal from the surface of hydrogen in Si–H bonds via the desorption of H_2, or the dissociation of SiH_2 or $SiCl_2$ on the surface.
2. The rate-controlling step for obtaining high crystallinity in the Si epilayer is the diffusion of Si on the growing surface.
3. Lattice defects are generated when the Si adsorption rate exceeds the rate at which Si can diffuse on the surface and be incorporated into the growing film. Si atoms then enter nonideal, higher-energy bonding configurations.
4. Si atoms compete with other species on the surface, such as dopant atoms or molecules and hydrogen, oxygen, or carbon atoms, for the available bonding sites to Si substrate atoms, thereby limiting the Si atom diffusion rate.

The termination of the growing Si surface by hydrogen in Si–H bonds can play a critical role in the CVD of Si by inhibiting epitaxial growth through the blocking of surface sites for the adsorption of reactive species such as SiH_2 and SiH_3. This is particularly important at T_s less than about 400 to 500°C.

Recently, the CVD of Si and of Si–Ge alloys has been combined with UHV techniques to achieve a very high level of system and substrate cleanliness (e.g., the elimination of oxygen and carbon surface impurities). The use of this growth method, known as *UHV/CVD*, allows the deposition of epitaxial Si and Si–Ge layers at much lower pressures, $P \approx 10^{-3}$ torr, and lower T_s, $\approx$ 500 to 550°C, than are ordinarily used. Operation at lower pressures has several advantages: the undesirable homogeneous pyrolysis of precursors in the vapor phase is minimized, the very low partial pressures of O_2 and H_2O necessary for the maintenance of an active, SiO_2-free Si surface are more readily achieved,[†] and molecular flow conditions are obtained, with the result that recirculating flows, eddy currents, and turbulence are avoided. Due to the clean and hydrogen-stabilized surfaces of the Si wafers when they are placed into

[†] For experimental results and discussions of the interactions of O_2 and H_2O with Si at high temperatures, see F. W. Smith and G. Ghidini, *J. Electrochem. Soc.*, **129**, 1300 (1982); G. Ghidini and F. W. Smith, *J. Electrochem. Soc.*, **131**, 2924 (1984).

the UHV/CVD system, no further in situ treatment at high temperatures is required to prepare the Si surface for epitaxial growth.

The use of lower substrate temperatures reduces problems associated with dopant atom redistribution via diffusion and also is a very effective method of reducing defect concentrations in the films. Growth at lower T_s will reduce the equilibrium concentrations of defects such as vacancies and will also reduce the mobility of point defects and hence their tendency to interact with each other to form extended defects. In addition, thermal stresses which can also lead to the generation of defects in the film will be reduced at lower T_s. Better film thickness uniformity is also expected at lower T_s since the deposition process changes from one controlled by vapor-phase transport at higher T_s to one controlled by surface reactions at lower T_s, as discussed in Section 21.3. It is still necessary to maintain T_s well above the range in which the film will become noncrystalline or amorphous.

Nonequilibrium structures and alloys can also be prepared at low T_s. These include strained Si–Ge epilayers grown on Si with thicknesses well above the critical values for the generation of misfit dislocations and also alloys of Si with concentrations of dopant atoms such as B which are several orders of magnitude above equilibrium concentrations. Sharp transitions, particularly in dopant profiles, between the substrate and the epilayer are essential as device dimensions continue to shrink. Both the layer growth rate and dopant diffusion rates decrease exponentially as T_s decreases. Since the activation energy for diffusion, $E_a(\text{diff}) \approx 3.5$ eV, is much greater than that for growth, $E_a(\text{growth}) \approx 1.5$ eV, reasonable growth rates, ≈ 0.1 to 10 nm/min, can still be obtained at $T_s \approx 500^\circ$C, where dopant diffusion has been effectively frozen out.

A schematic of the hot-wall apparatus used in the UHV/CVD method is shown in Fig. W21.11. The carefully cleaned Si wafers have surfaces passivated by H termination (i.e., Si–H bonds), which can be thermally desorbed from the Si surface at $T_s > 400^\circ$C. In the UHV/CVD of Si the vapor phase consists entirely of SiH_4.

Films that are "defect-free" (i.e., with defect densities less than ≈ 100 cm^{-2}) are readily achieved via CVD. The most sensitive quantitative method of determining

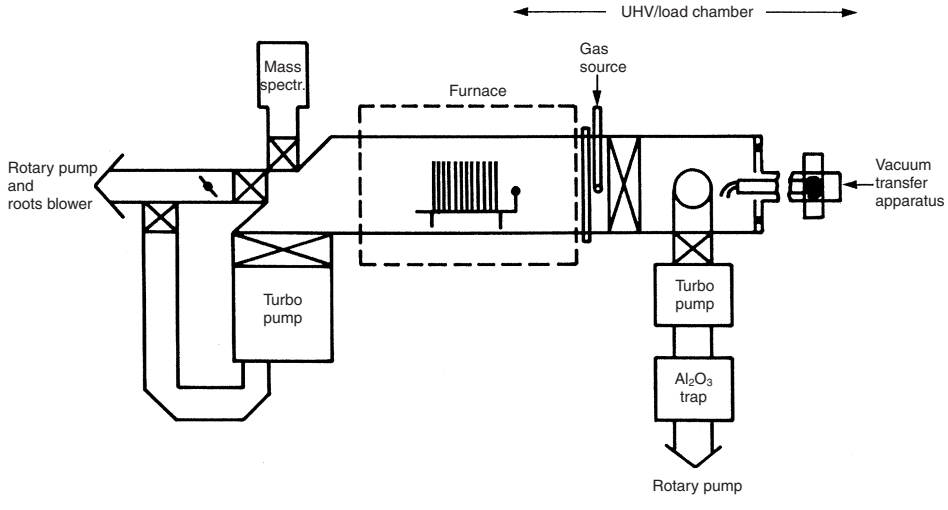

Figure W21.11. UHV/CVD system. (From B. S. Meyerson, *Appl. Phys. Lett.*, **48**, 797 (1987). Copyright 1987 by the American Institute of Physics.)

densities of structural defects such as dislocations in Si epitaxial layers is by means of chemical etching. Since the disordered regions of the lattice containing defects are in a state of higher energy, they are more rapidly attacked (i.e., etched) by appropriate acids. Optical microscopy can then be used to count the etch pits and also to identify the nature of the defects from the shape of the etch pit. Transmission electron microscopy (TEM) is the preferred method for probing the atomic perfection of the interface between the substrate and the epilayer. Electrically active defects such as impurity-related traps are not readily detected via etching or TEM. Their presence can be determined by the effects that they have on devices such as diodes, transistors, or metal–oxide–semiconductor (MOS) capacitors, which are fabricated from the Si epilayers.

Metallic elements such as Fe and other transition metals are undesirable impurities in Si due to the fact that they act as traps (i.e., as centers for the recombination of electrons and holes). Although they do not enter into CZ or FZ Si from the melt due to their very low distribution coefficients, they will diffuse rapidly into the bulk at elevated temperatures if they can reach the surface of the Si crystal through the vapor phase.

Other recent approaches to Si epitaxy via CVD include the use of intermediate layers such as cubic CaF_2, fluorite, whose lattice constant, $a = 0.546$ nm, matches that of Si, $a = 0.543$ nm, to within 0.6% at $T = 300$ K. The CaF_2 layer is deposited epitaxially onto the Si(100) surface first, followed by the deposition of the Si epilayer onto the CaF_2 layer. The top Si epilayer is then removed for further processing by dissolving the intermediate CaF_2 layer in an appropriate solvent. In this way the original Si(100) substrate can be reused.

A recent approach to understanding the growth of Si epilayers at low temperatures has involved the definition of a limiting *epitaxial thickness* h_{epi} above which the deposited films become amorphous. This is in contrast to the usual definition of a minimum *epitaxial temperature* T_{epi}, below which epitaxy is impossible, due to insufficient surface diffusion of atoms adsorbed on the surface. Epitaxial growth of Si can be observed in a very clean MBE system at all temperatures between $T = 50$ and 300°C, but only up to the thickness h_{epi}, which increases exponentially with increasing T and decreases with increasing growth rate. For Si films grown via MBE, h_{epi} was found to be 1 to 3 nm at room temperature. The transition from crystalline to amorphous growth at h_{epi} has been attributed to a surface-roughening effect, with the accumulation at the growing surface of impurity atoms such as hydrogen playing a major role in the roughening process.

W21.6 Molecular-Beam Epitaxial Growth of GaAs

The growth via *molecular-beam epitaxy* (MBE) of films of the group III–V semiconductor GaAs, as well as of other III–V and II–VI semiconductors, has many features in common with the CVD of epitaxial Si layers, including the steps of transport and adsorption of the appropriate precursor vapor species onto the substrate surface, nucleation and growth of the film, and removal of unwanted species from the substrate surface. In MBE molecular beams (i.e., beams of neutral molecules or atoms) are directed onto a heated substrate in a UHV system. Due to the low particle density of the beam and also to the very low background pressure in the growth chamber, the particles in the beam do not interact with each other and undergo essentially no collisions

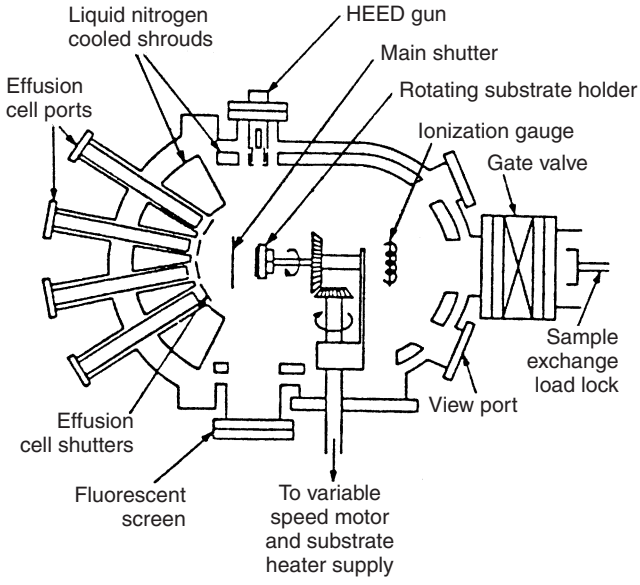

Figure W21.12. Typical MBE vacuum chamber. (Reprinted from A. Y. Cho, *Thin Solid Films*, **100**, 291 (1983), copyright 1983, with permission from Elsevier Science.)

with residual gas molecules on their path from the source to the substrate. A typical MBE growth chamber is shown schematically in Fig. W21.12. Along with the vacuum chamber and all the associated accessories, appropriate vacuum pumps and electronics for the control of the various components are required. The mass spectrometer is used for residual gas analysis. It can also be used to measure the fluxes of reactant species and can provide signals to be used for adjusting the effusion cell temperatures so that constant fluxes, and hence constant deposition rates, can be maintained.

Advances in UHV technology[†] have permitted the deposition via MBE of films at relatively low T_s with unparalleled control of composition, purity, and interface sharpness, involving literally atomic layer-by-layer growth. The low growth temperature has the advantage of reducing undesirable thermally activated processes such as diffusion, while the low growth rates ($\approx$ 10 nm/ min) offer the advantage of accurate control of film thickness. The UHV conditions employed in MBE also permit in situ monitoring of the film structure and thickness using high-energy electron beams reflected at very low angles from the surface of the growing film. This technique is known as *reflection high-energy electron diffraction* (RHEED). The chemical purity and composition of the substrate and of the film can also be monitored in situ using Auger electron spectroscopy (AES). Finally, the use of modulated-beam mass spectrometry (MBMS) employing separate beams of Ga and As_2 has allowed the detailed study of surface processes involved in the growth of GaAs via MBE.

The solids that are the source materials for the MBE of GaAs are contained in heated effusion cells within the vacuum chamber. Elemental Ga metal is used for the Ga flux, while solid GaAs is used for As_2 and solid elemental As for As_4. Additional elements

[†] See Weissler and Carlson (1979) for a useful description of UHV techniques.

used for doping, alloying, and for multilayer or junction depositions are contained in their own effusion cells. The nature and flux of the vapor species from each effusion cell are controlled by the temperature of the cell, with the flux directed through a small orifice in the wall of the cell toward the substrate. Shutters placed between each cell and the substrate are used to block individual beams when control of the composition or thickness of the growing film is desired. The substrates are mounted on heated holders whose temperature T_s can be controlled accurately by regulated internal heaters. The substrate holders can be rotated during growth in order to obtain extremely uniform epitaxial films.

Due to the very low background pressure in the MBE chamber during growth, $P \approx 10^{-9}$ torr ($\approx 10^{-7}$ Pa), very few unwanted residual gas molecules are incident on the substrate and incorporated into the films. Due to the cleanliness of the growth chamber, growth rates can be very low, 6 to 60 nm/min, which allows extremely thin layers with abrupt interfaces to be grown on surfaces that are essentially atomically smooth. Typical beam fluxes can be in the range 10^{11} to 10^{16} atoms (or molecules)/cm^2·s.

The substrates used for GaAs integrated-circuit fabrication are semi-insulating bulk GaAs crystals grown via the *liquid-encapsulated* Czochralski method. These undoped substrates typically contain 10^4 to 10^5 dislocations/cm^2. Before being placed in the growth chamber the substrates undergo a variety of polishing, etching, and rinsing procedures which are chosen carefully for each type of substrate. Further treatment of the substrate within the growth chamber is also possible and typically involves heating to about $T = 580^\circ$C to remove oxygen, followed by Ar ion bombardment to remove the less volatile carbon contamination. To obtain extremely clean growth surfaces, undoped epitaxial layers of GaAs are often grown in the MBE growth chamber on existing bulk substrates.

Stoichiometric GaAs films are typically grown in the range $T_s = 500$ to 600°C under an incident vapor flux that is enriched in As-containing species due to the instability of the heated GaAs surface with respect to the preferential loss of more volatile arsenic species. When As_2 is incident, stoichiometric GaAs films are obtained as long as the As_2 flux exceeds 50% of the Ga flux [i.e., as long as $R(As_2)/R(Ga) > 0.5$]. The sticking coefficient of Ga is equal to unity for T_s less than about 480°C and then decreases exponentially with an activation energy of $E_a \approx 2.5$ eV at higher temperatures. Under proper growth conditions any excess arsenic beyond that needed for stoichiometric growth is desorbed from the surface of the growing film. This is attributed to a high sticking coefficient for As_2 on a Ga-terminated surface and a low sticking coefficient for As_2 on an As-terminated surface, as observed experimentally. As a result, the growth rate of GaAs, which is controlled by the incident monoatomic Ga flux, can also be limited kinetically by the desorption of As-containing species that block sites for the incorporation of Ga atoms.

The GaAs growth process from Ga and As_2 has been shown by sensitive MBMS and RHEED studies to be limited by the first-order dissociative chemisorption of As_2 molecules when they encounter pairs of vacant As sites next to filled Ga sites. Growth of GaAs from Ga and As_4 has been shown to be more complicated, involving the dissociation of pairs of As_4 molecules on adjacent Ga atoms. Four of the resulting eight As atoms are incorporated into the growing film while the remaining four desorb as As_4. The doping of GaAs films for high-frequency and light-emitting device applications occurs during growth and is controlled by a variety of thermodynamic and kinetic

effects. For example, a dopant element such as Cd or Zn with a high vapor pressure can desorb from the growing surface and so may not be incorporated.

For a given substrate material there is a well-defined temperature range for the growth of high-quality epitaxial films. For example, MBE of GaAs is typically carried out for T_s between 500 and 600°C. The low-T_s limit is related to decreasing crystallinity, while the high-T_s limit is due to the high vapor pressure of As_2 and the resulting deviations from stoichiometry. The lower limit for T_s can be extended down to 200 to 300°C by using reduced arsenic fluxes, and the upper limit can be extended up to 700°C with the use of higher arsenic fluxes. Films deposited at $T_s = 700$°C are of higher quality (e.g., purer), due to reduced incorporation of impurities such as oxygen, which form volatile molecules that desorb from the growth surface at high T_s.

MBE systems are usually dedicated to the deposition of specific materials [e.g., either group III–V (GaAs, GaP, InP, etc.) or II–VI (ZnSe, CdTe, etc.) compound semiconductors]. For each group of materials the compositions and configurations of the films or superlattices deposited is essentially unlimited, with the only constraint being the imagination of the grower. MBE is a versatile deposition technique which, in addition to being used for group III–V and II–VI semiconductors, has also been used for the deposition of elemental semiconductors such as Si and Ge, for metals such as α-Fe, Co, and Al, and insulating layers such as CaF_2.

Other techniques used for the deposition of compound semiconductor thin films includes *metal–organic* CVD (MOCVD), metal–organic MBE (MOMBE), also known as *chemical beam epitaxy* (CBE), which make use of volatile organometallic compounds such as trimethyl gallium, $(CH_3)_3Ga$. When arsine, AsH_3, is used as the source of As, a typical reaction leading to the growth of GaAs is $(CH_3)_3Ga + AsH_3 \rightarrow GaAs + 3CH_4$.

W21.7 Plasma-Enhanced CVD of Amorphous Semiconductors

The use of energetic radio-frequency (RF) and microwave plasmas to produce highly-reactive chemical species (excited atoms, molecules, radicals, and ions) allows deposition of a wide variety of semiconducting and insulating thin films onto practically any substrate at low temperatures, typically in the range $T_s = 25$ to 500°C. Important advantages of this *plasma-enhanced* CVD (PECVD) method are that high-temperature materials such as oxides, nitrides, and carbides can be deposited without excessive heating of the substrate and also that large-area substrates can be coated. Low-temperature deposition is important because lower temperatures are required in integrated-circuit fabrication, due to the need to avoid diffusion of dopant atoms and due to the presence of the low-melting-point metal Al used for device interconnections. As a result of the lower T_s, the films deposited are usually *amorphous* and also often highly nonstoichiometric, with significant deviations from the nominal SiO_2, Si_3N_4, and SiC compositions in the case of Si-based films. Depending on the precursors employed and the substrate temperature, the films also can contain up to ≈ 40 at % hydrogen, which is chemically bonded in the random covalent network.

Despite the absence of long-range order, a considerable degree of short-range chemical order, corresponding to the strongest possible set of chemical bonds, is usually present in these films. This type of bonding results from the good atomic mixing taking place at the surface of the growing film as a result of energetic species (e.g., ions) incident from the plasma. This atomic mixing allows bonding configurations to be achieved

which correspond to a state of low enthalpy. The Gibbs free energy $G = H - TS$ for these amorphous films results from competition between achieving the lowest-possible enthalpy H, corresponding to the strongest set of chemical bonds in the network, and achieving the highest possible entropy S, corresponding to random bonding between the atoms in the network. A *free-energy model* for the bonding in amorphous covalent networks has been formulated which takes into account the effects of both enthalpy and entropy.[†]

Interesting and important examples of amorphous films deposited by PECVD include hydrogenated amorphous Si (i.e., a-Si:H), amorphous silicon oxide, nitride, and carbide (i.e. a-SiO_x:H, a-SiN_x:H, and a-SiC_x:H), and amorphous or diamond-like carbon (DLC) (i.e., a-C:H). One of the important advantages of the PECVD method is that films with a wide range of compositions can be deposited due to the wide variety of available gas-phase precursors and to the considerable range of deposition parameters such as T_s, discharge pressure and power, and substrate bias potential, which controls the bombardment of the film by ions. As a result, film properties such as the optical energy gap and the electrical conductivity at room temperature can be varied over wide ranges [e.g., between ≈ 0 and 5 eV and between 10^{-14} and $10^{-2}(\Omega\cdot\text{m})^{-1}$, respectively]. Available gaseous precursors include SiH_4, O_2, H_2O, NH_3, and hydrocarbons such as CH_4 and C_2H_2. Other precursors, such as borazine ($B_3N_3H_6$) and tetraethoxysilane [TEOS, $Si(OC_2H_5)_4$], can be generated from liquids. Gases such as diborane (B_2H_6) and phosphine (PH_3) can be added directly to the discharge when doping of the deposited layer (e.g., a-Si:H) is desired. Precursors that are typically used in the PECVD of thin films are listed in Table W21.4.

PECVD films have a wide range of semiconducting, dielectric, and protective-coating applications. Examples include n- and p-type a-Si:H in photovoltaic solar cells and thin-film transistors (TFTs), a-SiO_x:H as a dielectric layer and a-SiN_x:H as an encapsulating layer in semiconductor devices, p-type a-SiC_x:H as a window layer in a-Si:H solar cells, and a-C:H as a protective coating for magnetic-recording media, and so on.

As a specific example of the PECVD process, consider the deposition of hydrogenated amorphous silicon nitride, a-SiN_x:H, from SiH_4 and NH_3 mixtures using volume flow ratios $R = NH_3/SiH_4$. Under typical conditions [e.g., $T_s = 400^\circ$C and $P = 0.5$ torr ($= 66$ Pa) in RF discharges], the deposition rates of these a-SiN_x:H films are ≈ 0.1 to 0.5 nm/s and are controlled by the SiH_4 flow rate. This occurs because

TABLE W21.4 Typical Precursor Gases Used in PECVD

Film	Precursor Gases	Film	Precursor Gases
a-Si:H	SiH_4, SiH_4/H_2	a-Ge:H	GeH_4, GeH_4/H_2
a-C:H	C_2H_2, C_2H_4, C_6H_6	a-SiN_x:H	SiH_4/NH_3, SiH_4/N_2,
a-SiO_x:H	$Si(OC_2H_5)_4/O_2$,		SiH_2Cl_2/NH_3
	SiH_4/O_2,	a-SiC_x:H	SiH_4/C_2H_2
	$SiH_4/Ar/N_2O$	a-BN_x:H	$B_3N_3H_6$, B_2H_6/NH_3
a-C:F	CF_4, C_2F_4		

[†] For the application of the free-energy model to a-SiN_x:H, see Z. Yin and F. W. Smith, *Phys. Rev. B*, **43**, 4507 (1991); for a-C:H, see H. Efstathiadis, Z. L. Akkerman, and F. W. Smith, *J. Appl. Phys.*, **79**, 2954 (1996).

SiH_4 is dissociated much more rapidly than NH_3 in the plasma. For $R = 0$ a-Si:H films are deposited, and for $R \ll 1$ a fraction of the incorporated N atoms can act as substitutional donor impurities in a-Si:H. As R increases still further and more N is incorporated, the optical energy gap widens and the films become electrically more insulating. For very high ratios, $R \approx 60$, and for lower $T_s \approx 100^{\circ}C$, the films become N-rich, with N/Si ratios that can exceed the stoichiometric value of $\frac{4}{3}$ for Si_3N_4. These films do not correspond to a-Si_3N_4, even when N/Si $= \frac{4}{3}$ due to the incorporation of H in the range 10 to 30 at %.

The a-SiN_x:H films used in devices have N/Si ≈ 1 and typical compositions given by a-$Si_{0.4}N_{0.4}H_{0.2}$. Undesirable bonding configurations in these films include Si–Si bonds and Si–NH_2 bonding units. The former lead to an increase in the dielectric function and also cause optical absorption at low energies, while the latter lead to a lack of chemical and thermal stability. Films with higher H contents are in general not useful in devices. Films with compositions close to the compound silicon diimide [i.e., $Si(NH)_2$], the bonding analog of SiO_2, with NH units replacing O atoms, can be obtained at very high NH_3/SiH_4 flow ratios. Films of $Si(NH)_2$ are unstable in the presence of H_2O due to the chemical reaction $Si(NH)_2(s) + 2H_2O(g) \leftrightarrow SiO_2(s) + 2NH_3(g)$, particularly when Si–$NH_2$ bonding units are present. Films of a-SiN_x:H thus provide a typical example of how H incorporation can play a key role in controlling the properties of amorphous semiconducting and insulating films.

The plasmas used in PECVD processes include RF plasmas at 13.56 MHz (wavelength $\lambda = 22.1$ m) and microwave plasmas at 2.45 GHz ($\lambda = 12.2$ cm). The RF plasmas are often employed using a capacitively coupled parallel electrode configuration, as shown in Fig. W21.13, although inductive coupling is also used. The microwave plasmas typically consist of a plasma ball with dimensions of a few

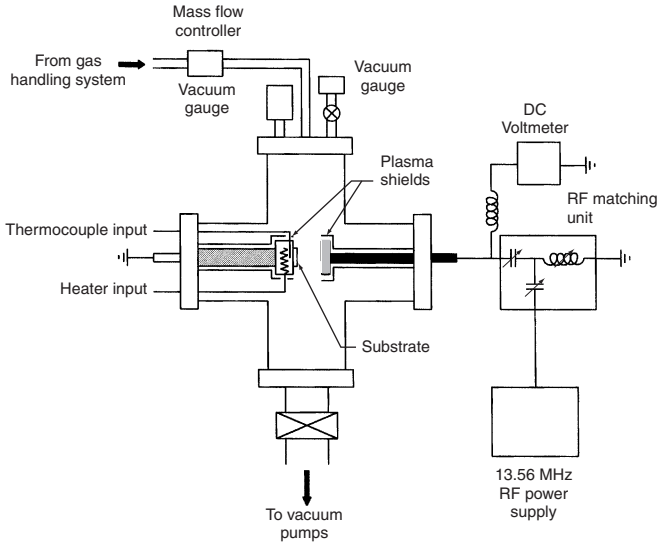

Figure W21.13. The RF plasmas used in plasma-enhanced CVD are typically employed in a capacitively coupled parallel electrode configuration, as shown here. (From K. Mui et al., *Phys. Rev. B*, **35**, 8089 (1987). Copyright 1987 by the American Physical Society.)

centimeters and are usually more confined in space than their RF counterparts. Electron cyclotron-resonance (ECR) plasmas which employ magnetic fields to aid in the coupling of energy into the plasma are also used in low-pressure discharges. Electron-impact dissociation of the feedstock gas in the plasma provides the excited neutral and charged species (i.e., free radicals and ions) needed for film deposition. Chemical reactions occurring in the gas phase and on the surface of the growing film can also produce species that are important for the deposition process.

A complete description and analysis of all the important processes occurring both in the plasma and on the surface of the growing film during PECVD is an extremely difficult task, due to the large number of possible species and processes and the often unknown rate constants and cross sections of these processes. A schematic model of the gas-phase and surface processes involved in the PECVD of a-Si:H from SiH_4 is shown in Fig. W21.14. The various ions, neutral radicals, and other molecular species present in the vapor phase are indicated, as are some of the surface reactions. The presence of the H-rich surface layer on the growing a-Si:H film is apparent. The net growth rate is the result of the competition between the deposition and etching rates. In most PECVD processes the substrate to be coated is mounted in a vacuum system on a heated substrate holder so that T_s can be varied from room temperature up to $\approx 400°C$. Typical discharge pressures are in the range 0.1 to 10 torr (13 to 1300 Pa) and typical plasma energy fluxes at the substrate are 10 to 100 mW/cm^2.

Hydrogen dilution (i.e., adding H_2 to the plasma) often has the advantage of actually reducing the hydrogen content of the deposited film by, for example, enhancing the removal from the growing surface of weakly bonded species such as SiH_2 or SiH_3.

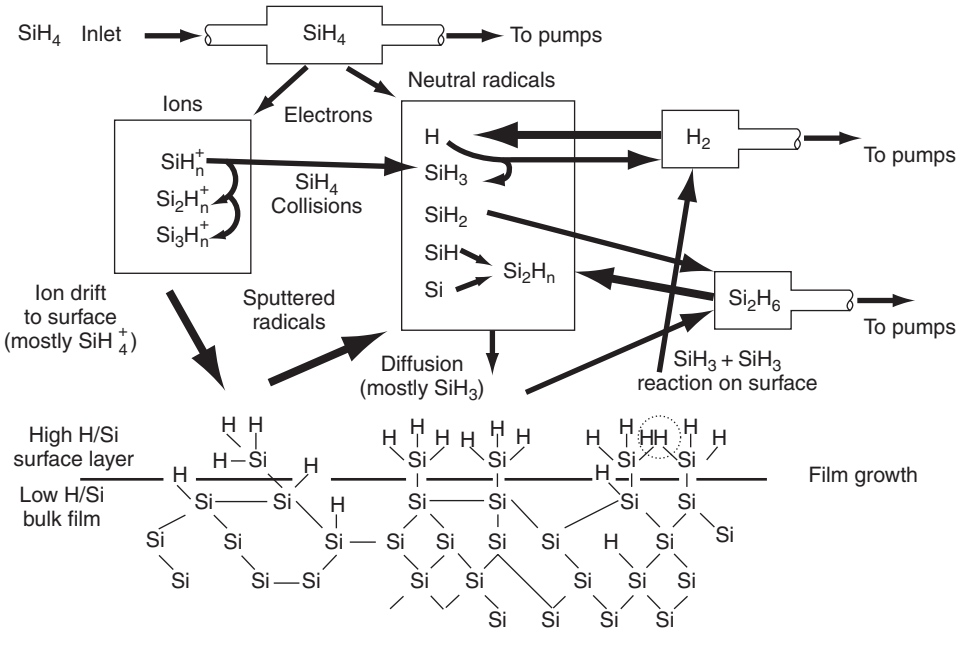

Figure 21.14. Gas-phase and surface processes involved in the plasma-enhanced CVD of a-Si:H from SiH_4. (From A. Gallagher, in *The Physics of Ionized Gases*, J. Puric and D. Belic, eds., World Scientific Press, 1987, p. 229.)

Another method used to reduce the hydrogen content is increasing T_s, which leads to increased mobility of the H atoms within the films, and their recombination into H_2 molecules, which can then diffuse to and desorb from the film surface. Higher deposition rates are also possible at higher T_s. The use of higher T_s allows greater atomic diffusion to occur in the films, which aids in the annealing (i.e., healing) of defects. Film stress and morphology are also strongly dependent on T_s as well as on ion bombardment.

Changes in the PECVD growth conditions, such as increasing the partial pressure of H_2 in SiH_4/H_2 mixtures, increasing the power density or the frequency of the plasma, or increasing the substrate temperature T_s, can lead to the deposition of *microcrystalline* (μc) films such as μc-Si:H. These μc-Si:H films have microstructures consisting of variable volume fractions of Si nanocrystals in an a-Si network. Preferential etching of the more weakly bonded amorphous component by H atoms is likely to play an important role in the deposition of μc-Si:H films.

In addition to deposition, reactive plasmas can also be used in a wide variety of etching processes, such as those used in the fabrication of Si devices. Some of these etching applications are discussed in Section W21.8. The plasma hardening of metal surfaces by the implantation of N or C ions, discussed in Section W21.13, and plasma doping by implantation of B ions into Si are also important materials processing procedures.

Another plasma-related mode of film deposition makes use of the *physical sputtering* of atoms from a target in, for example, an Ar plasma. The target material, as well as the deposited layer, can be a metal, semiconductor, or an insulator. The sputtered atoms are incident on the substrate, where they lead to the desired layer deposition. Physical sputtering is typically used for the deposition of metal films.

In another mode of operation, known as *reactive sputter deposition*, additional precursor gases are introduced into the plasma, where they are excited. These excited species contribute to the layer deposition since they can react with the target atoms both at the surface of the growing film and on the surface of the target. This method can readily be used to control the composition of the deposited layer. Reactive sputtering is typically used for the deposition of compound films such as oxides (including the high-T_c superconducting copper-based oxides), nitrides, carbides, and silicides. Typical precursor gases include O_2 and H_2O for oxygen, NH_3 and N_2 for nitrogen, CH_4 and C_2H_2 for carbon, SiH_4 for silicon, and H_2 when hydrogen is to be incorporated, as in a-Si:H.

W21.8 Fabrication of Si Devices

A brief overview of the important steps involved in the fabrication of Si-based electronic devices from Si wafers of sufficiently high resistivity is presented next. To illustrate the complexity of the process, consider the fabrication of a 256-Mbit dynamic random-access memory (DRAM). A wafer yields 16 chips, each 25 mm square and consisting of $\approx 3 \times 10^8$ devices with features as small as 0.25 μm. Due to the large number (≈ 300) of synthesis and processing steps involved in IC fabrication, it is not possible here to describe these procedures in detail. Wolf and Tauber (1990) and Maly (1987) provide useful descriptions of the steps involved in IC fabrication. Some of the important steps have already been described (e.g., the CVD of epitaxial Si films and the PECVD of silicon nitride dielectric films). The thermal oxidation of Si to form

passivating and protecting a-SiO_2 layers is discussed in Chapter 21. Other steps, such as diffusion (Chapter 6) and ion implantation (Section W21.3), are also discussed elsewhere. Therefore, only some additional details and current issues relevant to Si device fabrication are presented here.

Thermal Oxidation of Si. The *thermal oxidation* of Si to form layers of a-SiO_2 is repeated often during the fabrication of Si-based devices. In addition to protecting and passivating the surface of Si, oxide layers are also used as the surface for photoresist deposition, as masks for dopant diffusion, and as buried dielectric layers to isolate components of the device structure. Repeated oxidations of a given Si substrate can be carried out as often as necessary for the patterning of different circuit configurations via the photolithographic process, described later. For example, windows can be opened into an a-SiO_2 layer which can be used as diffusion masks, first for p-type doping into a n-type layer and then for n-type doping into the resulting p-type region in order to fabricate an *npn* transistor. This type of process is illustrated in Fig. W21.15.

The oxide dielectric layers include the thin *gate oxides* separating a metallic gate from, for example, the p-type region of a MOSFET, thicker *field oxides* which isolate transistors from metallic interconnecting wires, and dielectric caps which protect the device from the surrounding environment. Gate oxide thicknesses are typically ≈ 15 to 100 nm and are expected to decrease to the range 3.5 to 4.5 nm, and those of field oxides are ≈ 0.3 to 1 μm. These oxide layers are fabricated via the usual thermal oxidation process or via a plasma deposition process, discussed later. Thin gate oxides often include a region incorporating nitrogen (i.e., an oxynitride layer), which serves to suppress diffusion of boron from the polysilicon gate into the MOSFET channel.

The Si/a-SiO_2 interfaces can be prepared to be atomically or chemically abrupt, at least to within 0.5 nm, the dimensions of an $Si-O_4$ tetrahedron, and are flat on the scale of hundreds of nanometers. Nevertheless, the actual width of the interface (i.e., the region in which the properties of the Si and a-SiO_2 differ from their bulk

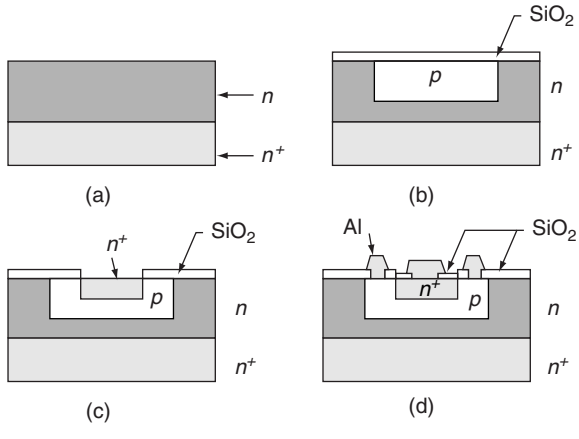

Figure W21.15. Fabrication of an *npn* transistor involving repeated oxidation, lithographic, and diffusion processing steps. In the case shown windows are created in an a-SiO_2 layer which can then be used as diffusion masks, first for p-type doping into a n-type layer and then for n-type doping into the resulting p-type region. (From B. Sapoval et al., *Physics of Semiconductors*, Springer-Verlag, New York, 1993.)

values) has been found to be ≈ 3 nm from sensitive core-level spectroscopies which can determine the strain in Si–O–Si bonding units. The properties of these interfaces are critically important for the operation of devices, and their physical and chemical structures and properties are discussed in Section 20.11.

Lithography. Optical lithography (i.e., *photolithography*) involves the patterning of two-dimensional circuits or designs onto Si wafers by means of the passage of light through a mask that corresponds to the outline of the desired circuit. This is illustrated in Fig. W21.16 and consists of the following sequence of steps:

1. A uniform a-SiO_2 layer is deposited onto the Si.
2. The a-SiO_2 layer is then covered by a layer of photosensitive polymeric material known as a *photoresist*. The photoresist is applied as a uniform liquid layer, using a spin-on procedure that is discussed in Section W21.24, and is then solidified via the application of heat.
3. The photoresist undergoes polymerization or cross-linking during exposure to light through a mask; this is the photoresist development step.
4. In the case illustrated involving the use of a negative photoresist, the unilluminated and hence unpolymerized areas of photoresist are removed via etching with an appropriate chemical solvent.
5. The exposed a-SiO_2 pattern is removed via etching using an acid that does not attack the polymerized photoresist.
6. The polymerized photoresist is finally removed via another suitable chemical solvent.

The patterned a-SiO_2 layer that remains on the surface can act as an insulating layer in the structure or can be used as a diffusion barrier in a subsequent processing step. The predominant method of photoresist removal is currently the use of oxygen plasmas which are described later in the discussion of etching processes.

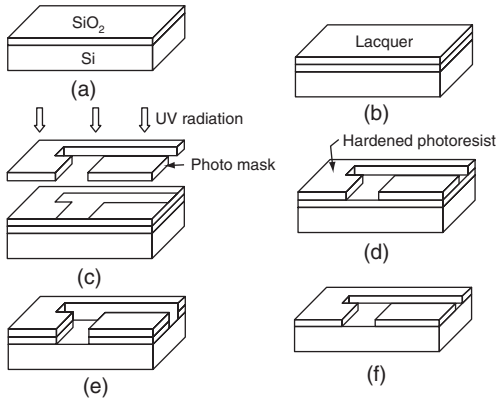

Figure W21.16. Optical lithography process involving the patterning of two-dimensional circuits or designs onto wafers through the use of light passing through a mask. (From B. Sapoval et al., *Physics of Semiconductors*, Springer-Verlag, New York, 1993.)

The interaction of light with photoresist materials such as the high-molecular-weight polymer polymethylmethacrylate (PMMA, also known as Plexiglas or Lucite) is discussed in Section 14.10. The light-induced breaking of bonds (i.e., photodissociation) in the long polymeric chains in the illuminated portions of the PMMA photoresist layer renders these regions susceptible to removal via etching. There are two types of photoresists in use: *negative photoresists*, which undergo light-induced cross-linking and so become insoluble and harder to remove after illumination, and *positive photoresists* like PMMA, which undergo light-induced chain breaking and so become more soluble and easier to remove after illumination. While negative photoresists are usually more photosensitive than positive photoresists and require less illumination, they have lower resolution and hence their use is not desirable in high-density ICs. PMMA is the photoresist with the highest-known resolution.

As the dimensions of features in ICs continue to decrease below 0.25 μm, optical lithography using UV light (e.g., the ArF laser line at $\lambda = 193$ nm) may no longer be possible since the minimum size of a feature is controlled by diffraction effects that limit the definition of the image to about one-half of the wavelength of the light used. The resolution limit D is given by

$$D = \frac{\lambda}{2 \sin \theta}, \tag{W21.20}$$

where θ is the angle subtended by the mask opening at a point on the surface and $\sin \theta$ is the numerical aperture (NA). For an opening of width w that is a height H above the substrate, $\tan \theta = w/2H$. The corresponding depth of focus, h, is given by

$$h = \frac{\lambda}{\sin^2 \theta}. \tag{W21.21}$$

Another important length scale governing the exposure depth is $1/\alpha$, the inverse of the absorption coefficient of the light in the photoresist.

Nanolithographic technologies (i.e., technologies with the higher resolution needed for producing geometrical circuit features with sizes below ≈ 0.1 μm) are based on shorter-wavelength beams of electrons or x-rays, or on the use of scanning probe microscopies such as scanning tunneling microscopy (STM) and atomic force microscopy (AFM). These advanced technologies are being explored as alternatives to optical lithography. Electron beams have the advantages of being able to be steered and focused rapidly using electric and magnetic fields. There are as yet no suitable photoresist materials for features smaller than 0.1 μm.

In the *LIGA process* (*lithographie galvanoformung abformung*), synchrotron radiation is employed to expose the photoresist polymer PMMA. Exceptionally sharp walls are produced, resembling steep cliffs. Metallization of the structure can even result in excellent molds from which replicas may be cast.

Diffusion. The thermal diffusion of dopants into a device in order to create junctions between n- and p-type regions, or just to change the electrical resistivity of a region, occurs repeatedly during device fabrication. Since solid-state diffusion is discussed in Chapter 6, only some details relevant to Si device fabrication are mentioned here.

Due to the need to limit the region of doping in the substrate, all diffusion processes are preceded by oxidation and mask-patterning lithographic steps. Layers

of a-SiO_2 serve as good mask materials for diffusion processes due to the low diffusion coefficients of typical dopants in the oxide. At typical diffusion temperatures of $T = 900$ to 1100°C, dopants present in a source at the Si surface will diffuse through the opening in the mask into the Si both vertically (i.e., normal to the surface), and laterally.

Two methods of dopant diffusion are typically used, constant-source diffusion or two-step diffusion. In the first method, used when shallow junctions are desired, a thick layer consisting of a mixture of B_2O_3 or P_2O_5 and SiO_2 is deposited onto the surface. This layer acts as a constant source of dopant atoms, so the dopant concentration at the surface remains essentially constant as diffusion occurs deeper and deeper into the substrate (see Fig. W6.2). The second method, used when deeper junctions are desired, starts with a predeposition step which is essentially the same as the constant-source method. After removal of the dopant source from the surface, a second, high-temperature step is used to drive the dopant atoms farther into the substrate (see Fig. W6.1).

Complicating the diffusion of acceptors such as B in Si are the effects known as *oxidation-enhanced diffusion* (OED) and *transient-enhanced diffusion* (TED). OED and TED both result from the injection of excess Si interstitials into the Si substrate and away from the Si/a-SiO_2 interface in the case of OED and out of a damaged ion-implanted layer in the case of TED. Dopants such as B must pair with defects such as vacancies or interstitials to move through the lattice, and as a result, their diffusion is affected by the motion of excess interstitials.

Ion Implantation. Ion implantation is used as an alternative to the introduction of dopants by diffusion in IC fabrication when the high temperatures associated with diffusion cannot be tolerated. In addition, the lateral spreading of dopants associated with the diffusion process is minimized when ion implantation is used, a significant advantage in high-density devices. As with diffusion, implantation occurs through a mask and extends into the Si for a characteristic distance known as the *range*. The mask is an opening in an a-SiO_2 overlayer or any other overlayer (metal, photoresist, etc.). Some of the important aspects of ion implantation are discussed in Section W21.3. The *dose* and *energy* of the implanted ions determine the doping level and the position of the resulting junction within the implanted Si. When desirable, implantation through a thin overlayer is possible as long as the incident ions are sufficiently energetic. A schematic phase-space map of the typical ion energies (in electron volts) and ion beam currents (in particle-amperes) used in semiconductor processing is illustrated in Fig. W21.17.

The lattice disorder created in the Si by the incident energetic ions can lead to dopant deactivation when the dopant atoms do not enter the lattice substitutionally or when traps are generated. A subsequent annealing step must then be carried out to repair the damage and for dopant activation.

When plasmas are used to excite the species to be implanted, the process is known as *plasma-immersion ion implantation* (PIII). In this method the substrate is immersed directly in the plasma, and rather than using accelerated beams of energetic dopant ions, high fluxes of relatively low-energy dopant ions are instead extracted from the plasma by applying pulsed high negative voltages, ≈ 2 to 4 kV, to the substrate. When PIII is used to form shallow p^+-n junctions, the n-type Si substrate is first converted to amorphous Si by using SiF_4 in the plasma, followed by the introduction of BF_3 to

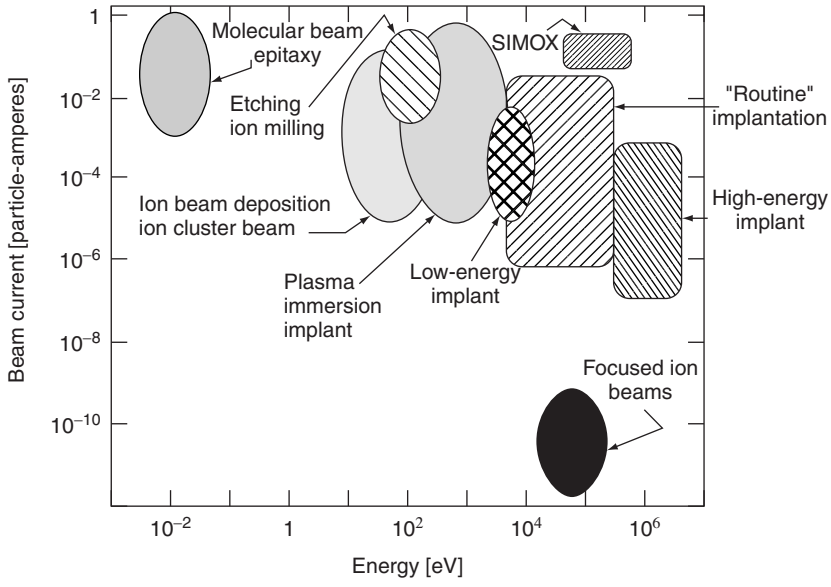

Figure W21.17. Schematic phase-space map of the typical ion energies (in electron volts) and ion beam currents (in particle-amperes) used in semiconductor processing. (From E. Chason et al., *J. Appl. Phys.*, **81**, 6513 (1997). Copyright 1997 by the American Institute of Physics.)

the plasma to implant B ions into the a-Si. An extremely shallow junction depth of 80 nm can be achieved following thermal activation of the dopant atoms using rapid thermal annealing of the implanted region at $T = 1060^{\circ}C$ for 1 s. The PIII process for dopant implantation is similar to the plasma carburizing and nitriding processes used to modify the surface properties of metals, as discussed in Section W21.13.

In the process known as *separation by implantation of oxygen* (i.e., SIMOX) a buried dielectric layer is created below the surface of a Si substrate via the implantation of oxygen ions. This process is a major candidate for the creation of Si-on-insulator (SOI) structures in which devices are isolated by being surrounded completely by an insulator rather than by using a reverse-biased *p-n* junction. The O^+ implantation consists of a high dose, $\approx 2 \times 10^{18}$ cm^{-2}, of ions, which leads to the formation of a continuous buried a-SiO_2 layer following an annealing step for 3 to 5 h at $T = 1100$ to $1175^{\circ}C$. The characteristic distance of the buried layer from the Si surface is 0.3 to 0.5 μm when O^+ ion energies of 150 to 180 keV are used.

Chemical and Physical Vapor Deposition. A variety of *chemical* and *physical* vapor deposition procedures are used to deposit the conducting, semiconducting, and insulating layers that are needed in device fabrication. Reactions between the incident vapor species and the substrate are not necessarily required to grow the desired films in these CVD and PVD procedures. As an example, a-SiO_2 layers must be deposited via PECVD when this dielectric layer is to be grown on a metallic layer instead of on Si. The CVD of epitaxial Si layers and the PECVD of the silicon oxide, nitride, and oxynitride layers used as dielectrics for interlevel isolation, for passivation, and as gate insulators have already been discussed. Si epilayers can be deposited on Si substrates with differing doping levels (e.g., an *n*-type Si epilayer deposited onto an

n^+ Si substrate). PVD in the form of electron-beam evaporation or sputtering is used for the deposition of Al layers.

A challenging problem is the deposition of conformal layers (i.e., layers of uniform thickness) on nonplanar substrates having steps, trenches, and holes. Examples of reliability problems in devices due to deposited layers with nonuniform thicknesses include inadequate electrical isolation in dielectric layers and nonuniform current densities in conducting layers, leading to enhanced electromigration in the conductors and hence open circuits. In the case of a-SiO_2 deposition, when mixtures such as $SiH_4/Ar/N_2O$ or $SiH_4/Ar/O_2$ are used, the sticking coefficients for SiH_n species are high, with the result that the a-SiO_2 layers tend not to be conformal. A method for obtaining conformal a-SiO_2 layers is plasma deposition using the liquid tetraethoxysilane (TEOS) as the source of the precursor in mixtures with O_2 or O_3 (ozone) and Ar. Oxide depositions using dilute TEOS/O_2 mixtures at $T = 200$ to 300°C result in lower deposition rates, < 50 nm/min, compared to SiH_4-based depositions, but the resulting layers have good conformality, due to the low sticking coefficients and higher surface mobility of the TEOS-based precursors.

Metallization. Aluminum and Al alloys have been the metals of choice for providing the electrical connections between circuit elements in ICs due to their desirable physical and chemical properties (e.g., excellent electrical conductivity, the ability to form both ohmic and Schottky barrier contacts to Si, good bonding and adherence to both Si and SiO_2 and also to diffusion barriers such as TiN and Ti, the ability to be patterned in Cl-based plasmas, and the ability to form a stable oxide, Al_2O_3, when exposed to air). Aluminum alloyed with 0.5 wt % Cu exhibits higher hardness and good electrical conductivity, along with improved resistance to electromigration, a process described in Section 12.9. The resistance to electromigration resulting from alloying Al with Cu is attributed to the precipitation of Cu at grain boundaries. This inhibits the harmful grain-boundary diffusion of Al, which leads to vacancy accumulation and void formation in the Al connecting lines. Even though Cu itself has low electrical resistivity and good resistance to electromigration, it has not been widely used so far as an interconnect metal because a successful dry-etching process has not been developed for patterning the Cu lines. In addition, diffusion barriers must be used between Cu lines and Si because Cu impurity atoms act as deep traps in Si.

Problems with Al layers deposited by PVD methods such as electron-beam evaporation and dc magnetron sputtering are associated with incomplete filling of vias and with poor step coverage for feature sizes below 0.5 μm. Other possible deposition procedures that may lead to improved via filling and step coverage include high-temperature Al-alloy sputtering processes, the use of Al reflow processes, and CVD at $T = 100$ to 200°C using Al-containing metal–organic molecules at deposition rates of 100 to 200 nm/min. Aluminum reflow processes involve the use of elevated deposition temperatures or postdeposition annealing to allow the deposited Al alloy to flow into and fill via/contact holes. The Al-alloy reflow temperatures lie below the alloy melting points by ≈ 150°C, with both temperatures decreasing with increased alloying of elements such as Cu or Ge.

The refractory metal W can be selectively deposited via CVD and allows much better step coverage and via and hole filling than Al. In addition, it exhibits excellent resistance to electromigration. Bilayers of Ti and TiN serve as *diffusion barriers* between W and Si and also as intermediate layers for the CVD of W. The initial Ti

layer is reacted with the underlying Si at $T \approx 700^{\circ}C$ to form a titanium silicide Ti_xSi_y phase with both good electrical conductivity and contact to the underlying Si. A TiN_x diffusion barrier layer is then deposited to prevent undesired reactions between the Ti_xSi_y layer and the fluorine involved in the CVD of W via the hydrogen reduction of the WF_6 precursor [i.e., $WF_6(g) + 3H_2(g) \rightarrow W(s) + 6HF(g)$]. When selective deposition of W and lower deposition temperatures are required, the silane reduction of WF_6 can be used [e.g., $2WF_6(g) + 3SiH_4(g) \rightarrow 2W(s) + 3SiF_4(g) + 6H_2(g)$].

Local interconnects formed from low-resistivity doped polycrystalline Si layers are useful because these layers can make good electrical contact to Si substrates and can also serve as diffusion barriers between Si and Al lines. Electrical contacts between pure Al and n^+ and p^+ Si are not stable at processing temperatures in the range $T = 350$ to $500^{\circ}C$, due to the solubility of Si in Al and also to the rapid diffusion of Si into the polycrystalline Al contacts. The reciprocal diffusion of Al into the Si layer can lead to the *spiking* (i.e., shorting) of shallow junctions. The use of polysilicon is restricted to buried contacts and to limited regions due to its relatively high sheet resistance of 20 to 30 Ω/square.

Etching Processes. Device fabrication involves a variety of processing steps employing the etching or controlled removal of material from the surface of the wafer. The etching or stripping process can employ either wet, liquid-phase or dry, gas-phase etchants. *Chemical etching*, in which the etchant reacts with the material to be removed, can occur in either the liquid or gas phases, is typically highly selective, and is isotropic (i.e., the etching occurs at the same rate in all directions). *Physical etching* is a gas-phase process in which material is removed by sputtering (i.e., via energy and momentum transfer from incident ions), is less selective than chemical etching, and is typically anisotropic (i.e., etching occurs preferentially in one direction). *Selectivity* refers to the ability of the etching process to remove some materials but not others. An example is positive-photoresist lithography, where liquid solvents etch away the illuminated portion of the photoresist while the unilluminated portion is unaffected, or as when an HF acid etch is used to remove a-SiO_2 but neither Si nor photoresist.

A plasma etching process with both chemical and physical components is *reactive-ion etching* (RIE), in which ions created in a plasma react with and also transfer kinetic energy to the material to be etched. An advantage of RIE is that it can be both selective and anisotropic. Plasma etching is used for the removal of Si, of a-SiO_2 and silicon nitride, of metals, and of photoresist. Appropriate etching species are chosen for each case: for example, F atoms and Ar^+ ions for etching Si or polysilicon (forming SiF_4) and O atoms for etching or stripping photoresist (forming CO, CO_2, and H_2O). The Ar^+ ions provide additional kinetic energy, which can greatly increase the yield of the etching process by enhancing chemical etching reaction rates on the surface. For example, a 1-keV Ar^+ ion can result in the removal of up to 25 Si atoms when a flux of F atoms is also incident on the surface. The use of Ar^+ ions can also increase the anisotropy of the etching but may decrease the etching selectivity.

Etch inhibitors are also used in RIE to prevent etching from occurring outside the area exposed to the ion beam. An example is the anisotropic etching of trenches and holes in Al using CCl_4/Cl_2 mixtures, where the CCl_4 molecules are the inhibitor precursors. A protective, etch-inhibiting amorphous chlorocarbon film is present on the areas of the Al surface not exposed directly to the ion beam, including on the sidewalls

of the features being etched. The presence of C in the etching mixture thus leads to an enhancement of the anisotropic etching of the desired trenches and holes.

Reactive-ion etching rates are very difficult to predict. This is due to difficulties associated with modeling the plasma processes giving rise to the incident fluxes of reactive atomic and molecular radicals and ions on the surface. There are also difficulties with modeling the many surface processes, including adsorption, diffusion, reaction, and desorption, involved in the generation of etching products. In addition, in the F etching of Si, a fluorinated SiF_x surface layer two to five monolayers thick is present and the diffusion of the etching species, F^- ions, through this layer plays an important role in the process. A rough estimate for the characteristic thickness of this layer is $d \approx D/R_e(\text{Si})$, where D is the diffusion coefficient for F^- ions in the surface layer and $R_e(\text{Si})$ is the etching rate in m/s.

The etching of Si by halogen atoms such as F and Cl is found to depend on the doping level and type of the Si substrate, with etching rates of n-type Si exceeding those of p-type Si by a factor of about 2 for F and by many orders of magnitude for Cl. These observations indicate that the position of the Fermi level and the concentrations of charge carriers near the Si surface can play important roles in the etching process. The current model is that electrons in n-type Si tunnel from the bulk through the SiF_x layer, leading to the formation of F^- or Cl^- ions that attack Si–Si bonds in either the surface layer or the bulk. Molecules such as CF_4 are typically used as etching precursors because the etching of Si by F_2 leads to roughening the surface through pitting. The overall etching reaction in this case can be written as

$$4CF_4 + Si \longrightarrow SiF_4 + 2C_2F_6. \tag{W21.22}$$

When wet chemical etching is used to remove an unprotected a-SiO_2 layer, the isotropic nature of the etching can cause unwanted undercutting of the oxide beneath the protective photoresist mask. As a result, the pattern obtained is not the one desired. Dry etching carried out at reduced pressures in the gas phase can combine the advantages of chemical etching in being selective and physical etching in being anisotropic, so that no undercutting of the oxide occurs.

The smallest feature size (e.g., the minimum trench width) that can be obtained via etching is

$$w \approx \frac{2d}{a_h}, \tag{W21.23}$$

where d is the depth of the trench and $a_h = R_{ev}/R_{eh}$ is the ratio of the vertical and horizontal etch rates of the material in which the trench is being etched. As an example, 0.2-μm-wide and 4-μm-deep trenches with the aspect ratio $d/w = a_h/2 = 20$ can be etched into single-crystal Si using F-based chemistry.

Remaining problems associated with the use of plasmas in device fabrication are related to ion-induced damage and plasma-induced contamination.

Annealing. Annealing at elevated temperatures is often required in IC fabrication for a variety of purposes:

1. To remove, or at least minimize, processing-induced defects (e.g., those created in the Si lattice during ion implantation).

2. To activate implanted dopants in Si or polysilicon following ion-implantation procedures.
3. To drive dopant atoms farther into the Si following their implantation in a shallow layer.
4. To promote the reactions between deposited metals such as Ti and the underlying Si in order to form desired silicides.
5. To deactivate deep trap-generating impurities such Cu and Fe via gettering, a process in which these impurities diffuse to and are immobilized in the strain fields of extended defects such as oxide precipitates or dislocations. In this way the traps are removed from the active area of the device.

The time and temperature of an anneal must be chosen so that unwanted dopant redistribution does not occur. Any exposure of the device to high temperatures must therefore be as brief as possible. A method for limiting the annealing time is the process of *rapid thermal annealing* (RTA), also known as *rapid thermal processing* (RTP). A typical RTA dopant drive-in procedure involves a rapid temperature increase to $T = 1050$ to 1150°C, a 10-s anneal, and a rapid decrease to temperatures at which diffusion is negligible.

W21.9 Processing of Microelectromechanical Systems

The fabrication of Si-based microstructures for use in *microelectromechanical systems* (MEMS) having typical dimensions ≈ 1 to 100 μm is an exciting new area of materials research.[†] In addition to its well-known and extremely versatile electronic properties, crystalline Si also possesses very useful mechanical and thermal properties, such as high durability, elasticity, and thermal conductivity, which can be exploited in very small electromechanical structures. With the development of MEMS, Si semiconductor device-fabrication technology can now also be exploited in sensors and actuators for measurement and control in the fields of thermodynamics, optics, magnetism, acoustics, and hydrodynamics. Besides Si, other materials used in MEMS include a-SiO_2, crystalline quartz, and other ceramics, such as SiC. Since MEMS technology is in a state of rapid development, only a brief survey is given here.

The fabrication of MEMS is involved primarily with the processing of Si wafers into the desired final forms using a variety of etching and micromachining procedures. These processing procedures currently include the following:

1. Anisotropic wet chemical etching, usually in KOH solutions
2. Dry etching (i.e., reactive-ion etching) with the etchant activated via plasma excitation
3. Surface micromachining involving the removal of a sacrificial layer of a-SiO_2 or porous Si via etching in HF
4. Porous Si technology, also involving surface micromachining but using much thicker sacrificial layers of porous Si, up to hundreds of micrometers thick

[†] A recent review article is W. Lang, *Mater. Sci. Eng.*, **R17**, 1 (1996).

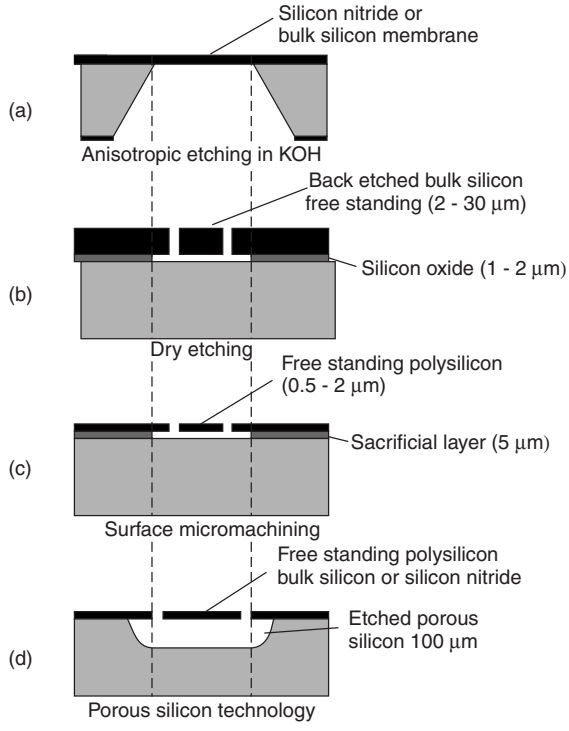

Figure W21.18. Micromachining processes currently used to fabricate microelectromechanical systems (MEMS) from Si wafers: (*a*) anisotropic wet chemical etching; (*b*) dry etching or reactive-ion etching; (*c*) surface micromachining involving a sacrificial layer of a-SiO_2; (*d*) porous Si technology, also involving surface micromachining but with much thicker sacrificial layers of porous Si. [Reprinted from W. Lang, *Mater. Sci. Eng.*, **R17**, 1 (1996). Copyright 1996, with permission from Elsevier Science.]

Examples of these processes are shown in Fig. W21.18. Free-standing features (e.g., Si cantilevers) are readily produced. The key to the rapid growth of MEMS technology is that most of these procedures involve deposition, lithography, and etching processes that have already reached an advanced level of development in Si electronic device fabrication. Porous Si, however, is a relatively new material consisting of variable volume fractions of crystalline Si filaments or wires and of empty pores, which is prepared by electrochemical anodic etching or anodization of crystalline Si in HF (see Fig. W11.9). The use of thick porous Si in MEMS is also compatible with Si device-fabrication techniques.

While Si electronic devices are essentially planar, containing circuit elements with typical thicknesses ≈ 1 μm, Si electromechanical devices or MEMS are truly three-dimensional and often contain free-standing structures such as cantilevers and bridges. The current trend in MEMS is to include several Si-based electronic devices and mechanical sensors and actuators in a single MEMS. The most widely used Si MEMS sensors at present are pressure transducers and thermopile radiation detectors. Other MEMS include micromotors, micromirrors in optical switches, accelerometers, microvalves, and flow sensors. In the future, MEMS actuators may be used to move STM tips in three dimensions as part of data storage systems at the near-atomic level.

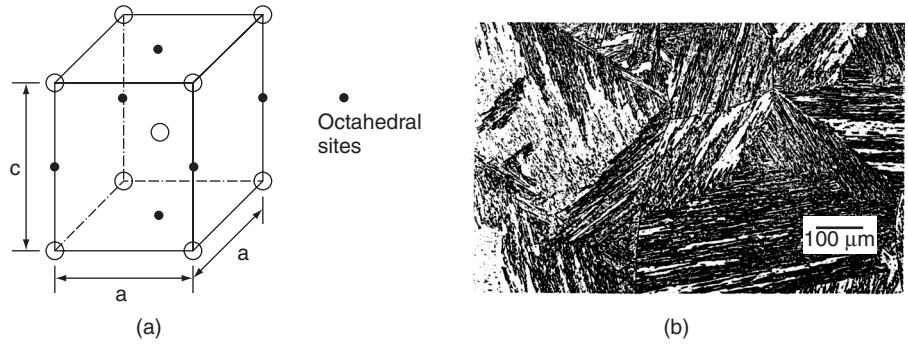

Figure W21.19. Martensite is a supersaturated solid solution of interstitial C in Fe. (*a*) Body-centered tetragonal (BCT) unit cell of martensite. The Fe atoms are actually displaced from their normal lattice sites to accommodate the C atoms in the octahedral sites. (*b*) Lath microstructure of martensite in a Fe–2Mn–0.03C wt % steel. (From ASM Handbook, 9th ed., Vol. 9, *Metallography and Microstructures*, ASM International, Materials Park, Ohio, 1985, p. 670.)

W21.10 Synthesis and Processing of Steels

While the simplest steels are just Fe–C alloys, steels in general can be very complex materials in both composition and microstructure. This complexity makes the design of a steel with a given set of properties quite challenging. It is useful first to review how the complex phases that may be present in steels are related to the simpler phases of pure Fe and Fe–C compounds and alloys.

Nonequilibrium Multicomponent Phases in Steels. The various nonequilibrium, multicomponent phases of Fe and Fe-based alloys and compounds which are the identifiable components of a wide variety of steels are described briefly next. These phases are all formed from the transformation or decomposition of *austenite* as the steel is cooled below the eutectoid temperature and include *pearlite*, *bainite*, *martensite*, and *acicular ferrite*. Table W21.5 summarizes the properties of these important phases and also of their multicomponent mixtures, which are found in the steels commonly used today.

Pearlite. *Pearlite* is a coarse, lamellar eutectoid mixture consisting of alternating layers of cementite and ferrite, shown in Fig. 21.11, which results from the decomposition of austenite as its temperature is lowered below $T_e \approx 727^\circ$C. Along with ferrite, it is a very common constituent of a broad range of steels in which it makes a substantial contribution to the strength of these materials. Pearlite also reduces the ductility and toughness of steels since cracks can nucleate at the ferrite–cementite interfaces.

The diffusion of C atoms is usually assumed to be the rate-controlling step for the nucleation and growth of pearlite in austenite. This is essentially a high-temperature reaction that occurs between T_e and $T \approx 550^\circ$C. Nucleation can take place at a variety of sites, including at austenite grain boundaries as well as on ferrite and cementite phases when they are already present in the austenite. At low transformation temperatures where the diffusion of C is slower, the lamellar spacing is much smaller and the resulting material is known as *fine pearlite*. The spacing of the lamellae in pearlite

TABLE W21.5 Important Phases of Fe, Fe–C Compounds and Alloys, and Their Multicomponent Mixtures Found in Steels

Phase	Structure and Description[a]	How Phase Is Obtained
	Equilibrium Phases of Pure Fe	
α-Fe (ferrite)	BCC, $a = 0.286$ nm at $T = 20^\circ$C; stable up to $T = 912^\circ$C; $T_C = 769^\circ$C	Stable phase at STP
γ-Fe (austenite)	FCC, $a = 0.364$ nm at $T = 912^\circ$C	Stable phase for $912 < T < 1394^\circ$C
δ-Fe (δ-ferrite)	BCC, $a = 0.293$ nm at $T = 1394^\circ$C; $T_m = 1538^\circ$C	Stable phase for $T > 1394^\circ$C
	Equilibrium Fe–C Compound	
Fe_3C (cementite)	Orthorhombic, $a = 0.509$, $b = 0.674$, $c = 0.452$ nm; a complex interstitial compound	Present in Fe–C alloys under conditions of metastable equilibrium (see Fig. 21.9)
	Equilibrium $Fe_{1-x}C_x$ Alloys	
α-Fe–C (ferrite)	Solubility limit of C in α-Fe at $T = 27^\circ$C: $x = 1.2 \times 10^{-6}$ (0.00012 at % or 1.2 ppm)	Present in Fe–C alloys under equilibrium conditions (see Fig. 21.9)
γ-Fe–C (austenite)	Solubility of C in γ-Fe at $T = 1150^\circ$C: $x \approx 0.09$ (9 at %)	Present in Fe–C alloys under equilibrium conditions (see Fig. 21.9)
	Nonequilibrium Multicomponent Phases	
Pearlite	A coarse, lamellar form of cementite in ferrite; a eutectoid structure	Formed between $T = 720$ and 550°C during cooling of austenite
Bainite	An intermediate structure composed of fine aggregates of ferrite plates (laths) and cementite particles	Formed between $T = 550$ and $\approx 250^\circ$C during cooling of austenite
Martensite	BC tetragonal, $c/a = 1 + 0.045$ wt % C; a supersaturated solid solution of interstitial C in ferrite, having a lath or lenticular microstructure	Rapid quenching of austenite to keep C in solution; formed between $T \approx 250^\circ$C and room temperature or below
Acicular ferrite	A disorganized structure of randomly oriented ferritic plates in a matrix such as martensite	Nucleation of ferrite at small, nonmetallic inclusions during cooling of austenite

[a]The range of thermal stability is given at $P = 1$ atm.

is larger at higher transformation temperatures due to the enhanced diffusion of C, with the resulting material known as *coarse pearlite*. The spacing is also controlled in part by the competition between the decrease in free energy associated with the more stable phase and the increases in surface energy associated with the interfaces between the ferrite and cementite lamellae and of any strain energy associated with the transformation.

Bainite. The term *bainite* refers to the intermediate structures found in steels, which are composed typically of fine aggregates of ferrite plates or laths and cementite particles. Bainite is formed at intermediate temperatures ($T \approx 250$ to 400°C for lower bainite and $T = 400$ to 550°C for upper bainite), below those at which pearlite ($T = 550$ to 720°C) is formed and above those at which martensite is formed (typically from room temperature up to $T \approx 250$°C). Bainite can also be formed when austenite is cooled too rapidly for the diffusion of C required for the formation of pearlite to occur and too slowly for martensite to be formed. Depending on the contents of C and of other alloying elements, the bainitic microstructure can be quite complicated, with austenite and martensite replacing cementite. There is a start temperature T_{Bs} for the austenite-to-bainite transition, with the amount of bainite that can be formed, increasing as T is lowered below T_{Bs}. The TTT diagram shown in Fig. 21.12 illustrates the formation of bainite at intermediate temperatures. Upper bainite is favored in low-carbon steels, while lower bainite is favored in high-carbon steels.

Martensite. *Martensite* is a supersaturated solid solution of interstitial C in Fe formed via the rapid quenching of austenite, which prevents the diffusion of C that would result in the formation of cementite. The body-centered tetragonal (BCT) crystal structure of martensite is shown in Fig. W21.19*a*. Carbon atoms are randomly distributed in the six equivalent octahedral interstitial sites at the midpoints of the edges along the c axis and in the centers of the basal faces. The lattice parameters of the BCT martensite unit cell depend on the C composition according to $a_{\text{mar}} = (0.286 \text{ nm})(1 - 0.0035 \text{ wt \% C})$ and $c_{\text{mar}} = (0.286 \text{ nm})(1 + 0.041 \text{ wt \% C})$, resulting in $c_{\text{mar}}/a_{\text{mar}} = (1 + 0.045 \text{ wt \% C})$. The lattice constant $a = 0.286$ nm of α-Fe has been used here for the zero-carbon limit.

The corresponding lath microstructure of martensite (Fig. W21.19*b*) can appear in a matrix of ferrite or pearlite. The martensitic transformation, known as a *diffusionless transformation*, involves the rapid appearance of shear strain in the FCC austenite lattice. The result is a change in shape of the unit cell from cubic to tetragonal. The preferential occupation of the octahedral sites by the C atoms distorts the structure, thus determining the c axis of the resulting BCT crystal structure. High densities of dislocations and also slip and twinning can occur in the martensite during its formation. Similar martensitic transformations or reactions occur in other alloys, such as Fe–Ni, In–Tl, and the shape-memory alloys discussed in Chapter W12.

The decomposition of metastable austenite to form martensite usually occurs over a well-defined range of temperatures, beginning at the *martensitic start temperature* T_{Ms} (often written as M_s), which ordinarily lies in the range from $T \approx 250$°C to below room temperature. Additional martensite is formed as the temperature is lowered further below T_{Ms}, until most of the austenite has been converted to martensite at the *finish temperature* T_{Mf} (or M_f). The transformation is an *athermal* one (i.e., it is not thermally activated and occurs essentially instantly once a nucleus of martensite is formed). Thus there is no time delay for the formation of martensite on the TTT diagram as observed for the formation of pearlite or bainite. The amount of austenite converted to martensite depends only on temperature and not on the time allowed for the transformation. Both T_{Ms} and T_{Mf} are lower when the austenite phase in the steel has been stabilized by carbon or other alloying elements. The cooling must occur rapidly enough so that the metastable austenite does not transform instead to ferrite, pearlite, or bainite at temperatures between T_e and T_{Ms}.

The actual microstructure present in a quenched steel will often exhibit spatial variations from the surface into the bulk, due to the fact that the cooling rate and temperature will be different at different depths within the sample. This is certainly the case in rapidly solidified steels, as discussed later.

Rapidly quenched steels that have both enhanced hardness and brittleness due to the formation of martensite from austenite are said to have good *hardenability*. The strength of the steel due to the martensite is enhanced as the C content is increased and can result from a variety of strengthening mechanisms, several of which are described later. When a martensitic steel is reheated so that the C can diffuse, the martensite will be transformed into other phases, such as pearlite and bainite. This process, known as *tempering*, is also described.

The cooling rates needed to transform a given steel completely to martensite can be determined from another type of temperature–time diagram, the continuous-cooling transformation (CCT) diagram shown in Fig. W21.20. This diagram provides information concerning the kinetics of the transformation which is not obtainable from the

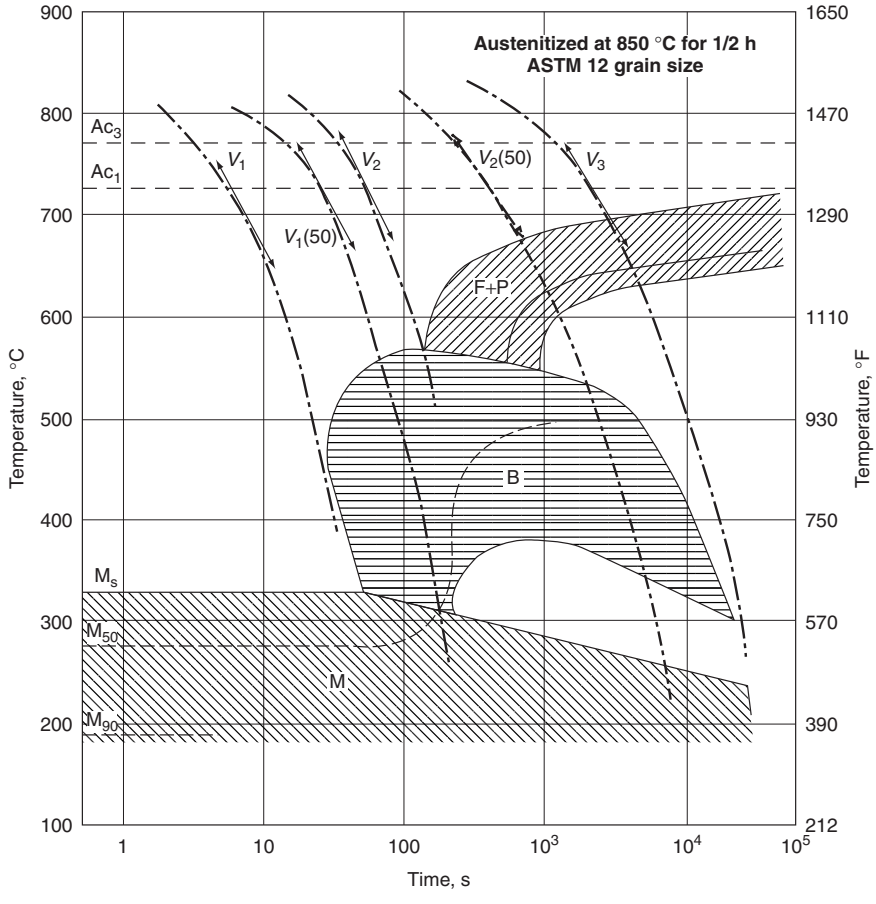

Figure W21.20. The cooling rates needed to transform a given steel completely to martensite (M) can be determined from the continuous-cooling or CCT diagram, shown here for 30 NC11 steel. The ferrite (F), pearlite (P), and bainite (B) phase regions are also shown. (From ASM Handbook, 9th ed., Vol. 4, *Heat Treatment*, ASM International, Materials Park, Ohio, 1991, p. 26.)

Figure W21.21. Coarse acicular ferrite, a disorganized structure of randomly oriented ferritic plates, is shown in a weld zone along with polygonal ferrite. The horizontal bar corresponds to 20 μm. (From ASM Handbook, 9th ed., Vol. 9, *Metallography and Microstructures*, ASM International, Materials Park, Ohio, 1985, p. 585.)

isothermal TTT diagram shown in Fig. 21.12. In the CCT diagram the ferrite, pearlite, and bainite phases are shown in addition to martensite.

Acicular Ferrite. *Acicular ferrite* is a nonequilibrium phase that has superior mechanical properties, including toughness, and consists of a disorganized structure of randomly oriented, interlocking ferritic plates in a matrix such as martensite. This phase can be obtained via the incorporation of small, nonmetallic inclusions that serve as nucleation sites for the plates. It can also appear in weld zones (Fig. W21.21). The morphology of this phase is three-dimensional since the ferritic plates can nucleate and grow in several different directions around an inclusion. Whether bainite or acicular ferrite is formed in a given steel as austenite is cooled depends on the ratio of nucleation sites at austenitic grain boundaries to those at the surfaces of inclusions, with grain-boundary nucleation leading preferentially to bainite. Ti_2O_3 and other oxide particles have been found to be especially effective in nucleating acicular ferrite, with the exact mechanism remaining unknown.

Processing Treatments for the Strengthening of Steels. A variety of processing treatments are used to strengthen steels and also other metals and alloys (e.g., Al alloys and Ni alloys). Important examples of these processes are given now, and a brief description of the strengthening mechanism is presented for each case. The strength of a given steel often results from contributions from more than one of these mechanisms. In practically every case the strengthening occurs via the pinning of dislocations, as discussed in Chapter 10. The specific application for which a given steel is designed will determine the conditions under which strength is needed (e.g., at high temperatures, under repeated loading, along with good ductility, etc.). Due to the large number of available processing variables, it is not possible to discuss here all of the important processing treatments that can be used to strengthen steels.

Mechanical Work Hardening. The tensile strength of a plain carbon steel that contains no other alloying elements can be increased up to 1500 MPa when it is drawn down

(e.g., to a wire) in a *work-hardening* or *cold-working* process in which its cross-sectional area is reduced by up to 95%. This large increase in strength produced by plastic deformation results from the generation of defects such as dislocations and dislocation arrays which reduce the mobility of other dislocations. The measured shear stress typically arises from two dislocation-pinning mechanisms, one arising from "small" defects, such as isolated dislocations, and the other from "larger" defects, such as dislocation arrays. The former mechanism decreases with increasing T, due to the thermally activated motion of dislocations around small defects while the latter is temperature independent. Work hardening is discussed in more detail in Section 10.13, where the dependence of the shear yield stress τ_y on dislocation density and strain is discussed in detail.

Solid-Solution Strengthening. Steels can also be strengthened or hardened by the presence of *interstitial* or *substitutional* impurities. The strong, attractive interactions between dislocations and the interstitial impurities C and N play an important role in this strengthening mechanism. Since interstitial C and N atoms as well as dislocations produce their own strain fields in the material, the attractive interaction arises from an overall reduction in strain energy when the C and N atoms reside in the strain field of a dislocation. The binding energy of a C atom to a dislocation in Fe is ≈ 0.5 eV. At high interstitial concentrations the resulting distribution of interstitial atoms surrounding the dislocation, known as the *Cottrell atmosphere*, can condense at the dislocation core. The movement of dislocations under the influence of an external stress will clearly be impeded by this interaction since the Cottrell atmosphere of interstitials has the effect of increasing the effective mass or inertia of the dislocation.

The condensation of interstitial atoms near dislocations can occur in steels at temperatures even as low as room temperature, due to the high diffusivity of C and N through defect-free regions of the material. Under applied stress and at higher temperatures, thermal activation of dislocations away from the atmosphere of interstitials can lead to a reduction of the yield strength. The strengthening process known as *strain aging* occurs under an applied stress after the yield point has been reached when interstitial atoms condense on newly generated dislocations.

The martensite structure, formed by rapid quenching, is usually very hard, due primarily to interstitial C and the resulting solid-solution strengthening but also due to the high densities of dislocations caused by the transformation of austenite to martensite. Martensite can, however, be brittle and not very ductile. The process known as *tempering*, (discussed later), is often used to increase its ductility and toughness.

The strengthening resulting from solid solutions of substitutional impurities such as Si, Mn, Cr, and Mo in steels results from the strain introduced into the structure by these impurities and thus is greater for impurity atoms, whose sizes are quite different from that of the host Fe atom. The increase of yield stress $\Delta\sigma_y$ of steel for various interstitial and substitutional impurities is illustrated in Fig. W21.22. The interstitial impurities C and N can be seen to have a much larger effect on σ_y than the substitutional impurities Si, Mn, Mo, and Ni due to the tetragonal distortions introduced into the lattice by C and N. These tetragonal distortions allow the stress fields of C and N impurities to interact with both edge and screw dislocations, while substitutional impurities have spherically symmetric stress fields and so can interact only with edge dislocations. Since substitutional alloying elements are usually added to the steel for other reasons

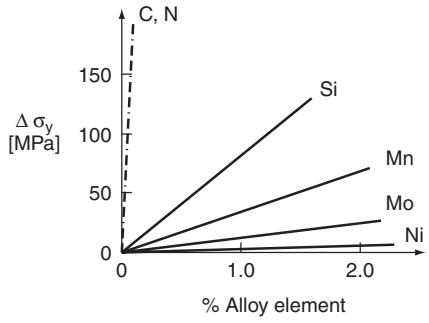

Figure W21.22. Increase $\Delta\sigma_y$ of the yield stress of steel for various interstitial and substitutional impurities. (From ASM Handbook, Vol. 1, *Properties and Selection: Iron, Steels, and High-Performance Alloys*, ASM International, Materials Park, Ohio, 1990, p. 400.)

(e.g., to improve corrosion resistance or to combine with oxygen or sulfur), the increase in strength associated with their presence can be considered a bonus.

Strengthening via Grain-Size Reduction. The reduction of grain size and the resulting increase in the number of grain boundaries are some of the most effective ways of increasing the strengths of steels. The *Hall–Petch relation* between the yield stress σ_y and the average grain size d of a material,

$$\sigma_y(d) = \sigma_0 + \frac{k_y}{\sqrt{d}}, \tag{W21.24}$$

is described in Section 10.14. Here σ_0, the yield stress for a single crystal with no grain boundaries, and k_y are constants that are independent of d for a given steel. The strengthening effect of grain boundaries results from their ability to pin dislocations. Reduction of the grain size in steels into the range 2 to 10 μm can produce yield stresses of over 500 MPa. This reduction is typically achieved via hot rolling and the addition of small amounts of certain alloying elements. The grain size can also be controlled by varying the cooling rate (i.e., the time available for the grains to grow). The kinetics of grain growth in metals are discussed in Section 21.5.

The growth of larger grains can be inhibited by the addition of small amounts, < 0.1 wt %, of grain-refining elements such as V, Al, Nb, and Ti, which form carbides, nitrides (e.g., VC and AlN), or carbonitrides. The 3 to 10-nm carbide and nitride particles that are formed tend to pin grain boundaries, thus helping to prevent grain growth. The resulting steels, which also contain 0.008 to 0.03 wt % C and up to 1.5 wt % Mn, have yield strengths in the range 450 to 550 MPa and are known as high-strength low-alloy (HSLA) steels or micro-alloyed steels.

Dispersion Strengthening. The strengthening of steels through the introduction of more than one structural phase in the ferrite matrix is known as *dispersion strengthening*. The typical phases present in plain carbon steels include carbides such as cementite, nonequilibrium phases such as pearlite, bainite, and martensite, and the precipitates formed by tempering. In alloy steels the thermodynamically more stable carbides of Si, Mn, and V often replace iron carbides. Other possible phases in steels include nitrides, other intermetallic compounds, and graphite.

A simple relation has been developed by Orowan for the yield stress σ_y of an alloy containing a random distribution of spherical particles of a different phase which are impenetrable by dislocations. With an average interparticle spacing Λ, the result is

$$\sigma_y(\Lambda) = \sigma_0 + \frac{2T_L}{b\Lambda}, \tag{W21.25}$$

where σ_0 is the yield stress of the particle-free matrix and T_L and b are the line tension (i.e., energy per unit length) and Burgers vector of a typical dislocation, respectively. An order-of-magnitude estimate for the line tension is $T_L \approx Gb^2 \approx 1.7 \times 10^{-9}$ J/m $\approx$ 10 eV/nm, using $G \approx 82$ GPa as the shear modulus and $b = a/2 = 0.144$ nm for Fe. The term $2T_L/b\Lambda$ is the stress required to move a dislocation past a second-phase particle via bowing. This process leaves a dislocation loop around each such particle. Equation (W21.25) is only approximately valid for steels in which the precipitates are plates or rods. In pearlite where the microstructure consists of a lamellar mixture of cementite and ferrite, the parameter controlling the strength is usually the average size of the uninterrupted ferritic regions, known as the *mean free ferrite path* (MFFP). In this case the flow stress is proportional to $(\text{MFFP})^{-1/2}$, a relationship of the Hall–Petch type [see Eq. (W21.24)]. Thus the fine pearlite formed at lower T will be stronger than the coarse pearlite formed at higher T.

The extent of the dispersion strengthening in a given steel is controlled by the C content, by alloying, and by the processes that determine which phases are present (e.g., heat treatment, tempering, etc.). When steels are quenched in order to form martensite, they are typically very strong but also tend to be quite brittle. Subsequent reheating or tempering of martensitic steels at an intermediate temperature between $T \approx 150$ and 700°C (i.e., below the eutectoid temperature T_e) is used to improve their ductility and toughness without at the same time causing too large a decrease in strength. The tempering process is controlled by the diffusion of carbon, which comes out of the supersaturated solid solution found in martensite and forms finely divided carbide phases. The martensite is thus converted to ferrite and the resulting material is then a dispersion of fine particles of cementite or transition metal (TM) carbides in a ferrite matrix. The formation of TM carbides such as MoC, Mo_2C, WC, W_2C, and VC_x ($x \approx 0.75$) occurs via precipitation and at much higher temperatures, $T \approx 500$ to 600°C, than that of cementite due to the much lower diffusivities in ferrite of these substitutional impurities as compared to that of C. This process, which can involve the conversion of cementite to TM carbides, is known as *secondary hardening* and is a type of age hardening.

Alloying elements such as Ni, Mn, and Si are often added to steels to make them heat treatable (i.e., to facilitate the heat treatment of austenite to produce martensite). This occurs because the formation of pearlite is retarded and so the desired martensite is more easily formed.

When the steel includes a high TM content (e.g., 18 to 25 wt % Ni along with Mo and Ti), particles of intermetallic compounds such as Ni_3Mo and Ni_3Ti can be formed via precipitation. Such materials are known as *maraging steels* and can have very high yield stresses, $\sigma_y \approx 2000$ MPa, along with good ductility and toughness.

The nucleation and growth of particles, often of a second phase, in a matrix is a recurrent theme in steels, especially in the discussion of dispersion-strengthening. This topic is also discussed in Section 21.5, where the Johnson–Mehl equation for the annealing and recrystallization (i.e., grain growth) of metals is discussed.

In addition to their uses in the strengthening processes just described, *heat treatments* of steels are used for a variety of other purposes. Various heat treatments are given to plain carbon steels containing pearlite in order to achieve the desired pearlite microstructures. As an example, *spheroidizing annealing* at just below T_e is used to transform the lamellar pearlite structure into one in which the pearlite takes on a spheroidal microstructure (i.e., the cementite lamellae have been spheroidized). This process leads to improved ductility and machinability of the steel. The driving force for this process is the reduction of the surface energy between the cementite and ferrite phases. This process is similar to the tempering of martensite discussed earlier, which, however, results in much smaller cementite particles, due to the lower temperatures used for tempering.

As just described, *tempering* is the term often used for the heat treatment or annealing of steels to achieve desired changes in microstructure and mechanical properties such as improved ductility. For example, the strength of martensite falls quickly and its ductility improves during tempering, due to the precipitation of C in carbides or carbon-containing intermetallic compounds. In contrast, tempering has little effect on bainite because there is not much C in solid solution. The effects of tempering on the mechanical properties of a steel are illustrated in Fig. W21.23. Similar behavior is observed for the tempering or annealing of nonferrous metals and alloys.

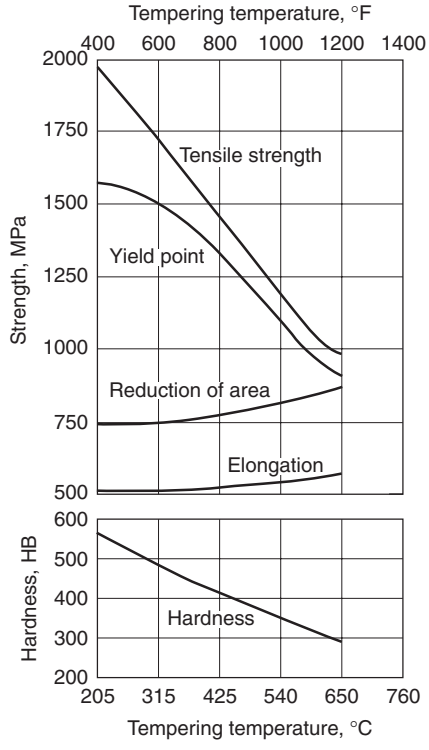

Figure W21.23. Effects of tempering at various temperatures on the mechanical properties (Brinell hardness, tensile and yield strengths, reduction of area, and elongation) for a 4340 steel bar. (From ASM Handbook, 9th ed., Vol. 4, *Heat Treating*, ASM International, Materials Park, Ohio, 1991, p. 123.)

Thermomechanical processing treatments involve the simultaneous use of both heat and plastic deformation to achieve desired changes in both the external shape and the microstructure of a material. The *hot rolling* of steels in the range $T = 1200$ to 1300°C, for example, achieves several purposes: the reduction in cross section of a large steel ingot, the breaking down of the original coarse microstructure in the as-cast material, the reduction of compositional inhomogeneities, and the redistribution of impurities. As hot rolling is carried out at successively lower temperatures, the precipitation of carbides, nitrides, and carbonitrides occurs, leading to the pinning of grain boundaries. As a result, grain refinement (i.e., the achievement of lower average grain sizes) and dispersion strengthening can both occur during hot rolling, leading to significant increases in the yield strength of the steel.

The *welding* of steels to fabricate structural forms is often an unavoidable processing step which can cause unwanted changes in the microstructure and properties of the steel in the vicinity of the weld. Fusion welding involves the melting of the steel in regions near the weld, known as the *fusion zone*, as well as large increases of temperature in surrounding areas known as the *heat-affected zone*. Significant changes in the microstructure of the steel can occur in both zones, affecting both its corrosion resistance and strength. Many of the phase transformations and processes already described in this section occur in and near the weld. Honeycombe and Bhadeshia (1996, Chapter 13) present a brief summary of the important effects associated with the generation of weld microstructures in steels.

W21.11 Precipitation Hardening of Aluminum Alloys

Pure FCC Al metal has the following properties: a low density, $\rho \approx 2700$ kg/m^3, and a low melting point, $T_m = 660$°C; high electrical and thermal conductivities; high ductility in the annealed state; high corrosion resistance due to the thin coating of the protective oxide Al_2O_3. Because of the relatively low strength of pure Al, its alloys with elements such as Cu, Si, and Mg have found a wider range of applications. The microstructures of these alloys are characterized by a solid-solution phase, α-Al, and by intermetallic compounds such as $CuAl_2$ and Al_3Mg_2.

Al alloys are typically strengthened by the mechanism of *precipitation* or *age hardening*. The precipitation-hardening process involves the use of heat treatments, which result in precipitation within the original matrix of a uniform dispersion of very small particles of a second phase. Although a heat-treatment process, precipitation hardening involves a distinctly different sequence of steps than occur in the heat treatment of steels, which results in the formation of martensite, for example. Two heat treatments are typically required, the first for creating a solid solution and the second for accelerating the process of precipitation or aging. The first heat treatment takes place at a temperature near T_e and for a time long enough to produce a solid solution. The alloy is then quenched to room temperature to obtain a supersaturated solid solution. The second heat treatment is then carried out at a lower T to allow the diffusion to occur which is necessary for formation of the precipitates of the second phase, which results in the strengthening of the alloy. Precipitation hardening is more commonly carried out in Al–Cu, Al–Si, Cu–Be, Cu–Sn, and Mg–Al alloys and in Ni_3Ti and Ni_3Al compounds than in ferrous alloys. Precipitation hardening in Ni_3Al is discussed in Section 12.8.

To illustrate a specific example of the precipitation-hardening process in Al alloys, consider the Al-rich side of the Al–Cu equilibrium phase diagram (Fig. W21.24). The

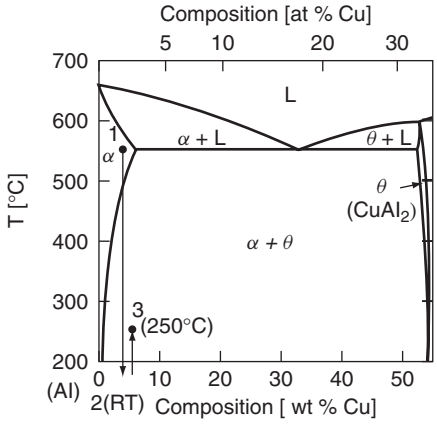

Figure W21.24. Al-rich side of the Al–Cu equilibrium phase diagram shown to illustrate the precipitation-hardening process. The two stable solid phases present are the solid-solution α-Al phase and the θ phase (i.e., the intermetallic compound $CuAl_2$). The sequence of treatments used for precipitation hardening of an Al–1.5Cu wt % alloy is also shown: 1, solid-solution heat treatment at $T \approx 550^{\circ}C$; 2, quench to room temperature; 3, precipitation heat treatment at $T \approx 250^{\circ}C$. (From ASM Handbook, 9th ed., Vol. 3, *Alloy Phase Diagrams*, ASM International, Materials Park, Ohio, 1992, p. 244.)

two stable solid phases present are α-Al, which is a solid solution of Cu in Al, and the θ phase corresponding to the intermetallic compound $CuAl_2$. The solubility of Cu in α-Al reaches a maximum value of $x_e = 5.6$ wt % at $T_e = 548^{\circ}C$ and then decreases rapidly with decreasing T, reaching ≈ 0.02 wt % at room temperature. The initial heat treatment for obtaining a solid solution takes place near T_e for $Al_{1-x}Cu_x$ alloys with $x < x_e$. Following quenching to room temperature, the Al–Cu alloy then undergoes a precipitation heat treatment. If the alloy is left either at room temperature for a few days or is reheated to $T \approx 100$ to $150^{\circ}C$, the Cu atoms are not able to undergo sufficient diffusion to form precipitates of $CuAl_2$. Instead, they rearrange themselves locally within the lattice on {100} planes in two-dimensional platelets or disks known as *Guinier–Preston* (GP) *zones*. The first structures formed, known as *GP-1 zones*, are coherent with the Al lattice and are essentially randomly distributed in the alloy. They are typically 3 to 6 nm long with thicknesses of 0.5 to 1 nm. Their Cu contents are deficient with respect to $x = \frac{1}{3}$, the fraction found in $CuAl_2$.

Additional aging of the alloy leads to the gradual growth of the GP-1 zones and then to the formation of a series of phases or precipitates. The larger *GP-2 zones*, also known as the θ'' *phase*, with lengths ≈ 10 nm, widths ≈ 1 to 4 nm, and Cu contents $x \approx \frac{1}{3}$ are formed next, followed by their conversion into an intermediate θ' *phase*, which is metastable and incoherent with the Al lattice. The stable θ equilibrium phase finally forms from the θ' phase when the aging temperature is raised to $T \approx 200$ to $250^{\circ}C$. The θ' and θ phases both have the $CuAl_2$ stoichiometry but have different crystal structures. The hardness and strength of precipitation-hardened Al–Cu alloys reach maximum values when the GP-2 zones (i.e., the θ'' phase) are formed and then decreases with further heat treatment as the θ' and then the θ phases appear.

The sequence of microstructures of the supersaturated α-Al solid solution and of the θ'' and θ phases are illustrated schematically in Fig. W21.25. Precipitation-hardened

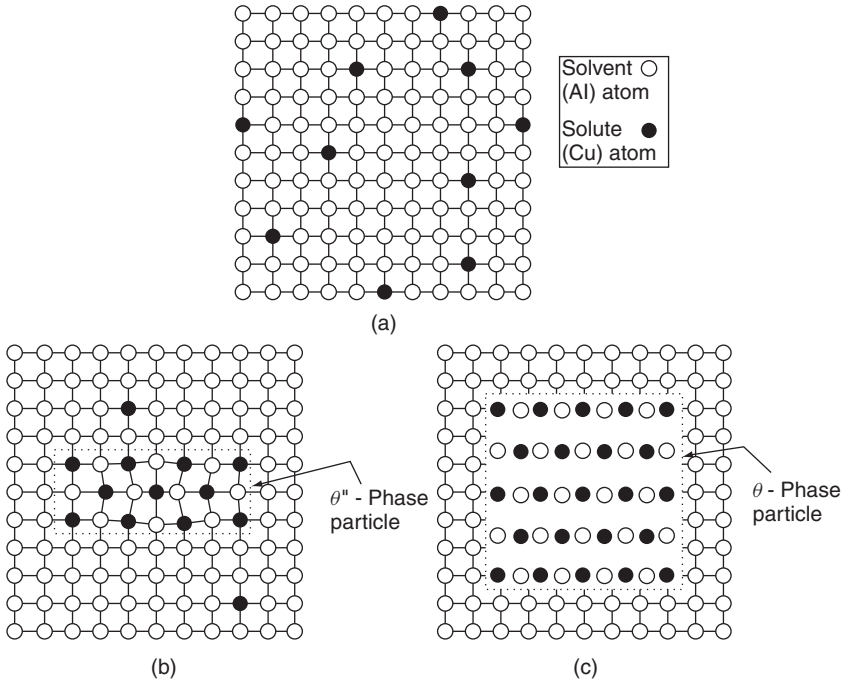

Figure W21.25. Microstructures of (*a*) the supersaturated α-Al solid solution and of (*b*) the θ'' and (*c*) the θ phases. The θ phase has the $CuAl_2$ stoichiometry. The actual particle or zone sizes are much larger than shown here. (From W. D. Callister, Jr., *Materials Science and Engineering*, 2nd ed., copyright 1991 by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

Al alloys can in general have complicated microstructures corresponding to mixtures of the phases mentioned earlier. The strengthening of the alloy can be described by the Orowan expression, Eq. (W21.25), with Λ the average distance between precipitate particles. Strengthening is enhanced when significant lattice strain exists at the interface between the precipitates and the surrounding matrix. This lattice strain is particularly effective in impeding the motion of dislocations. When aging proceeds to the extent that the $CuAl_2$ precipitates become too large and too few in number, they are much less effective in impeding the motion of dislocations. When this happens, the strength of the alloy can actually decrease, a phenomenon known as *overaging*.

W21.12 Synthesis of Metals via Rapid Solidification

As the name indicates, *rapid-solidification processing* (RSP) of metals involves a rapid transition from the liquid to the solid state. RSP usually involves the cooling of liquid metals at sufficiently high rates, $\approx 10^3$ to 10^9 K/s, so that nonequilibrium compositions, phases, or microstructures that are not ordinarily obtainable at "normal" cooling rates of $\approx 10^{-2}$ to 10^2 K/s (≈ 10 to 10^5 K/h) can be synthesized. The amorphous or nanocrystalline microstructures often resulting from the RSP of metals have led to the use of the term *metallic glass*. It is ordinarily extremely difficult to produce elemental metals in an amorphous state due to the ease with which liquid metals crystallize due

to their low viscosities and high diffusivities and the ease with which solid metals recrystallize. By contrast, materials based on $Si-O_4$ tetrahedra, such as silicates, form glasses relatively easily on cooling due to the high viscosity of the liquid.

Metals that have been synthesized via RSP include hard and soft magnetic materials; high-strength Al, Mg, and Ti alloys; tool steels; shape-memory alloys; Ni-based superalloys and brazing materials. Some of the properties of metallic glasses are discussed in Chapter W12. The random close-packing model for the short-range order found in metallic glasses is discussed in Chapter 4.

Techniques that are used in RSP to obtain extremely high cooling rates include the following:

1. *Splat cooling*. A small, molten drop of metal is incident at high speed onto a metallic substrate (e.g., copper) held at room temperature or below. A related method involves the trapping of the molten drop between two cooled surfaces (e.g., a hammer and an anvil).
2. *Melt spinning*. A molten stream of metal is projected against a rapidly rotating surface.
3. *Twin-roller quenching*. A molten stream of metal is forced between a pair of rapidly rotating rollers.
4. *Plasma or flame spraying*. The metal in the form of a powder is introduced into a high-temperature plasma or flame and then sprayed onto a cooled substrate.
5. *Surface melting*. A source of thermal energy such as a laser, ion beam, or electron beam causes a thin surface layer of a metal to melt. The surface layer then undergoes rapid resolidification as soon as the source of heat is removed.

In the first three techniques listed above, and in similar techniques not mentioned here specifically, the rapid solidification is achieved by placing as thin a layer of molten metal as possible in contact with a cooled surface of high thermal conductivity to obtain as high a rate of heat extraction as possible from the molten metal. As a result, the materials are typically thin foils or thin, continuous ribbons. The small dimension of the rapidly solidified material is typically ≈ 25 to 50 μm.

Another technique for achieving the rapid solidification of a metal is through the use of strong undercooling of several hundreds of degrees celsius, as when small, molten metallic particles are cooled well below their normal melting point by avoiding nucleation of the solid phase. This RSP technique, known as *atomization*, involves breakup of a stream of molten metal into fine particles. In this case once a solid nucleus forms in a given particle, solidification occurs extremely rapidly due to the high velocity of the solid–liquid interface, which passes through the particle. The resulting solid powder usually needs additional processing (e.g., consolidation) before it can be used to form a solid object. Additional processing of RSP materials is often needed to develop microstructures with the desired mechanical properties. Strong undercooling can, of course, also occur during the rapid cooling processes listed above.

A necessary condition for obtaining nonequilibrium compositions via RSP is that the growth rate or solidification velocity v_{sl} be greater than the diffusive speed $v_d = D/d_a$ of the solute in the liquid metal. Here D is the thermal diffusivity, $\approx 10^{-9}$ m^2/s, of the solute and d_a is the interatomic distance, $\approx 3 \times 10^{-10}$ m. Other important materials parameters that influence the degree of solute incorporation in the solid phase include the solid–liquid interface energy density σ_{sl} and the latent heat ΔH_m and entropy

change ΔS_m for the liquid–solid transition. When $v_{sl} > v_d \approx 0.03$ m/s, it follows that solute can be trapped at above-equilibrium levels in the solidifying solvent. In the limit $v_{sl} \gg v_d$, the solute distribution coefficient K will approach 1. This has been observed in doped Si and in metallic alloys when $v_{sl} > 5$ m/s. For comparison, a typical value for the normal cooling of a steel ingot is $v_{sl} \approx 3 \times 10^{-5}$ m/s.

It is useful to discuss RSP in terms of the equilibrium phase diagram of the system in question even though the process of rapid solidification leads to nonequilibrium solid products. Consider the solid solution and eutectic binary phase diagrams shown schematically in Fig. W21.26. Indicated in each diagram is the curve of T_0 versus composition, where T_0 is the temperature at which the liquid and solid phases of the same composition have the same Gibbs free energy. For the eutectic system shown in the middle, where the two solid phases have the same crystal structure, there is a smooth T_0 curve. In the right-hand phase diagram where the two solid phases have very limited mutual solid solubilities, the T_0 curves do not intersect. In all three cases shown in Fig. W21.26, the solid formed will have the same composition as the liquid when cooling is rapid enough so that solidification occurs at $T < T_0$. Under these conditions the solidification rate can exceed the diffusion rate in the liquid so that the components cannot redistribute themselves in the liquid phase. The *glass-transition temperature* T_g is shown in the right-hand phase diagram. In a glass-forming system where T_g is so low that it cannot be readily reached via rapid solidification, a dispersion of particles of a second phase can then occur in the primary matrix.

Metastable phases can also be formed when cooling rates are sufficiently high. In addition to the important example of the Fe–C system, where Fe_3C is a metastable product, a wide variety of interesting icosohedral metastable phases of Al with fivefold rotational symmetry (e.g., Al_6Mn_{1-x}, $Al_6Mn_{1-x}Fe_x$, $Al_{12}Fe_{1-x}Mo_x$, and $Al_{62}Cu_{26}Fe_{12}$) have been prepared via RSP. An RSP phase diagram using information obtained by heating the surfaces of Al-rich Al–Mn alloys with a scanned electron beam is presented in Fig. W21.27. Here the solid phases obtained for a range of scan (i.e., solidification) velocities v_{sl} from 0.001 to 1 m/s and for Mn concentrations from 0 to 30 wt % are shown. Icosohedral (fivefold symmetry) and decagonal (tenfold symmetry) phases in the form of dendrites in an Al-rich matrix are obtained for v_{sl} greater than about 0.02 m/s and for more than $\approx$ 18 wt % Mn. The solid-solution phase α-Al extends up to $\approx$ 14 wt % Mn for v_{sl} greater than about 0.03 m/s, well beyond the equilibrium eutectic

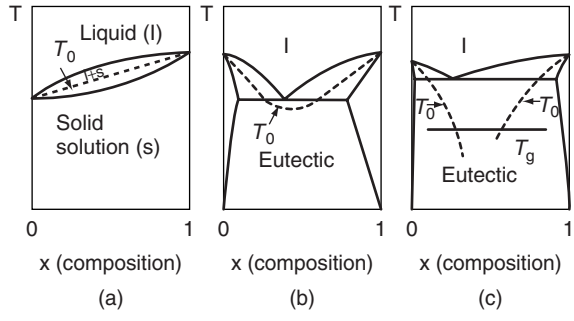

Figure W21.26. Solid-solution and eutectic binary phase diagrams are shown schematically, with the temperature T_0 at which the liquid and solid phases of the same composition have the same Gibbs free energy indicated.

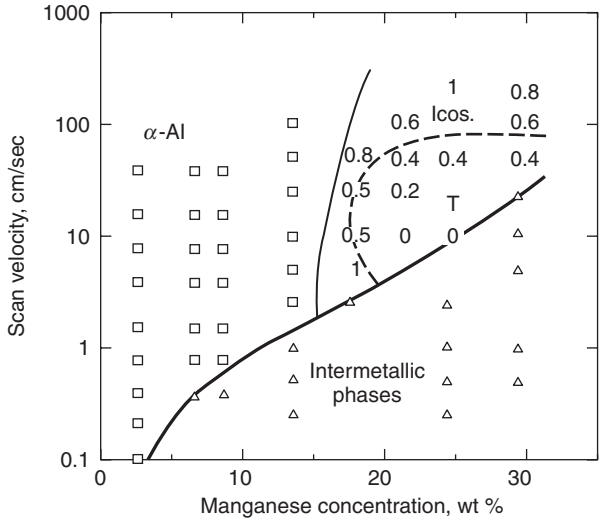

Figure W21.27. RSP phase diagram. The numbers indicate the relative fractions of the intermetallics that are icosohedral; from x-ray diffraction intensities. The region labeled T is a decagonal region. [From R. J. Schaefer et al., *Metall. Trans.*, **17A**, 2117 (1986).]

limit of 1.8 wt % at $T_e = 658°C$. The possibility of obtaining metastable phases in Al–Mn alloys is enhanced due to the many different intermetallic compounds found in Al-rich alloys and also due to their relatively low growth velocities.

Despite the initial and continuing enthusiasm for the RSP technique, many of the hoped-for applications have not yet materialized, due in part, perhaps, to a lack of fundamental knowledge concerning the processes occurring during rapid solidification. It is, of course, an extremely difficult problem to control the microstructure, morphology, and stoichiometry of a rapidly solidified material under processing conditions that are so far from equilibrium. The consolidation of RSP-generated materials into useful forms without causing a degradation of their desirable as-synthesized properties has also proven to be difficult.

W21.13 Surface Treatments for Metals

Most pure metals are thermodynamically unstable with respect to oxidation and other environmental chemical reactions. As a result, a wide variety of physical and chemical processing procedures is used to modify the surface properties of metals in order to improve their corrosion resistance, wear resistance, and surface hardness. Some of these procedures have been mentioned in Chapter W12 and include electroplating, chemical reactions, vapor deposition, ion implantation, and thermal reactions. In addition, the electrolytic anodization of Al resulting in the formation of an oxide layer has been discussed in Section 19.11. Two additional surface-treatment procedures are discussed briefly here: surface carburizing and nitriding and the intense-pulsed-ion-beam (IPIB) surface treatment.

The *surface carburizing* and *nitriding* of metals are both processes that involve changing the chemical composition of the metal in a surface layer. They can be achieved using a variety of techniques for introducing C and N into the material.

Gas carburizing (in the austenite region near $T = 1000^{\circ}C$) and nitriding (in the ferrite region near $T = 500^{\circ}C$) of low-carbon steels typically involve heating the steel in C- or N-containing atmospheres (CH_4 or NH_3), which leads to the rapid diffusion of C or N atoms into the near-surface region known as the *case* (hence the use of the term *casehardening*). Other metallic substitutional alloying elements, such as Ni, Mn, and Cr, are not affected by this treatment, due to their much lower diffusivities in iron. Analogous processes known as *carbonitriding* (or *nitrocarburizing*) and *boronizing* can also be used for surface hardening.

The resulting spatial distribution of C in the steel depends on both the temperature and time of the carburizing process. The carbon concentration is given approximately by the solution of Fick's second law of diffusion [see Eq. (W6.2)]. Typical C concentrations obtained in the surface layer are ≈ 0.8 to 1 wt % (i.e., well below the solubility limit of C in austenite). As the steel is cooled from the carburizing temperature, the microstructure that develops varies with depth into the material due to the varying C concentration. Pearlite and cementite are formed at and just below the surface, then only pearlite when the C concentration has fallen to the eutectoid composition, followed by a mixture of pearlite and ferrite at greater depths. For most steels carburized for 5 to 10 h, the thickness of the carburized surface layer is from 0.5 to 2 mm.

Following the carburizing step, additional heat treatments known as casehardening are necessary to form precipitates of martensite, which result in the formation of a wear-resistant surface layer on the steel. This subsequent heat treatment usually takes place in the austenite phase near $T = 850^{\circ}C$ and is followed by rapid quenching to form martensite. A martensite tempering heat treatment is then carried out in the range $T = 150$ to $200^{\circ}C$ to relieve stresses.

Surface nitriding procedures are ordinarily employed for steels containing the alloying elements Al, V, Cr, and Mo and result in surface layers which are harder than those which are obtained by carburizing. Nitriding is usually carried out in an NH_3 atmosphere and at lower temperatures, and therefore for longer times, than for the case of carburizing since the eutectoid temperature T_e in the Fe–N system is only $\approx 590^{\circ}C$. The possible microstructures appearing in the Fe–N system are more complicated than in the Fe–C system since more than one stable iron nitride (e.g., Fe_4N, Fe_3N, and Fe_2N) can exist in the nitrided surface layer, depending on the processing conditions. The relatively N-rich compound Fe_3N is typically found near the surface, while Fe_4N is found at a greater depth where the diffused N concentration is lower. In addition, precipitates of the nitrides of the alloying elements Al, V, Cr, and Mo are also found in the nitrided surface layer. As a result, the surface layer can be quite hard due to the dispersion-strengthening mechanism. In contrast to carburizing, no additional heat treatment is required to harden the nitrided surface layer.

In the case of surface hardening via carbonitriding or boronizing, carbonitrides and borides are formed instead of carbides or nitrides. The Fe_2B phase is preferred over the FeB phase because it is less brittle and also because the resulting casehardened surface is under compressive stress. Boronized layers on plain carbon steels are typically two or three times harder than carburized layers on the same steels.

The carburizing and nitriding of steels can also be carried out in $CH_4/Ar/H_2$ or CO_2 and NH_3 or N_2/H_2 plasmas, respectively, with the result that the necessary treatment times and temperatures can be greatly reduced. In addition, the plasma can clean the surface via sputtering, activate the chemical species so that they interact more readily with the surface to be hardened, and even heat the surface. Plasma nitriding is also used

to improve the surface hardness and wear resistance of Ti alloys containing Al and V. Four distinct layers can be found in the surface region following plasma nitriding at $T = 800^{\circ}C$ for 13 to 15 h: a 0.3 to 0.5-μm surface layer of FCC δ-TiN, a 1.7 to 2-μm layer of tetragonal ε-Ti_2N, a thin layer of Ti_2AlN, and then the diffusion zone containing nitrogen-stabilized α-Ti. An alternative source of energy is employed in the laser nitriding of Fe and Ti in a N_2 atmosphere which leads to improved hardness and corrosion resistance.

The *intense-pulsed-ion-beam* (IPIB) surface treatment is a recently developed thermal process that causes rapid heating and melting of the surface layer of a metal, followed by extremely rapid cooling, $\approx 10^9$ K/s, of the layer. This procedure, which can be considered to be a type of rapid-solidification processing, results in nonequilibrium microstructures such as amorphous, metastable, or nanocrystalline layers in the surface region. Such surface layers on tool steels and high-temperature Ti alloys have greatly improved surface hardnesses and wear and corrosion resistances. The plasma-immersion ion-implantation (PIII) procedure used to implant dopant ions into semiconductors is also used to implant N into the surfaces of metals in order to improve wear resistance.

The intense pulsed ion beams are typically composed of H or heavier ions. A single ion pulse containing $\approx 10^{13}$ to 10^{14} ions/cm^2 leads to the implantation of ionic species at the level of only $\approx 10^{-5}$ at % in the implanted surface region, which can be $\approx 10^2$ to 10^3cm^2 in area. The depth of the IPIB treatment can be ≈ 2 to 10 μm for H ions but a factor of 20 less than this for heavier ions. IPIB-induced shock waves due to the use of heavier ions such as N can lead to greatly improved mechanical and chemical properties to a depth of up to 100 μm.

As an example of the IPIB treatment, the surface cross section of a tool steel sample treated with a 40-ns-duration 10-J/cm^2 pulsed beam of 0.5–1 to MeV C and H ions is shown in Fig. W21.28. The treated depth is ≈ 5 μm. In this near-surface layer which originally consisted of ferrite and large cementite particles, the carbon has been dissolved into the Fe matrix during the melting. Following rapid resolidification of this region, 20-nm carbide grains have been observed.

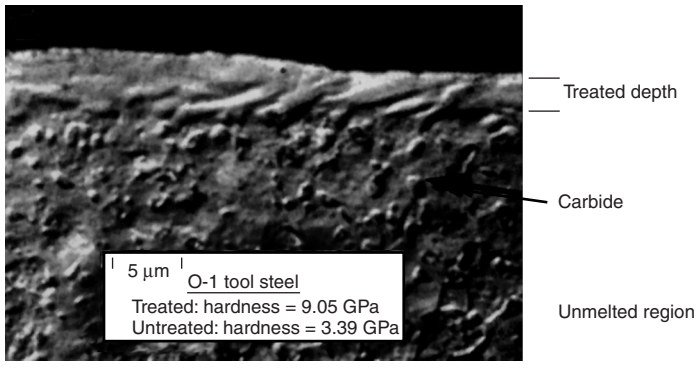

Figure W21.28. As an example of the intense-pulsed-ion beam (IPIB) treatment, the surface cross-section of a O1 tool steel sample treated with a 40-ns-duration 10-J/cm^2 pulsed beam of 0.5- to 1-MeV C and H ions is shown. [From H. A. Davis et al. *Mater. Res. Soc. Bull.*, **21**(8), 58 (1996).]

W21.14 Chemical Vapor Deposition of Diamond

The synthesis of crystalline diamond films via CVD has become an important area of research over the last 15 to 20 years. The growth of diamond takes place either at atmospheric pressure (10^5 Pa), as in the case of the oxygen–acetylene or plasma torches, or at reduced pressures of about 10^3 to 10^4 Pa (7.6 to 76 torr) when microwave plasmas or hot filaments are used. The substrates employed are Si crystals, transition metals such as Mo and W, and ferrous-based materials such as tool steels. Substrate temperatures T_s are normally in the range 800 to 1100°C, although growth of diamond has been observed up to ≈ 1250°C and down to ≈ 500°C. Graphite is deposited at higher T_s while amorphous carbon is deposited at lower T_s. Typical chemical compositions of the CVD environment as expressed by the ratios of the feedstock gas flow rates are $H_2/CH_4 \approx 100{:}1$ or $H_2/CH_4/O_2 \approx 100{:}4{:}0.4$ in the microwave plasma or the hot filament method and $C_2H_2/O_2 \approx 101{:}100$ (i.e., slightly carbon-rich) in the oxygen–acetylene torch.

An understanding of the growth of diamond under conditions where graphite is the thermodynamically stable form of carbon can be obtained by recognizing that the competing forms of solid carbon, graphite, and amorphous carbon have higher solubilities in the vapor phase relative to diamond in reactive environments containing large amounts of either atomic hydrogen or oxygen (or both). The thermodynamic *quasiequilibrium* (QE) *model*[†] has been applied to the carbon–hydrogen (C–H) and C–H–O systems to provide the basis for an analysis of the CVD of diamond. In this approach the dominant vapor species (H, C_xH_y, O) in equilibrium with either the diamond or graphite surfaces and also the deposition and etching rates of diamond or of graphite can be determined. When the kinetic effects associated with the enhanced etching of graphite by atomic hydrogen and oxygen are included in the model, regions in the CVD phase diagram of the C–H and C–H–O systems are predicted where diamond is the only stable form of solid carbon present.

The key assumption of the QE model is that *thermochemical equilibrium* exists between the solid carbon surface and the vapor species desorbed from it. Kinetic theory is employed to determine the rates at which vapor species arrive at and leave the carbon surface. The standard Gibbs free energies of formation $\Delta_f G^0(C_xH_y, T)$ of the vapor species are employed to obtain the needed equilibrium constants $K(C_xH_y, T)$ using the expression

$$K(C_xH_y, T) = \exp\left[-\frac{\Delta_f G^0(C_xH_y, T)}{RT}\right]. \tag{W21.26}$$

These in turn provide the equilibrium vapor pressures of the $C_xH_y(g)$ species for the reactions

$$xC(s) + \frac{y}{2}H_2(g) \leftrightarrow C_xH_y(g), \tag{W21.27}$$

using

$$P_{eq}(C_xH_y, T) = K(C_xH_y, T)[P(H_2)]^{y/2}, \tag{W21.28}$$

[†] J. C. Batty and R. E. Stickney, *J. Chem. Phys.*, **51**, 4475 (1969).

where $P(H_2)$ is the partial pressure of H_2 in the system. The pressures in this equation are expressed in atmospheres.

By requiring conservation of H atoms in the fluxes of atoms and molecules incident on and leaving either the diamond or the graphite surface, predictions for the evaporation rates $R_e(C_xH_y, T)$ can be obtained. Deposition rates are then obtained from

$$R_d(C) = I(C) - R_e(C), \tag{W21.29}$$

where $I(C)$ is the net flux of incident C atoms and $R_e(C)$ is the net flux of C atoms leaving the surface. The evaporation rates $R_e(C_xH_y, T)$ and deposition rates R_d of diamond and graphite are presented as functions of temperature in Fig. W21.29 for a mixture of 1% CH_4 in H_2 at $P = 5 \times 10^3$ Pa. It can be seen that the evaporation rates of C_xH_y species are predicted to be higher above diamond (dashed curves) than above graphite (solid curves), as expected from the slightly higher free energy of formation of diamond relative to graphite. Under the conditions presented in Fig. W21.29, there exists an intermediate temperature range, from $T = 910$ to 2295 K, where diamond is stable relative to hydrogen. For $T < 910$ K diamond is etched via the formation of $CH_4(g)$ while for $T > 2295$ K etching via the formation of $C_2H_2(g)$ dominates.

The data presented in Fig. W21.29 can be used to construct the *CVD phase diagram* for the C–H system shown in Fig. W21.30. Here the regions of stability of solid carbon (i.e., diamond or graphite) are presented at 5×10^3 Pa as functions of temperature and reactant ratio $C/(C + H)$. In this case there exists a region where diamond is predicted to be the only stable phase of solid carbon. This occurs because the phase boundary of graphite has been shifted to the right by taking into account the enhanced etching of graphite by atomic hydrogen. Experimental data points for the deposition of diamond

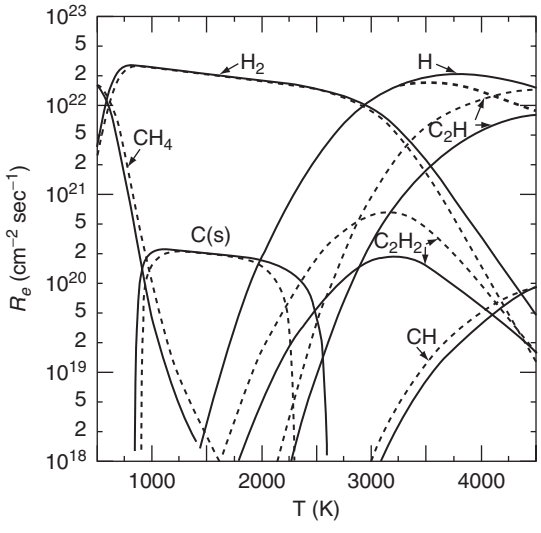

Figure W21.29. Predictions of the quasiequilibrium model for the evaporation rates $R_e(C_xH_y, T)$ of C_xH_y vapor species and the deposition rates $R_d(T)$ of either diamond or graphite are presented as functions of temperature for a mixture of 1% CH_4 in H_2 at $P = 5 \times 10^3$ Pa. [From M. Sommer and F. W. Smith, *High Temp. Sci.*, **27**, 173 (1989). Reprinted by permission of Humana Press, Inc.]

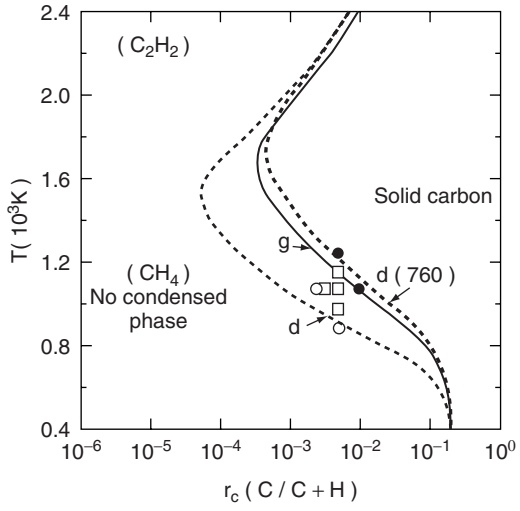

Figure W21.30. CVD phase diagram for the C–H system. The regions of stability of solid carbon (i.e., diamond or graphite) are presented at $P = 5 \times 10^3$ Pa as functions of temperature T and reactant ratio $C/(C + H)$. [Reprinted from M. Sommer, K. Mui, and F. W. Smith, *Solid State Commun.*, **69**, 775 (1989). Copyright 1989, with permission from Elsevier Science.]

are also presented and can be seen to be in very good agreement with the predicted region of stability of diamond. Similar predictions for the C–H–O system have been obtained for the deposition of diamond via the oxyacetylene torch.[†]

Problems remaining in the CVD of diamond films are related to obtaining films with fewer defects and with lower levels of nondiamond components, such as graphitic carbon, amorphous carbon, and impurities such as hydrogen and nitrogen. The successful preparation of *n*-type films is also an important goal for the eventual use of diamond as an active element in electronic devices. The *p*-type doping of diamond by substitutional B acceptors is well established.

In addition to the CVD of diamond films, the synthesis of diamond at high temperatures ($\approx$ 2000 K) and pressures ($\approx$ 60 atm) (i.e., under HPHT conditions) in the form of small single crystals or abrasive grains is a well-developed technology, with several tons of diamond being prepared yearly. Under these conditions, diamond is apparently thermodynamically stable with respect to graphite, although the phase boundary between diamond and graphite is still not very well known in the HPHT region. The HPHT method relies on the solubility of carbon in molten transition metals such as Ni at high T and P and its subsequent controlled precipitation as diamond crystals.

Cubic BN (c-BN) with the zincblende crystal structure is similar in many respects to diamond, having essentially the same lattice constant, a wide bandgap ($\approx$ 6.4 eV) and also very high hardness and thermal conductivity. c-BN is actually superior to diamond for electronic applications due to the fact that it can be doped both *n*- and *p*-type with Si and Be, respectively. The ceramic c-BN also has excellent potential for use as a hard, wear-resistant coating for tools since its solubility in ferrous materials is much

[†] R. B. Wang, M. Sommer, and F. W. Smith, *J. Cryst. Growth*, **119**, 271 (1992).

lower than that of carbon. So far a successful technique for preparing single-phase c-BN in thin film or bulk form has not been developed.

W21.15 Synthesis of $YBa_2Cu_3O_{7-x}$

Early methods of synthesizing the high-temperature superconductor $YBa_2Cu_3O_{7-x}$ (YBCO or 1:2:3) involved a solid-state self-flux reaction resulting in a metastable compound. Typically, a mixture of $BaCO_3$, CuO, and Y_2O_3 with the molar ratios Y/Ba/Cu = 1:4:10 was mixed and ground in a zirconia crucible, pressed into pellets, and heated at 890°C for a day. The process was repeated a second time. Finally, the material was annealed at 1000°C while being subjected to flowing O_2 for three days. The cooling rates had to be slow to obtain crystals of size ≈ 1 mm. The parent compound is $YBa_2Cu_3O_7$, which is nonstoichiometric. This compound is enriched with oxygen as the O atoms intercalate into the crystal and order. The oxygen content of the crystals $(7 - x)$ was found to be a function of the oxygen partial pressure during annealing. Optimal values of T_c (≈ 90 K) were obtained for $x = 0.3$. To obtain crystal growth the temperature had to be sufficiently high to obtain a partial melt, yet sufficiently low so as not to decompose the crystals to more thermodynamically stable forms (such as Y_2BaCuO_5). A ternary phase diagram is given in Fig. W21.31.

The deposition of thin films of YBCO requires a different approach. Methods such as magnetron sputtering, pulsed excimer-laser ablation, and metal-organic chemical vapor deposition (MOCVD) have been developed. A proper choice of substrate has to be made so that epitaxial growth will occur. YBCO is an orthorhombic crystal with lattice constants a, b, and $c = 0.383$, 0.389, and 1.169 nm. Suitable substrates for growing crystals with the c axis normal to the substrate surface are the (100) faces of $SrTiO_3$ ($a = 0.39$ nm) and $LaAlO_3$ ($a = 0.536$ nm $\approx a\sqrt{2}$). These substrates have high melting temperatures, $T_m = 2030$°C and 2110°C, respectively, and also have low microwave loss, which is important in designing superconducting microwave filters and cavities.

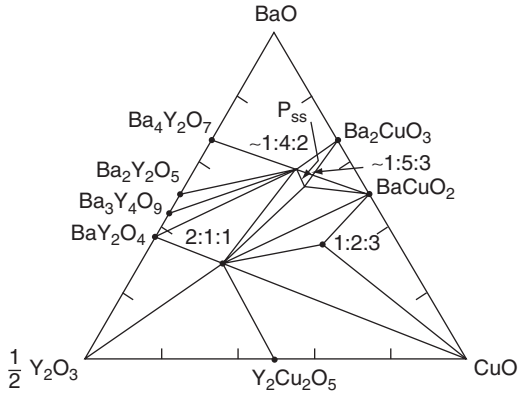

Figure W21.31. Ternary phase diagram for Y–Ba–Cu–O. The numbers x:y:z refer to the Y:Ba:Cu stoichiometry. P_{ss} denotes a solid solution region. The temperature is $T \approx 975$ to 1000°C. (Adapted from L. F. Schneemeyer et al., Barium yttrium copper oxide crystals, in D. W. Murphy and L. V. Interrante, eds., *Inorganic Synthesis*, Vol. 30, Wiley, New York, 1995.)

In the MOCVD method the cations are bound to organic anions, and the resulting metal–organic precursor has a high vapor pressure at relatively low temperatures ($\approx 250^\circ$C). An inert carrier gas such as Ar is used. Precursors such as $Ba(THD)_2$, $Cu(THD)_2$, and $Y(THD)_3$ are used, where THD is bis(2,2,6,6-tetramethyl-1,3,5-heptanedione). Oxygen is introduced. The vapor is transported to the heated substrate, where the organic components are pyrolyzed and the oxides of the cations are deposited. In this method the substrate is the hottest part of the system, so the deposition takes place only on the substrate, not on the walls of the system.

In pulsed laser deposition (PLD) a Kr–F excimer laser generates a $\lambda = 248$ nm pulse of duration 30 ns with a fluence of $\approx 5 \times 10^4$ J/m^2. The pulse is absorbed near the surface of a block of material that is to be deposited on a substrate some distance away. The pulse has sufficient energy to vaporize (and partially ionize) several hundred layers of atoms. A plume of ablated material is cast off primarily in the forward direction perpendicular to the target. The substrate temperature is in the range 500 to 700°C, which provides sufficient atomic mobility for crystal growth to occur. The deposited layer retains the chemical composition of the target. Unlike the MOCVD method, one is not dependent on all the precursors having a high vapor pressure.

W21.16 Synthesis of Si_3N_4

There exist a variety of methods for synthesizing Si_3N_4 powders, but only three methods are used commercially. They are carbothermal reduction and nitridation, direct nitridation of silicon, and a liquid-phase process. Si_3N_4 exists in two phases, a low-temperature metastable α-phase and a high-temperature stable β-phase. Both phases have a hexagonal unit cell, but the stacking sequences of the planes along the c axis are different. The α-phase has the stacking sequence ABABABAB... , whereas the β-phase has the sequence ABCDABCD... . The α-phase can readily accommodate cations within its structure.

In the carbothermal reduction and nitridation process silica reacts with carbon in a nitrogen atmosphere according to the overall formula

$$3SiO_2(s) + 2N_2(g) + 6C(s) \longrightarrow Si_3N_4(s) + 6CO(g). \tag{W21.30}$$

The reaction occurs at temperatures in excess of 1420°C at atmospheric pressure and is endothermic with $\Delta H = 1270$ kJ/mol. To prevent the reverse reaction from occurring, the CO gas must be removed. Unless impurities are added the reaction strongly favors the production of the α-phase. The reaction proceeds in several stages. One possible route is to produce a supersaturated SiO gas and then have this gas react with carbon and nitrogen. Some possible pathways are

$$SiO_2 + C \longrightarrow SiO + CO, \qquad 3SiO + 3C + 2N_2 \longrightarrow Si_3N_4 + 3CO, \tag{W21.31}$$

or

$$SiO_2 + C \longrightarrow SiO + CO, \qquad 3SiO + 3CO + 2N_2 \longrightarrow Si_3N_4 + 3CO_2, \tag{W21.32}$$

followed by

$$CO_2 + C \longrightarrow 2CO. \tag{W21.33}$$

Other reactions are possible as well.

The direct nitridation of silicon involves "burning" solid silicon in a nitrogen atmosphere to promote the exothermic reaction

$$3\mathrm{Si(s)} + 2\mathrm{N_2(g)} \longrightarrow \mathrm{Si_3N_4(s)}. \tag{W21.34}$$

The enthalpy of formation is $\Delta_f H^\circ = -730$ kJ/mol at a temperature of 1320°C. This reaction produces a mixture of both the α-phase and the β-phase. It is possible to reaction-sinter the Si_3N_4 by slowly raising the temperature to $\approx$ 1400°C and keeping it at that temperature for several days. Then the β-phase is produced.

The liquid-phase process involves pouring liquid silicon tetrachloride into liquid ammonia at $P = 230$ Pa and $T = -29$°C. The $SiCl_4$ is dissolved in an organic solvent composed of cyclohexane and benzene. A polymer of silicon diimide is formed at the interface according to the reaction

$$n(\mathrm{SiCl_4}) + 18n(\mathrm{NH_3}) \longrightarrow (\mathrm{Si(NH)_2})_n + 4n(\mathrm{NH_4Cl{\cdot}3NH_3}). \tag{W21.35}$$

The $(Si(NH)_2)_n$ dissolves in the organic solvent. The solvent, ammonia, and NH_4Cl are then removed and the silicon diimide is heated to 1000°C to convert it to amorphous silicon nitride according to the reaction

$$3(\mathrm{Si(NH)_2})_n \longrightarrow n(\mathrm{Si_3N_4}) + 2n(\mathrm{NH_3}). \tag{W21.36}$$

Further heating anneals the amorphous material to crystalline α-Si_3N_4.

Densification of Si_3N_4 can be accomplished, for example, by liquid-phase sintering. The Si_3N_4 is mixed with silica and additives such as alumina and yttria which are used to lower the melting temperature of the silica. This is important because Si_3N_4 has a low dissociation temperature (1500°C) and it is desirable to keep the liquid temperature at around 1470°C. In the sintering process the silica and additives melt, some of the Si_3N_4 goes into solution and precipitates out, ultimately causing the grains to merge and to eliminate the intergranular void spaces. The microstructure that results is strongly influenced by the additives used.

It is also possible to densify Si_3N_4 powders by means of hot pressing, hot isostatic pressing, or gas pressure sintering. Oxides, such as MgO or Y_2O_3, or $BeSiN_2$ are added as sintering aids.

It is possible to deposit Si_3N_4 films by means of CVD. The precursors are ammonia (NH_3) and dichlorosilane ($SiCl_2H_2$). The operating temperature is 700 to 800°C. Unfortunately, this is too high for application to electronic VLSI chips. PECVD is used to reduce the operating temperatures to below 450°C, in which case amorphous films also containing H are deposited.

Laser reactions may also be used to synthesize Si_3N_4. A mixture of NH_3 and SiH_4 is irradiated with infrared radiation from a CO_2 laser. The SiH_4 is vibrationally excited and the net endothermic reaction

$$3\mathrm{SiH_4}(g) + 4\mathrm{NH_3} \longrightarrow \mathrm{Si_3N_4}(s) + 12\mathrm{H_2}(g) \tag{W21.37}$$

is able to proceed. Particles of size $\approx$ 20 to 100 nm are produced.

W21.17 Synthesis of SiC

At low temperatures ($T < 1800^{\circ}$C) one may sinter powders of Si and C to produce the β (zincblende) form of SiC via the reaction $Si + C \rightarrow SiC$. For temperatures higher than 2000°C, hot pressing may be used. Silicon carbide is most commonly synthesized using the Acheson process, which employs a resistance furnace. A mixture consisting of carbon, NaCl, SiO_2 sand, and some sawdust is placed around a core of graphite. An electrical current is passed through the graphite, heating it to a temperature of around 2600 to 2700°C (below the melting temperature of 2830°C). The reaction $SiO_2 + 3C \rightarrow SiC + 2CO$ is highly exothermic, with $\Delta H = 14,700$ kJ/kg, and this helps create the high temperature. The α (wurtzite, high temperature) form of SiC grows around the graphite core. The NaCl helps to remove impurities from the material. The sawdust creates sufficient pore space so that the CO gas may escape.

The Lely process is another way of synthesizing SiC. Amorphous SiC granules are placed inside a hollow graphite tube and the combination is heated to $\approx 2500^{\circ}$C in an inert gas such as Ar. Some of the SiC sublimes, forming a vapor. From this vapor SiC crystals nucleate on the granules and then continue to grow.

Silicon carbide may also be grown by chemical vapor deposition on a hot substrate. The temperatures are typically much cooler than used in the Acheson and Lely processes. The precursor gases that are used are silane, (SiH_4) and methane (CH_4) or propane (C_3H_8). Typical net reactions are $SiH_4 + CH_4 \rightarrow SiC + 4H_2$ or $3SiH_4 + C_3H_8 \rightarrow 3SiC + 10H_2$. Laser-induced reactions are also possible, such as $2SiH_4 + C_2H_4 \rightarrow 2SiC + 6H_2$. It is also possible to use single molecules called *carbosilanes*, containing Si and C in a 1:1 ratio, as the precursor. Included are molecules such as 1,3-disilacyclobutane. It is possible to produce β-SiC at temperatures $\approx 1000^{\circ}$C and even lower. Other molecules in use include 1,3-disila-*n*-butane and methylsilane (CH_3SiH_3).

Silicon carbide powders may be formed into shapes using methods such as extrusion, injection molding, and hot isostatic pressing, among others. SiC may be sintered using the hot-pressing technique at temperatures in excess of 2000°C.

Of the various methods for preparing SiC, CVD produces the highest-quality crystals. For example, a thermal conductivity of $\kappa = 300$ W/m·K at $T = 300$ K is attainable, compared with values in the range 15 to 120 for sintered SiC, 120 to 170 for reaction-bonded SiC, and 50 to 120 for hot-pressed SiC. CVD SiC also yields the material with the highest elastic modulus, $E = 466$ GPa, and the lowest coefficient of thermal expansion, 2.0×10^{-6} K^{-1}, at room temperature.

W21.18 Synthesis of the Zeolite ZSM-5

Sol–gel synthesis is also used to produce the zeolite ZSM-5, introduced in Section 13.6. This zeolite is an aluminosilicate in which the silicon-to-aluminum ratio is very high. A typical procedure is to first prepare $NaAlO_2$ by $Al_2O_3 + 2NaOH \rightarrow 2NaAlO_2 + H_2O$ and then put it in a solution of NaOH and H_2O. A second solution is prepared by dissolving a small amount of tetrapropylammonium bromide in H_2SO_4 and water. The solutions are combined with a sol consisting of silica, Na_2O, and water. The silica-to-alumina ratio can be kept high to make the resulting crystal almost entirely silica. The resulting solution is kept at 95°C for up to two weeks and the sol–gel reaction is monitored closely to see when crystallites of the zeolite form. When the crystallization is complete, the organic molecules can be slowly pyrolyzed in oxygen at elevated

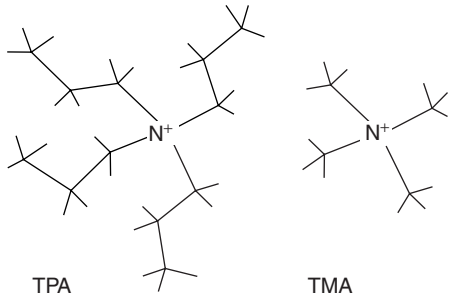

Figure W21.32. Tetrapropylammonium (TPA) ion and tetramethylammonium (TMA) ion. [Adapted from D. W. Lewis et al., *Nature*, **382**, 604(1996).]

temperatures. The resulting crystal consists mainly of tetrahedrally coordinated silica with aluminum ions incorporated into the framework. Sodium ions (equal in number to the Al ions for charge balance) reside outside the framework in the pore spaces. The zeolite serves as an ion exchanger, so other ions may be substituted for the sodium.

The tetrahedral molecule tetrapropylammonium (TPA) ion (Fig. W21.32) serves as a template molecule upon which the zeolite nucleates. The framework and pore size of the crystal are determined by the geometry of this ion. The steric hindrance presented by the ion guarantees a large pore size. Since the charge on the TPA ion is $+1e$, it serves to compensate for the valence deficit that occurs when an Al^{3+} ion replaces a Si^{4+} ion. Each of the TPA propyl groups extends into one of the four channels that emanate from each intersection, with the nitrogen atom residing at the junction. If other ions are used, the pore size will be different. This gives the chemist the opportunity to custom design zeolite structures based on the template molecule employed.

Recent observation of the formation and growth of a similar material, zeolite A ($Na_{12}[(AlO_2)_{12}(SiO_2)_{12}]\cdot 27H_2O$), identified the steps involved in the formation of crystals.[†] The monomers polymerized to form small amorphous clusters of aluminosilicate particles with diameters in the range 5 to 10 nm in solution. Tetramethylammonium (TMA) (see Fig. W21.32) is used as a template for zeolite A. When TMA is added to the solution, the solution becomes basic and the particles aggregate to form amorphous gel particles, with sizes in the range 40 to 80 nm. The aggregation is presumably due to the screening of the Coulomb repulsion between the particles by the ions in solution, allowing the long-range van der Waals forces to bring the particles together. After three days at room temperature, single crystals nucleate within the gel particles and grow to the size 10 to 30 nm. After a week there is complete conversion of the gel particles to the single crystals, of size 40 to 80 nm. Presumably the high supersaturation present in the amorphous gel particles is the driving force for the nucleation and growth of the crystals. If the temperature is then elevated to 80°C, there is transport through the solution and the crystals undergo Ostwald ripening. Larger crystals in the range 200 to 400 nm are formed within one day.

By using micelles as the templating agent it is possible to produce mesoporous films of transition metal oxides with variable pore sizes.[‡] The micelles are rodlike structures

[†] S. Mintova et al., *Science*, **283**, 958 (1999).
[‡] T. Sun and J. Y. Ying, *Nature*, **389**, 704 (1997).

that are self-assembled from hexylamine molecules, with the molecular axes directed perpendicular to the rod axis. The length of the hydrocarbon chain determines the radius of the rod.

It has been reported[†] that zeolite-like materials with helical pores could be constructed by using inorganic cations or amines as templates. In place of the aluminosilicate structure of ZSM-5, the inorganic framework is based on zinc and beryllium arsenate and gallium germanate. Such structures could serve as chiral catalysts that would yield products with enantioselectivity (i.e., with a fixed handedness).

W21.19 Synthesis of the Perovskite PLZT

Lead zirconate titanate, $Pb(Zr,Ti)O_3$ (PZT), is a ceramic material used in the electronics and optics industries. There are a number of ways of synthesizing it. Powders can be made by the solid-state reaction method or by wet chemical synthesis.

In the solid-state reaction method the reagent powders PbO, ZrO_2, and TiO_2 are mixed, ground, and then heated at 850°C for about 3 hours, during which time crystallization takes place. To create the daughter compound $Pb_{1-x}\ La_x\ (Zr_y\ Ti_{1-y})_{1-x/4}O_3$ (PLZT) one uses a combination of La_2O_3 and ZrO_2 instead of pure zirconia.

Wet chemical methods include coprecipitation, hydrothermal synthesis, and sol–gel synthesis. An example of the coprecipitation method is to mix various salts together with ammonium hydroxide and water, for example,

$$\begin{aligned}(1-x)PbCl_2 + xLaCl_3 + y\left(1-\frac{x}{4}\right)ZrCl_4 + (1-y)\left(1-\frac{x}{4}\right)TiCl_4 + 6NH_4OH \\ \longrightarrow (1-x)Pb(OH)_2 + xLa(OH)_3 + y\left(1-\frac{x}{4}\right)ZrO(OH)_2 \\ + (1-y)\left(1-\frac{x}{4}\right)TiO(OH)_2 + \left(1-\frac{x}{4}\right)H_2O + 6NH_4Cl. \qquad \text{(W21.38)}\end{aligned}$$

The various hydroxides form a gel precipitate. The solution is washed to eliminate the ammonium chloride salt. It is then heated at a temperature of 550°C for an hour during which time the hydroxide groups are converted to water and the PLZT crystals are formed through the reaction

$$\begin{aligned}(1-x)Pb(OH)_2 + xLa(OH)_3 + y\left(1-\frac{x}{4}\right)ZrO(OH)_2 + (1-y)\left(1-\frac{x}{4}\right)TiO(OH)_2 \\ \longrightarrow Pb_{1-x}La_x(Zr_yTi_{1-y})_{1-x/4}O_3 + \left(2+\frac{x}{4}\right)H_2O. \qquad \text{(W21.39)}\end{aligned}$$

Hydrothermal synthesis allows the reaction to occur at lower temperatures (350°C), but at higher pressures.

The sol–gel synthesis of PLZT utilizes precursors typically consisting of metal salts (lead acetate hydrate and lanthanum acetate hydrate) and alkoxides (zirconium *n*-propoxide and titanium isopropoxide). Salts are used because the alkoxides of lead and

[†] T. E. Gier et al., *Nature*, **395**, 154 (1998).

lanthanum are not soluble. The acetates are added to an alcohol such as methoxyethanol in water. Reactions such as

$$Pb(OAc)_2 + ROH \longrightarrow Pb(OAc)\,(OR) + HOAc \tag{W21.40}$$

or

$$4Pb(OAc)_2 + H_2O \longrightarrow 3Pb(OAc)_2{\cdot}PbO + 2HOAc \tag{W21.41}$$

occur, with corresponding ones for $La(OAc)_3$. Here $R = C_2H_4OCH_3$ and it is seen that the reaction replaces the OAc ion by an OR ion. Typical condensation reactions that can occur are

$$A(OR)_4 + B(OAc)_n \longrightarrow (OR)_3A{-}O{-}B(OAc)_{n-1} + ROAc, \tag{W21.42}$$

where $n = 2$ or 3, A = Ti or Zr, B = Pb or La, and $R = C_3H_7$ or $C_2H_4OCH_3$.

Thin films of PLZT created by the sol–gel process may be spun onto silica or MgO substrates while still wet and then dried. The films may be processed further for various applications.

W21.20 Synthesis of Glasses: Pilkington Process

The synthesis of glass involves essentially three steps. In the first step a batch of raw materials is prepared. The principal ingredient is SiO_2. Modifier oxides, such as Na_2O or K_2O, are added to lower the melting temperature. Other oxides, such as CaO, are added to provide chemical stability. If a glass-ceramic with controlled crystallinity is to be produced, Al_2O_3 is also added.

In the second stage the mixture is melted. For common glasses the temperature is elevated to 1300 to 1400°C, while for glass-ceramics the temperature range is 1400 to 1500°C. Volatile gases leave the liquid. The viscosity of the liquid decreases rapidly with increasing temperature, so the rate of escape of the gas bubbles is sensitive to temperature.

The final stage involves forming the glass into the desired shape. Techniques such as rolling, blowing, casting, pressing, and drawing are used. In creating ordinary glass the cooling rate is as fast as it can be without producing cracking. If it is too high, the temperature differential between the surface and interior portions of the glass produces stress fields that could lead to cracking. In creating glass-ceramics, slower cooling is required. The cooling rate is critical in determining the amount of crystallization that will occur. Residual stresses may be eliminated or reduced by annealing the glass.

In some cases the surface of the glass is tempered to enhance its mechanical properties. For example, one may heat the glass uniformly in a furnace, remove it, and then rapidly cool the outer surface. Due to the poor thermal conductivity of glass, the interior remains hot for some time. Viscoelastic relaxation allows the atoms in the interior to assume new configurations to relieve the stress. Upon further cooling the glass becomes so viscous that relaxation no longer can occur and the interior develops a tensile stress in response to the thermal contraction. Correspondingly, the surface region is put in a state of compressive stress. The existence of the internal stress field permits the glass to withstand larger flexural stresses that may be imposed on it.

In addition to thermal tempering, chemical tempering is also possible. For example, by exchanging the Na^+ ions for smaller Li^+ ions near the surface, the surface is placed

under compressive stress. It is also possible to remove Na by exposing the glass to gases such as SO_2 and H_2O. Effectively, the Na^+ ions are replaced by protons from the water, producing OH radicals.

The *Pilkington process*, or *float process*, represents more of a manufacturing process than a change in the microstructure of a material. It is included here because it shows the importance of surface tension in a practical manufacturing setting. The process provides an economical means for large-scale production of finished sheets of glass. Molten glass is poured onto a tub of molten metal and is allowed to float until it cools below the glass temperature, T_g. If T_m is the melting temperature of the metal, then if $T_m < T < T_g$, the solidified glass that forms will float on the molten metal and may readily be removed. Tin is usually used as the metal because it melts at a sufficiently low temperature ($T_m = 232°C$).

Let ρ_m and ρ_g be the densities of the metal and glass. For flotation it is required that $\rho_m > \rho_g$. For tin and glass the specific gravities are 6.5 and 2.2, respectively. The interfacial surface tensions are denoted by γ_{mv}, γ_{gv}, and γ_{mg}, where the subscript v refers to the surrounding atmosphere (without oxygen). The thickness of the glass sheet will be denoted by t and its base area by A. The base area of the vat of metal is A'. The geometry is depicted in Fig. W21.33.

To find t one minimizes the total potential energy, consisting of gravitational and surface contributions,

$$U = \frac{1}{2}\rho_m g[A(y-h)^2 + (A'-A)y^2] + \rho_g g A t\left(y - h + \frac{t}{2}\right) + \gamma_{gv}A + \gamma_{mg}A + \gamma_{mv}(A'-A), \tag{W21.43}$$

subject to the constraints of constant glass and metal volumes

$$V_m = A(y-h) + (A'-A)y, \tag{W21.44}$$

$$V_g = At. \tag{W21.45}$$

The surface energy associated with the vertical sides of the slab is small and is neglected. Introducing Lagrange multipliers λ and μ, one has

$$\delta(U - \mu V_m - \lambda V_g) = 0. \tag{W21.46}$$

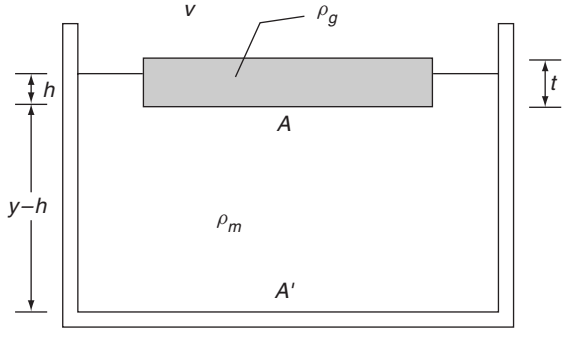

Figure W21.33. Slab of molten glass floating on a bath of molten metal in the Pilkington process.

The partial derivatives are taken independently with respect to the variables y, h, t, and A to obtain the four equations

$$\rho_m g(A'y - Ah) + \rho_g Agt - \mu A' = 0, \tag{W21.47}$$

$$\rho_m g(y - h) + \rho_g gt = \mu, \tag{W21.48}$$

$$\rho_G g(y - h + t) = \lambda, \tag{W21.49}$$

$$\frac{1}{2}\rho_m g(h^2 - 2hy) + \rho_g gt\left(y - h + \frac{t}{2}\right) + \gamma_{gv} + \gamma_{gm} - \gamma_{mv} + \mu h - \lambda t = 0. \tag{W21.50}$$

Eliminating the Lagrange multipliers results in

$$\rho_m h = \rho_g t, \tag{W21.51}$$

which could have been deduced from Archimedes principle, and a formula for t,

$$t = \sqrt{\frac{2\rho_m(\gamma_{gv} + \gamma_{gm} - \gamma_{mv})}{g\rho_g(\rho_m - \rho_g)}}, \tag{W21.52}$$

independent of the volume of the glass. Note that it is necessary for $\gamma_{gv} + \gamma_{gm} > \gamma_{mv}$; otherwise, the glass would spread, with $A \to A'$. Since the interfacial surface tensions are dependent on T, one has some control over the thickness of the sheet by varying the temperature and the cooling rates.

By applying a tensile stress to the sheet of glass while it is cooling, it is possible to stretch it and thereby make it thinner, according to the relation $t = V_g/A$. Let a constant external tension per unit thickness $\tau = S/t$ be introduced, where S is the tension. This is equivalent to adding the term $-\tau A$ to the potential energy, so replacing the term γ_{gv} by $\gamma_{gv} - \tau$ gives the result

$$t = \sqrt{\frac{2\rho_m(\gamma_{gv} + \gamma_{gm} - \gamma_{mv} - \tau)}{g\rho_g(\rho_m - \rho_g)}}. \tag{W21.52$'$}$$

As the parameter τ increases, the thickness t decreases.

W21.21 Synthesis of Polycarbonate

Polycarbonate is synthesized by means of a polymerization reaction that occurs at the interface between two immiscible liquids. One liquid is an organic solvent (such as methylene chloride, CH_2Cl_2) and the other is a basic solution (such as NaOH in water) which acts as the initiator for the reaction. The starting material for the monomer from which the polymer is built is bisphenol-A, $C(C_6H_4)_2(CH_3)_2(OH)_2$, and is soluble in the organic solvent. The organic solvent is dispersed into small globules in the alkaline solution. Phosgene gas, CCl_2O, is bubbled through the emulsion. The primary reaction is

$$C(C_6H_4)_2(CH_3)_2(OH)_2 + CCl_2O + NaOH \longrightarrow C(C_6H_4)_2 \times (CH_3)_2OHCClO_2 + H_2O + NaCl. \tag{W21.53}$$

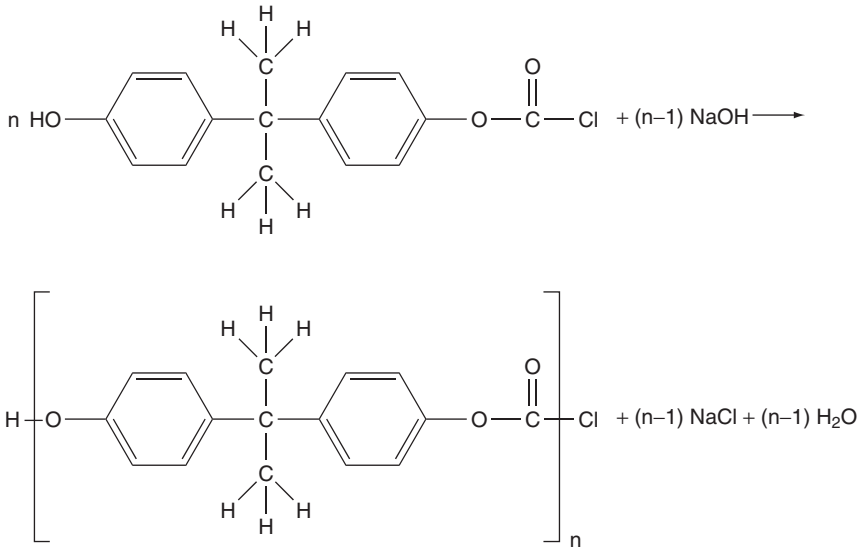

Figure W21.34. Synthesis of polycarbonate.

Successive reactions form the polymer, as indicated in Fig. W21.34. The polycarbonate that is formed is soluble in the organic solvent. Catalysts used to accelerate the reaction include triethylbenzyl ammonium chloride and various tertiary amines. The size (n) of the monomer is regulated by adding a monofunctional phenol such as *p-tert*-butyl phenol, which serves to terminate the chain. Since there is only one functional group on this terminator, it is utilized to attach to the polymer and to stop the polymer growth. The solvent globule size also serves to limit the size of the polymer. A typical synthesis sequence might include an initiation and reaction stage, followed by removal of the aqueous solution, removing the solvent (and recycling it), drying the product, and finally, extruding the polycarbonate residue.

W21.22 Synthesis of Polystyrene

Several methods are available for the synthesis of polystyrene (see Fig. 14.1). One may polymerize it in a solution, suspension, or in bulk. The monomer is styrene ($C_6H_5CHCH_2$), which will be denoted by M. In free-radical polymerization a free-radical initiator such as benzoyl peroxide [$(RCOO)_2$, with $R = C_6H_5$] is added. The initiator thermally dissociates into two free radicals according to the reaction

$$(RCOO)_2 \longrightarrow 2(RCOO\cdot) \longrightarrow 2R\cdot + 2(CO_2). \tag{W21.54}$$

The free radicals are like ordinary molecules but have one unpaired electron. The free radical combines with the monomer to create a longer radical:

$$R\cdot + M \longrightarrow M_1\cdot. \tag{W21.55}$$

This in turn can link up to another monomer in a chain-growth process to create a still longer radical:

$$M_1\cdot + M \longrightarrow M_2\cdot, \tag{W21.56}$$

and the polymer continues to grow, one monomer at a time. After n steps one has

$$M_{n-1}\cdot + M \longrightarrow M_n\cdot. \tag{W21.57}$$

The growth eventually stops by a competing termination reaction. Two such reactions are chain coupling and disproportionation. In the chain-coupling reaction, two polymeric radicals combine, the electron spins become paired, and the growth is terminated:

$$M_n\cdot + M_m\cdot \longrightarrow M_{m+n}. \tag{W21.58}$$

In the disproportionation reaction a hydrogen atom is transferred from one radical to another. The net result is

$$M_n\cdot + M_m\cdot \longrightarrow M_n + M_m. \tag{W21.59}$$

This could involve changes of bond order within the resulting molecules. Alternatively, termination could be induced through a transfer reaction by introducing a molecule of the generic form $R'R''$ with an acceptor group R':

$$M_n\cdot + R'R'' \longrightarrow M_nR' + R''. \tag{W21.60}$$

Typical chain-transfer molecules include ethylbenzene, CCl_4, and CBr_4.

To prevent spontaneous polymerization, styrene is introduced into the processing environment with a retardant, such as benzoquinone. This retardant is stripped away before the styrene is introduced into the reactor. Also introduced are the initiator and the transfer molecule. After the polymerization is complete, the volatile chemicals are distilled out and the polymer is fed to an extruder, which produces the end product — pellets.

W21.23 Synthesis of Electro-active Polymers

Included among the electro-active polymers are *cis*- and *trans*-polyacetylene [$(CH)_x$], polythiophene (PT), polyaniline (PAN), polypyrrole (PPY), poly(*para*-pyridine) (PPPyr), poly(*para*-phenylene) (PPP), and its decorated variant poly(2-methoxy-5-(2′-ethylhexyloxy)-1,4-phenylene vinylene) (MEH-PPP) (Fig. W21.35). They tend to have delocalized π electrons distributed over the polymer and filling the valence band. The conduction band consists of the antibonding π^* orbitals and is empty. The polymers are therefore one-dimensional semiconductors.

The electrical activity comes about when the polymers are doped. Unlike the case of solid-state semiconductors where the dopant atoms are introduced into the lattice, these polymers are doped by placing atoms into the space between polymer chains. If the dopant is a donor, it donates an electron to the π^* conduction band. The polymer is thereby reduced. Typical elemental donor atoms are the alkalis Li, Na, and K, and the alkaline earth Ca. If the dopant is an acceptor, it creates a hole in the π valence band.

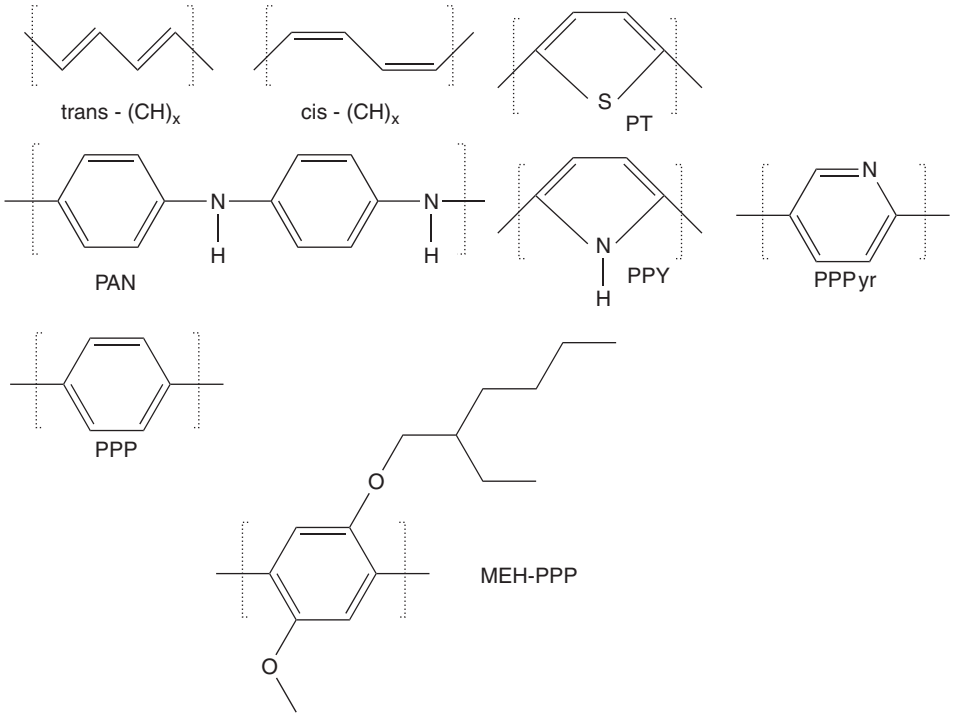

Figure W21.35. Some electroactive polymers.

The polymer is thereby oxidized. Typical elemental acceptors include the halogens Cl and I.

The doping may be introduced simply by exposing the polymer to a vapor or solution containing the dopant and allowing the dopant to diffuse into the bulk. Alternatively, electrochemical injection could be used. A "battery" is constructed with one electrode being the polymer and the other electrode containing the dopant. For example, to dope t-$(CH)_x$ p-type one takes one electrode to be t-$(CH)_x$ and the other electrode to be Li. An electrolyte such as $LiClO_4$ dissolved in propylene carbonate could be used. One proceeds to charge this battery. The Li electrode is biased sufficiently negative to reduce Li^+ ions. The battery pumps n electrons from the $(CH)_x$, leaving it in as positively charged $(CH)_x^{n+}$, thereby oxidizing it. For each polymer molecule n ClO_4^- ions drift through the electrolyte over to the $(CH)_x^{n+}$ and diffuse into it to preserve charge neutrality. The reactions for p-doping may be summarized as $nLi^+ + ne^- \rightarrow nLi$ and $(CH)_x \rightarrow (CH)_x^{n+} + ne^-$.

Light-emitting diodes can be fabricated by placing a film of polymer between two electrodes. The electrodes are chosen so that the anode Fermi level lies slightly above the π valence band of the polymer and the cathode Fermi level lies below the π^* conduction band. Schottky barriers are formed. Application of a forward bias then allows electrons to tunnel via Fowler–Nordheim tunneling from the cathode into the conduction band. Similarly, electrons tunnel from the valence band into the anode, leaving behind holes. The electrons and holes drift in opposite directions and radiatively combine in the interior of the film. A typical polymer that is used is MEH-PPV. Indium

tin oxide (ITO) is used as an anode because of its transparency to light. A typical cathodic material is Ca. A pattern of ITO anodes can be deposited on a substrate using photolithographic techniques. The polymer is then spin cast onto the device. Finally, the cathodic metal can be deposited from a vapor.

In light-emitting diodes the color of the light is determined by the bandgap. By using other polymers or adding side groups onto the polymer, one may tune this bandgap to other values. For example, MEH-PPV emits red light, whereas its mother, PPV, emits green light. The polymers PVK [poly(*N*-vinylcarbazole)] and PPP emit blue light.

W21.24 Spin Coating

The rheology of polymers puts severe constraints on possible processing methods. Above the glass-transition temperature, T_g, the polymer is a viscoelastic fluid that cannot simply be poured into molds. Of course, the temperature could be elevated to reduce the viscosity, but this also entails the risk of breaking the polymer chains or thermally inducing other unwanted chemical reactions. The methods used to form the materials into useful shapes involve various forms of extrusion, injection molding, blow molding, compression molding, coating, fiber spinning, thermoforming, and calendering. These manufacturing processes are not considered here. Instead, the focus of attention will be on the technique of spin coating, which has proven to be valuable in fabricating microelectronic circuits.

For a simple Newtonian fluid the stress and the strain rate are proportional to each other. To obtain a high strain rate, one must apply a large stress. If one wished to push the fluid through a die, one could do so by maintaining a pressure differential ΔP across its ends. For a die in the shape of a circular pipe of radius R and length Δz, Poisseuille's law gives the flow rate Q (volume/time) as

$$Q = -\frac{\pi R^4 \Delta P}{8\eta \Delta z}, \tag{W21.61}$$

where η is the viscosity. Clearly, a high-pressure differential favors a rapid throughput of material.

For polymers above T_g, the stress and the strain rates are not linearly related. An empirical formula is

$$\sigma = \lambda \left(-\frac{\partial u}{\partial r}\right)^n, \tag{W21.62}$$

where u is the axial velocity component, r the radial distance from the axis, and λ a constant depending on the polymer. The exponent n can be as small as 0.1. One may regard this as a formula defining a strain-rate-dependent viscosity $\eta = \lambda(-\partial u/\partial r)^{n-1}$. As the strain rate increases, the polymer becomes less viscous. By balancing forces on a coaxial cylindrical section of length Δz and radius r, one finds that $\sigma = -r\Delta P/(2\Delta z)$. Combining this with the previous formula gives

$$u(r) = -\frac{1}{2\lambda}\frac{\Delta P}{\Delta z}\frac{n}{n+1}(R^{1+1/n} - r^{1+1/n}). \tag{W21.63}$$

The flow rate is found to be

$$Q = \frac{n\pi}{3n+1}\left(-\frac{1}{2\lambda}\frac{\Delta P}{\Delta z}\right)^{1/n} R^{3+1/n}. \tag{W21.64}$$

For $n = 1$ this reduces to the Poisseuille law, with $\lambda = n$. For $n \neq 1$ it is a suitable generalization.

In the spin-coating process the substrate is mounted on a rotating turntable that is spun at angular velocity ω (Fig. W21.36). A polymer is poured on the axial region and it flows radially outward as it revolves around the axis. The centripetal force on any volume of polymer is due to the viscous force.

The radial component of the velocity is determined by solving the equation

$$\frac{\partial}{\partial z}\left[\eta\left(\frac{\partial v_r}{\partial z}\right)\right] = -r\rho\omega^2, \tag{W21.65}$$

where v_r is zero at the surface of the substrate ($z = 0$) and reaches a maximum value at the surface of the coating. Integration leads to

$$v_r(r, z) = \frac{n}{n+1}\frac{r\omega^2\rho}{\lambda}\left[h^{1+1/n} - (h-z)^{1+1/n}\right], \tag{W21.66}$$

where h is the thickness of the coating. The volume flow rate, Q, is given by

$$Q(r) = \int_0^h v_r 2\pi r\, dz = \frac{n}{2n+1}\frac{2\pi r^2\omega^2\rho}{\lambda}h^{2+1/n}, \tag{W21.67}$$

so

$$h(r) = \left(\frac{2n+1}{n}\frac{\lambda Q}{2\pi\rho r^2\omega^2}\right)^{n/(2n+1)}. \tag{W21.68}$$

When the spinning stops, the surface tension will flatten the surface so as to minimize the surface energy. The time scale for this relaxation may be expressed using dimensional analysis in the form $\tau = \eta D/\sigma$, where D is a characteristic dimensionless scale length formed from R and the volume of the film.

Suppose that there is a uniform coating of thickness h_0 at time $t = 0$. Allow the spinning to occur so that an excess of polymer will flow over the edge. Apply continuity

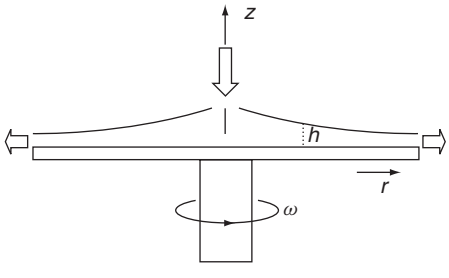

Figure W21.36. Spin-coating process.

concepts to develop an equation for h as a function of time. Consider the rate of change of the volume above the annulus lying between r and $r + dr$:

$$\frac{\partial(dV)}{\partial t} = 2\pi r\, dr \frac{\partial h}{\partial t}. \tag{W21.69}$$

This change would be caused by a net flux into the cylinder. The flux coming from the inside is

$$\Phi(r) = \int_0^h v_r(r, z) 2\pi r\, dz = 2\pi r^2 \frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} h^{(2n+1)/n}. \tag{W21.70}$$

The continuity equation states that

$$\Phi(r) - \Phi(r + dr) = \frac{\partial(dV)}{\partial t}, \tag{W21.71}$$

so

$$\frac{\partial h}{\partial t} = -\frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} \frac{1}{r} \frac{\partial}{\partial r} (r^2 h^{(2n+1)/n}). \tag{W21.72}$$

A solution of this equation $h(t)$, independent of r, is

$$h(t) = \left(h_0^{-(n+1)/n} + \frac{2n+2}{2n+1} \frac{\rho \omega^2}{\lambda} t \right)^{-n/(n+1)}, \tag{W21.73}$$

where $h(0) = h_0$. In the case of a Newtonian fluid, $n = 1$ and $\lambda = \eta$. The volume flow per unit time over the outer rim of the disk, of radius R, is

$$\Phi(R) = \int_0^h v_r(R, z) 2\pi R\, dz = 2\pi R^2 \frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} h^{(2n+1)/n}. \tag{W21.74}$$

W21.25 Microwave and Plasma Processing of Polymers

Microwave Processing. Often, it is advantageous to heat a material to remove defects, alter the crystallinity, control the morphology, promote drying, accelerate a polymerization reaction, and so on. The problem with polymers, however, is that they tend to be poor thermal conductors. The temperature that needs to be reached is usually close to the ceiling temperature T_c, the temperature at which there is irreversible chemical damage to the polymer. Heating the surface of the material leads to thermal gradients, so the heating rate must be made very slow to avoid having parts of the material with $T > T_c$. Microwaves offer the advantage of being able to heat a sample uniformly, at least over a skin depth, making it possible to achieve much more rapid heating rates.

Microwave radiation couples to molecular groups on the polymer that possess electric-dipole moments, such as OH, CN, or NH_2. Once local vibrations, librations, or rotations of these groups are excited, they dissipate the energy to the rest of the polymer chain by radiating one-dimensional phonons. The radiation also couples to ions and produces ac currents. The moving ions collide with the polymer chain, also creating phonons. In the liquid phase the dipole moments are able to reorient themselves readily, and this gives rise to strong coupling to the microwave field. As the temperature is

lowered and the polymer becomes more rubbery, the groups lose some of their flexibility and the absorption is reduced. If the liquid is cooled below the glass-transition temperature T_g, more of the degrees of freedom are removed and the dipolar coupling becomes still less. In the crystalline phase the steric hindrance becomes even more severe and the coupling drops even further. Often, additives are added to enhance the dipolar absorption during processing. These include metallic particles, carbon black, or carbon fibers. If conducting particles are added and these percolate throughout the material, eddy currents may be established by the microwave fields. These currents produce ohmic heating.

The skin depth δ is given in terms of the ac conductivity σ by the formula $\delta = 1/\sqrt{2\omega\mu_0\sigma}$. The value of σ is determined by the concentration of carbon black, so the value of δ may be chosen so it is larger than the sample thickness. Within the sample the intensity of the radiation falls off according to Beer's law, $I(z) = I_0 \exp(-z/\delta)$. In the limit of a material of small thickness, obeying $d \ll \delta$, the power absorbed per unit area is given by $P/A = \sigma|E_0|^2 d/2$, where E_0 is the electric field in the medium.

An important application of microwave radiation is in curing the polyimide thin films used in the design of electronic chips. Microwaves may also be used to ensure a uniformity of epoxy resins that are thermoset. In composite materials uniform curing is important to improve the fracture properties. Thus microwave radiation provides a valuable processing tool for polymers.

Plasma Processing. Whereas microwaves are most useful in processing the bulk of polymers, plasma processing is used to modify the surfaces of polymers. The plasma may be used to clean the polymer surface. It may be used selectively to remove unwanted polymers of low molecular weight (oligomers). It is often used to remove photoresist from electronic chips as part of the lithography process. It may serve to etch new topographic features into the surface or to sputter away old ones. It may be employed to alter the surface chemically, such as by adding new chemical groups to the polymer. It may also be used to deposit protective coatings or coatings on the surface that modify its physical properties.

The reason for the vulnerability of small molecules to plasma etching has to do with their inability to dissipate energy via phonons. The longer-chain molecules may conduct thermal energy along their chains to the interior of the polymer. The shorter chains are not able to do so. Instead, the vibrational temperature of the oligomers is increased by exposure to the plasma until they dissociate. Frequently, gaseous products such as CH_4 and H_2 or free radicals such as CH_2 are produced. For the sputtering process Ar^+ plasmas are used. Since Ar is inert, it does not contaminate the underlying surface by chemically bonding to it.

Surfaces may be reactively etched by using plasmas with ions such as O_2^+, SF_6^+, or CF_4^+. The microstructure of the surface may be suitably changed. One may roughen a surface to increase its surface area so that a coating will adhere to it better.

At times the surface is chemically altered. It is first exposed to process gases that adsorb on the surface and create free radicals. Gases such as O_2, N_2, NH_3, and CF_4 are used. Subsequent exposure to the plasma provides the activation energy that permits the polymer chain to acquire new functional groups.

An example of a coating that may be deposited using plasmas is PMMA. The polymerization reaction, in which gaseous methylmethacrylate molecules are polymerized

into long chains of polymethylmetacrylate (PMMA), could be initiated by exposure to a plasma.

REFERENCES

Synthesis and Processing Procedures

Venables, J. A., in C. B. Duke, ed., *Surface Science: The First Thirty Years*, North-Holland, Amsterdam, 1994, p. 798.

Venables, J. A., G. D. T. Spiller, and M. Hanbucken, *Rep. Prog. Phys.*, **47**, 399 (1984).

Voorhoeve, R. J. H., Molecular beam deposition of solids on surfaces: ultrathin films, in N. B. Hannay, ed., *Treatise in Solid State Chemistry*, Vol. 6A, Plenum Press, New York, 1976.

Synthesis and Processing of Semiconductors

Lieberman, M. A., and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, Wiley, New York, 1994.

Maly, W., *Atlas of IC Technologies: An Introduction to VLSI Processes*, Benjamin-Cummings, Menlo Park, Calif., 1987.

Meyerson, B. S., Low-temperature Si and Si:Ge ultrahigh-vacuum/chemical vapor deposition: process fundamentals, *IBM J. Res. Dev.*, **14**, 806 (1990).

Pamplin, B. R., ed., *Molecular Beam Epitaxy*, Pergamon Press, Oxford, 1980.

Shimura, F., ed., *Oxygen in Silicon*, Vol. 42 of R. K. Willardson, A. C. Beer, and E. R. Weber, eds., *Semiconductors and Semimetals*, Academic Press, San Diego, Calif., 1994.

Stringfellow, G. B., *Organometallic Vapor-Phase Epitaxy: Theory and Practice*, Academic Press, San Diego, Calif., 1989.

Weissler, G. L., and R. W. Carlson, eds., *Vacuum Physics and Technology*, Vol. 14 of *Methods of Experimental Physics*, Academic Press, San Diego, Calif., 1979.

Wolf, S., and R. N. Tauber, *Silicon Processing for the VLSI Era*, Vol. 1, *Process Technology*; S. Wolf, *ibid.*, Vol. 2, *Process Integration*, Lattice Press, Sunset Beach, Calif., 1986 (Vol. 1), 1990 (Vol. 2).

Synthesis and Processing of Metals

Honeycombe, R. W. K., and H. K. D. H. Bhadeshia, *Steel: Microstructure and Properties*, 2nd ed., Edward Arnold, London, 1996.

Jacobson, L. A., and J. McKittrick, Rapid solidification processing, *Mater. Sci. Eng.*, **R11**, 355 (1994).

Silicon Nitride

Weimer, A. W., ed., *Carbide, Nitride and Boride Materials Synthesis and Processing*, Chapman & Hall, London, 1997.

PLZT

Beltram, T., M. Kosec, and S. Stavber, *Mater. Res. Bull.*, **28**, 313 (1993).

Plasma Processing of Polymers

Coates, D. M., and S. L. Kaplan, Modification of polymeric surfaces with plasmas, *Mater. Res. Soc. Bull.*, Aug. 1996, p. 43.

Microwave Processing of Polymers

Lewis, D. A., and J. M. Shaw, Recent development in the microwave processing of polymers, *Mater. Res. Soc. Bull.*, Nov. 1993, p. 37.

PROBLEMS

W21.1 Calculate the equilibrium constant K for the reaction $SiO_2(s) \leftrightarrow Si(s) + O_2(g)$ at $T = 1300$ K and also the $O_2(g)$ equilibrium vapor pressure $P_{eq}[O_2(g), -1300\ K]$. Use $\Delta_f G^o[SiO_2(s), 1300\ K] = -678.535$ kJ/mol for high cristobalite, the stable form of $SiO_2(s)$ at $T = 1300$ K. This problem illustrates that $SiO_2(s)$ is very stable against decomposition into its elemental components $Si(s)$ and $O_2(g)$.

W21.2 The heat or enthalpy of formation $\Delta_f H^o$ of $Si(g)$ vapor is essentially independent of temperature from $T = 0$ K up to the melting temperature $T_m = 1414°C$ and is approximately equal to (448 ± 2) kJ/mol. Convert this result from kJ/mol to eV/atom and compare the result with the Si–Si bond energy given in Chapter 2, [i.e., $E(Si–Si) = 2.34$ eV]. From what you know about the crystal structure and bonding in $Si(s)$, what can you conclude about the vaporization of $Si(s)$ from this comparison?

W21.3 Consider the equilibrium phase diagram for the Si–A system shown schematically in Fig. W21.8, where the liquidus and solidus curves are both nearly straight lines for low concentrations of element A in Si, with negative slopes s_L and s_S. Show that the distribution coefficient K_A of A in Si is given by the ratio of the slopes of these two lines [i.e., by Eq. (W21.14)]. (*Hint*: Review the discussion of equilibrium binary phase diagrams in Section 6.5.)

W21.4 For the thermal oxidation of Si using dry O_2:

(a) Derive the expression for the thickness $x(t)$ of the growing a-SiO_2 layer given in Eq. (21.53).

(b) Derive the expressions for $x(t)$ given in Eq. (21.54) for thin layers and in Eq. (21.55) for thicker layers.

W21.5 Calculate the mass densities of α-Fe at $T = 20°C$, γ-Fe at $T = 912°C$, δ-Fe at $T = 1394°C$, and Fe_3C at $T = 20°C$ (four formula units per orthorhombic unit cell), and of martensite at $T = 20°C$ for $x = 0.02$ and 0.04 using the structural data given in Table W21.5.

W21.6 Propose and explain a mechanism by which the alternating layers of ferrite and cementite found in the pearlite shown in Fig. 21.11 can be formed when austenite is slowly cooled through the eutectoid temperature T_e.

W21.7 For the situation illustrated in Fig. W21.28, where an ion beam deposits 10 J/cm^2 of energy into the surface of a tool steel, calculate the thickness d of the surface layer that can be melted. To simplify the calculation, assume that the steel is pure Fe, its specific heat is constant from $T = 300$ K up to $T_m = 1538$ K, and the energy is deposited uniformly within the thickness d that is melted.