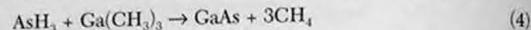


To grow GaAs, we can use metalorganic compounds such as trimethylgallium [$\text{Ga}(\text{CH}_3)_3$] for the gallium component and arsine (AsH_3) for the arsenic component. Both chemicals can be transported in vapor form into the reactor. The overall reaction is



For Al-containing compounds, such as AlAs, trimethylaluminum [$\text{Al}(\text{CH}_3)_3$] can be used. During epitaxy, the GaAs is doped by introducing dopants in vapor form. Diethylzinc [$\text{Zn}(\text{C}_2\text{H}_5)_2$] and diethylcadmium [$\text{Cd}(\text{C}_2\text{H}_5)_2$] are typical *p*-type dopants, and silane (SiH_4) is an *n*-type dopant for III-V compounds. The hydrides of sulfur and selenium or tetramethyltin are also used for *n*-type dopants, and chromyl chloride is used to dope chromium into GaAs to form semiinsulating layers. Since these compounds are highly poisonous and often spontaneously inflammable in air, rigorous safety precautions are necessary in the MOCVD process.

Figure 8.4 shows a schematic of an MOCVD reactor.³ Typically, the metalorganic compound is transported to the quartz reaction vessel by hydrogen carrier gas, where it is mixed with AsH_3 in the case of GaAs growth. The chemical reaction is induced by heating the gases to 600°C to 800°C above a substrate placed on a graphite susceptor using radio frequency heating. A pyrolytic reaction forms the GaAs layer. The advantages of using metalorganics are that they are volatile at moderately low temperatures and that there are no troublesome liquid Ga or In sources in the reactor.

8.1.2 Molecular Beam Epitaxy

MBE is an epitaxial process involving the reaction of one or more thermal beams of atoms or molecules with a crystalline surface under ultrahigh-vacuum conditions ($\sim 10^{-8}$ Pa).⁴

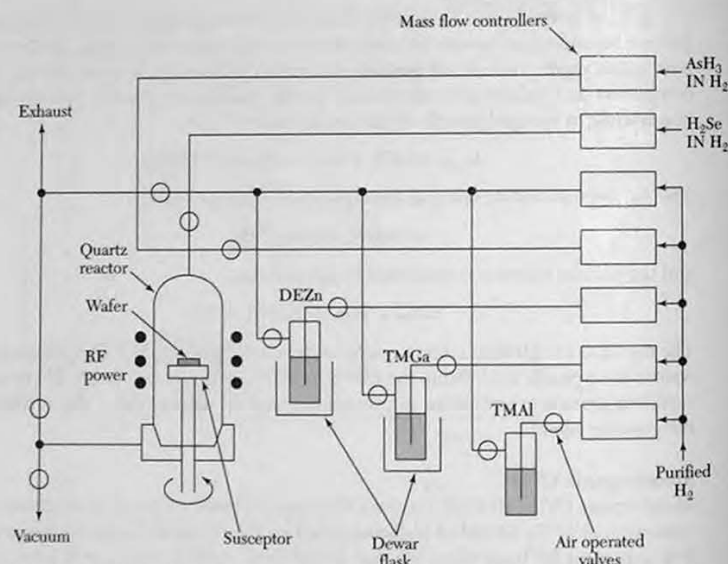


Figure 8.4 Schematic diagram of a vertical atmospheric-pressure MOCVD reactor.³ DEZn, diethylzinc [$\text{Zn}(\text{C}_2\text{H}_5)_2$]; TMGa, trimethylgallium [$\text{Ga}(\text{CH}_3)_3$]; TMAI, trimethylaluminum [$\text{Al}(\text{CH}_3)_3$].

MBE can achieve precise control of both chemical compositions and doping profiles. Single-crystal, multilayer structures with dimensions on the order of atomic layers can be grown using MBE. Thus, the MBE method enables the precise fabrication of semiconductor heterostructures having thin layers from a fraction of a micron down to a monolayer. In general, MBE growth rates are quite low. For GaAs, for example, a value of 1 $\mu\text{m/hr}$ is typical.

Figure 8.5 shows a schematic of an MBE system for gallium arsenide and related III-V compounds, such as $\text{Al}_x\text{Ga}_{1-x}\text{As}$. The system represents the ultimate in film deposition control, cleanliness, and in situ chemical characterization capability. Separate effusion ovens made of pyrolytic boron nitride are used for Ga, As, and the dopants. All the effusion ovens are housed in an ultrahigh-vacuum chamber ($\sim 10^{-8}$ Pa). The temperature of each oven is adjusted to give the desired evaporation rate. The substrate holder rotates continuously to achieve uniform epitaxial layers (e.g., $\pm 1\%$ in doping variations and $\pm 0.5\%$ in thickness variations).

To grow GaAs, an overpressure of As is maintained, since the sticking coefficient of Ga to GaAs is unity, whereas that for As is zero, unless there is a previously deposited Ga layer. For a silicon MBE system, an electron gun is used to evaporate silicon. One or more effusion ovens are used for the dopants. Effusion ovens behave like small-area sources and exhibit a $\cos\theta$ emission, where θ is the angle between the direction of the source and the normal to the substrate surface.

MBE uses an evaporation method in a vacuum system. An important parameter for vacuum technology is the molecular impingement rate; that is, how many molecules impinge on a unit area of the substrate per unit time. The impingement rate (ϕ) is a function of the molecular weight, temperature, and pressure. The rate is derived in Appendix H and can be expressed as⁵

$$\phi = P(2\pi mkT)^{-1/2} \quad (5)$$

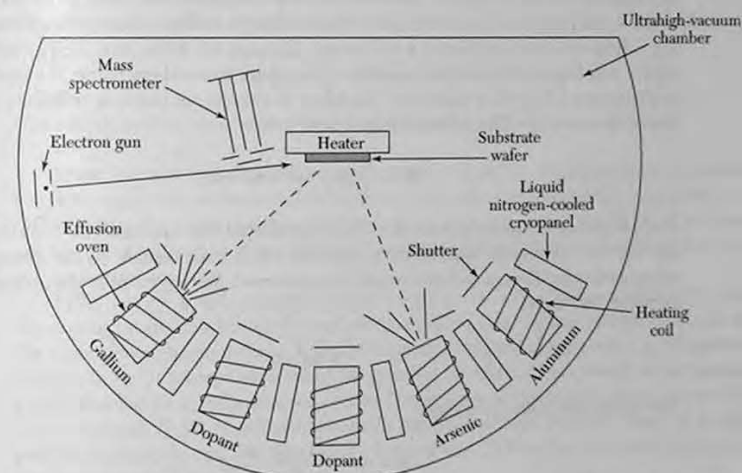


Figure 8.5 Arrangement of the sources and substrate in a conventional molecular beam epitaxy (MBE) system. (Courtesy of M. B. Panish, Bell Laboratories, Lucent Technologies.)

or

$$\phi = 2.64 \times 10^{20} \left(\frac{P}{\sqrt{MT}} \right) \text{ molecules/cm}^2\text{-s} \quad (5a)$$

where P is the pressure in Pa, m is the mass of a molecule in kg, k is Boltzmann's constant in J/K, T is the temperature in degrees Kelvin, and M is the molecular weight. Therefore, at 300 K and 10^{-4} Pa pressure, the impingement rate is 2.7×10^{14} molecules/cm²-s for oxygen ($M = 32$).

EXAMPLE 1

At 300 K, the molecular diameter of oxygen is 3.64 Å, and the number of molecules per unit area N_s is 7.54×10^{14} cm⁻². Find the time required to form a monolayer of oxygen at pressures of 1, 10^{-4} , and 10^{-5} Pa.

SOLUTION The time required to form a monolayer (assuming 100% sticking) is obtained from the impingement rate:

$$t = \frac{N_s}{\phi} = \frac{N_s \sqrt{MT}}{2.64 \times 10^{20} P}$$

Therefore,

$$\begin{aligned} t &= 2.8 \times 10^{-4} = 0.28 \text{ ms} \quad \text{at } 1 \text{ Pa} \\ &= 2.8 \text{ s} \quad \text{at } 10^{-4} \text{ Pa} \\ &= 7.7 \text{ hr} \quad \text{at } 10^{-5} \text{ Pa} \end{aligned}$$

To avoid contamination of the epitaxial layer, it is of paramount importance to maintain ultrahigh-vacuum conditions ($\sim 10^{-5}$ Pa) for the MBE process.

During molecular motion, molecules will collide with one another. The average distance traversed by all the molecules between successive collisions with each other is defined as the *mean free path*. It can be derived from a simple collision theory. A molecule having a diameter d and a velocity v will move a distance $v\delta t$ in the time δt . The molecule suffers a collision with another molecule if its center is anywhere within the distance d of the center of another molecule. Therefore, it sweeps out (without collision) a cylinder of diameter $2d$. The volume of the cylinder is

$$\delta V = \frac{\pi}{4} (2d)^2 v \delta t = \pi d^2 v \delta t \quad (6)$$

Since there are n molecules/cm³, the volume associated with one molecule is on the average $1/n$ cm³. When the volume δV is equal to $1/n$, it must contain on the average one other molecule. Thus, a collision would have occurred. Setting $\tau = \delta t$ as the average time between collisions,

$$\frac{1}{n} = \pi d^2 v \tau \quad (7)$$

The mean free path (λ) is then

$$\lambda = v\tau = \frac{1}{\pi n d^2} = \frac{kT}{\pi P d^2} \quad (8)$$

A more rigorous derivation gives

$$\lambda = \frac{kT}{\sqrt{2} \pi P d^2} \quad (9)$$

and

$$\lambda = \frac{0.66}{P(\text{in Pa})} \text{ cm} \quad (10)$$

for air molecules (equivalent molecular diameter of 3.7 Å) at room temperature. Therefore, at a system pressure of 10^{-5} Pa, λ would be 660 km.

EXAMPLE 2

Assume an effusion oven geometry of area $A = 5$ cm² and a distance L between the top of the oven and the gallium arsenide substrate of 10 cm. Calculate the MBE growth rate for the effusion oven filled with gallium arsenide at 900°C. The surface density of gallium atoms is 6×10^{14} cm⁻², and the average thickness of a monolayer is 2.8 Å.

SOLUTION On heating gallium arsenide, the volatile arsenic vaporizes first, leaving a gallium-rich solution. Therefore, only the pressures marked "Ga-rich" in Figure 2.11 are of interest. The pressure at 900°C is 5.5×10^{-2} Pa for gallium and 1.1 Pa for arsenic (As₂). The arrival rate can be obtained from the impingement rate (Eq. 5a) by multiplying it by $A/\pi L^2$:

$$\text{Arrival rate} = 2.64 \times 10^{20} \left(\frac{P}{\sqrt{MT}} \right) \left(\frac{A}{\pi L^2} \right) \text{ molecules/cm}^2\text{-s}$$

The molecular weight (M) is 69.72 for Ga and 74.92×2 for As₂. Substituting values of P , M , and T (1173 K) into the above equation gives

$$\begin{aligned} \text{Arrival rate} &= 8.2 \times 10^{14} / \text{cm}^2\text{-s for Ga} \\ &= 1.1 \times 10^{16} / \text{cm}^2\text{-s for As}_2 \end{aligned}$$

The growth rate of gallium arsenide is governed by the arrival rate of gallium. The growth rate is

$$\frac{8.2 \times 10^{14} \times 2.8}{6 \times 10^{14}} = 3.8 \text{ Å/s} = 23 \text{ nm/min}$$

Note that the growth rate is relatively low compared with that of VPE.

There are two ways to clean a surface in situ for MBE. High-temperature baking can decompose the native oxide and remove other adsorbed species by evaporation or diffusion into the wafer. Another approach is to use a low-energy ion beam of an inert gas to sputter-clean the surface, followed by a low-temperature annealing to reorder the surface lattice structure.

MBE can use a wide variety of dopants (compared with CVD and MOCVD), and the doping profile can be exactly controlled. However, the doping process is similar to the vapor-phase growth process: A flux of evaporated dopant atoms arrives at a favorable lattice site and is incorporated along the growing interface. Fine control of the doping profile is achieved by adjusting the dopant flux relative to the flux of silicon atoms (for silicon epitaxial films) or gallium atoms (for gallium arsenide epitaxial films). It is also possible to dope the epitaxial film using a low-current, low-energy ion beam to implant the dopant (see Chapter 7).

The substrate temperatures for MBE range from 400°C to 900°C, and the growth rates range from 0.001 to 0.3 $\mu\text{m}/\text{min}$. Because of low-temperature processing and the low growth rate, many unique doping profiles and alloy compositions not obtainable from conventional CVD can be produced in MBE. Many novel structures have been made using MBE. These include the *superlattice*, which is a periodic structure consisting of alternating ultrathin layers with a period less than the electron mean free path (e.g., GaAs/Al_xGa_{1-x}As, with each layer 10 nm or less in thickness), and heterojunction field-effect transistors.

A further development in MBE has replaced the group III elemental sources by metalorganic compounds such as trimethylgallium (TMG) or triethylgallium (TEG). This approach is called *metalorganic molecular beam epitaxy* (MOMBE) and is also referred to as *chemical beam epitaxy* (CBE). Although closely related to MOCVD, it is considered a special form of MBE. The metalorganics are sufficiently volatile that they can be admitted directly into the MBE growth chamber as a beam and are not decomposed before forming the beam. The dopants are generally elemental sources, typically Be for *p*-type and Si or Sn for *n*-type GaAs epitaxial layers.

► 8.2 STRUCTURES AND DEFECTS IN EPITAXIAL LAYERS

8.2.1 Lattice-Matched and Strained-Layer Epitaxy

For conventional homoepitaxial growth, a single-crystal semiconductor layer is grown on a single-crystal semiconductor substrate. The semiconductor layer and the substrate are the same material having the same lattice constant. Therefore, homoepitaxy is, by definition, a lattice-matched epitaxial process. The homoepitaxial process offers one important means of controlling the doping profiles so that device and circuit performance can be optimized. For example, an *n*-type silicon layer with a relatively low doping concentration can be grown epitaxially on an *n*⁺ silicon substrate. This structure substantially reduces the series resistance associated with the substrate.

For heteroepitaxy, the epitaxial layer and the substrate are two different semiconductors, and the epitaxial layer must be grown in such a way that an idealized interfacial structure is maintained. This implies that atomic bonding across the interface must be continuous. Therefore, the two semiconductors must either have the same lattice spacing or be able to deform to adopt a common spacing. These two cases are referred to as *lattice-matched epitaxy* and *strained-layer epitaxy*, respectively.

Figure 8.6a shows a lattice-matched epitaxy where the substrate and the film have the same lattice constant. An important example is the epitaxial growth of Al_xGa_{1-x}As on a GaAs substrate where for any *x* between 0 and 1, the lattice constant of Al_xGa_{1-x}As differs from that of GaAs by less than 0.13%.

For the lattice-mismatched case, if the epitaxial layer has a larger lattice constant and is flexible, it will be compressed in the plane of growth to conform to the substrate spacing. Elastic forces then compel it to dilate in a direction perpendicular to the interface. This type of structure is called *strained-layer epitaxy* and is illustrated in Figure 8.6b. On the other hand, if the epitaxial layer has a smaller lattice constant, it will be dilated in the plane of growth and compressed in a direction perpendicular to the interface. In the above strained-layer epitaxy, as the strained-layer thickness increases, the total number of atoms under strain of the distorted atomic bonds grows, and at some point, misfit dislocations are nucleated to relieve the homogeneous strain energy. This thickness is referred to as the *critical layer thickness* for the system. Figure 8.6c shows the case in which there are edge dislocations at the interface.

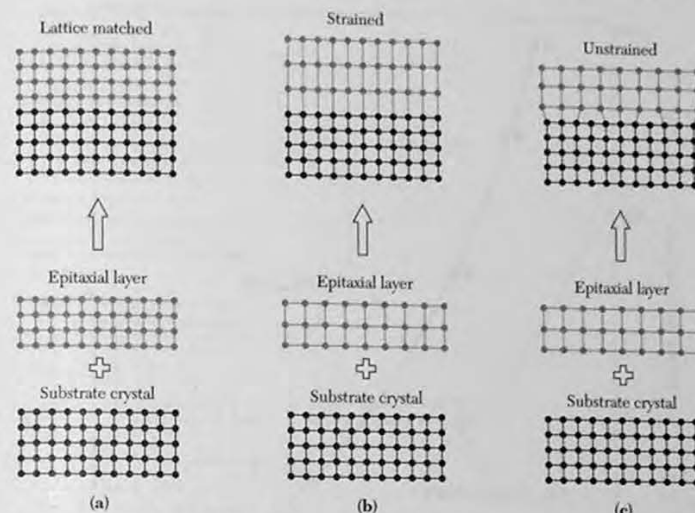


Figure 8.6 Schematic illustration of (a) lattice-matched, (b) strained,⁶ and (c) relaxed heteroepitaxial structures. Homoepitaxy is structurally identical to the lattice-matched heteroepitaxy.

The critical layer thicknesses for two material systems are shown in Figure 8.7.⁷ The upper curve is for the strained-layer epitaxy of a Ge_xSi_{1-x} layer on a silicon substrate, and the lower curve is for a Ga_{1-x}In_xAs layer on a GaAs substrate. For example, for Ge_{0.3}Si_{0.7} on silicon, the maximum epitaxial thickness is about 70 nm. For thicker films, edge dislocations will occur.

A related heteroepitaxial structure is the *strained-layer superlattice* (SLS). A superlattice is an artificial one-dimensional periodic structure constituted by different materials with a period of about 10 nm. Figure 8.8 shows⁴ an SLS having two semiconductors with different equilibrium lattice constants $a_1 > a_2$, grown in a structure with a common inplane lattice constant b , where $a_1 > b > a_2$. For sufficiently thin layers, the lattice mismatch is accommodated by uniform strains in the layers. Under these conditions, no misfit dislocations are generated at the interfaces, so high-quality crystalline materials can be obtained. These artificially structured materials can be grown by MBE. These materials provide a new area in semiconductor research and permit new solid-state devices, especially for high-speed and photonic applications.

8.2.2 Defects in Epitaxial Layers

Defects in epitaxial layers will degrade device properties. For example, defects can result in reduced mobility or increased leakage current. The defects in epitaxial layers can be categorized into five groups:

1. *Defects from the substrates.* These defects may propagate from the substrate into the epitaxial layer. To avoid these defects, dislocation-free semiconductor substrates are required.

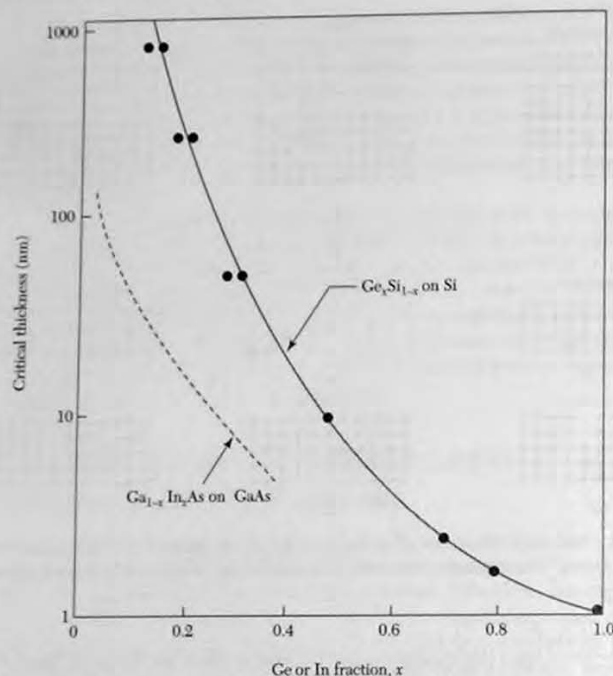


Figure 8.7 Experimentally determined critical layer thickness for defect-free, strained-layer epitaxy of $\text{Ge}_x\text{Si}_{1-x}$ on Si, and $\text{Ga}_{1-x}\text{In}_x\text{As}$ on GaAs.⁷

2. *Defects from the interface.* Oxide precipitates or any contamination at the interface of the epitaxial layer and substrate may cause the formation of mis-oriented clusters or nuclei containing stacking faults. These clusters and stacking faults may coalesce with normal nuclei and grow into the film in the shape of an inverted pyramid. To avoid these defects, the surface of the substrate must be thoroughly cleaned. In addition, an in situ etchback may be used, such as the reversible reaction of Eq. 1.
3. *Precipitates or dislocation loops.* Their formation is due to supersaturation of impurities or dopants. Epitaxial layers containing very high intentional or unintentional dopants or impurity concentrations are susceptible to such defects.
4. *Low-angle grain boundaries and twins.* Any misoriented areas of an epitaxial film during growth may meet and coalesce to form these defects.
5. *Edge dislocations.* These are formed in the heteroepitaxy of two lattice-mismatched semiconductors. If both lattices are rigid, they will retain their fundamental lattice spacings, and the interface will contain rows of misbonded atoms described as misfit or edge dislocations. The edge dislocations can also form in a strained layer when the layer thickness becomes larger than the critical layer thickness.

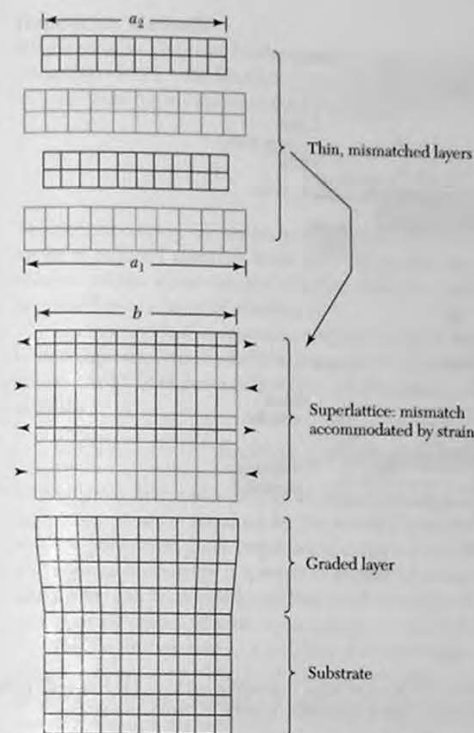


Figure 8.8 Illustration of the elements and formation of a strained-layer superlattice.⁴ Arrows show the direction of the strain.

► 8.3 DIELECTRIC DEPOSITION

Deposited dielectric films are used mainly for insulation and passivation of discrete devices and integrated circuits. There are three commonly used deposition methods: atmospheric-pressure CVD, low-pressure CVD (LPCVD), and plasma-enhanced chemical vapor deposition (PECVD, or plasma deposition). PECVD is an energy-enhanced CVD method in which plasma energy is added to the thermal energy of a conventional CVD system. Considerations in selecting a deposition process are the substrate temperature, the deposition rate and film uniformity, the morphology, the electrical and mechanical properties, and the chemical composition of the dielectric films.

The reactor for atmospheric-pressure CVD is similar to the one shown in Figure 3.2, except that different gases are used at the gas inlet. In a hot-wall, reduced-pressure reactor like the one shown in Figure 8.9a, the quartz-tube is heated by a three-zone furnace, and gas is introduced at one end and pumped out at the opposite end. The semiconductor wafers are held vertically in a slotted quartz boat.⁸ The quartz-tube wall is hot because it is adjacent to the furnace, in contrast to a cold-wall reactor such as the horizontal epitaxial reactor that uses radio frequency (rf) heating.

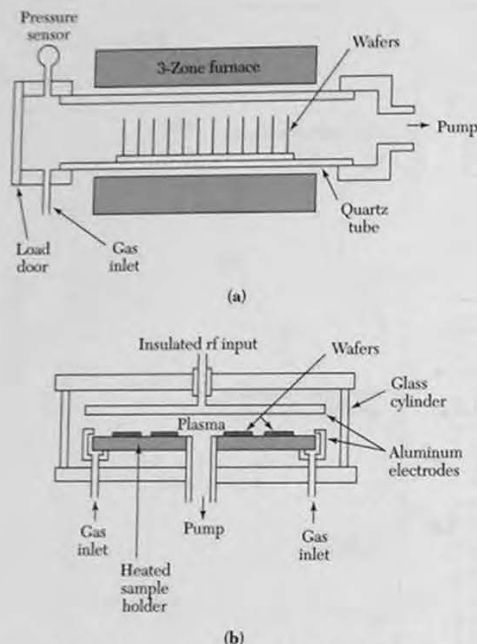


Figure 8.9 Schematic diagrams of chemical vapor deposition reactors. (a) Hot-wall, reduced-pressure reactor. (b) Parallel-plate plasma deposition reactor. rf, radio frequency.

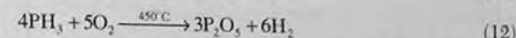
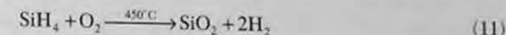
The parallel-plate, radial-flow PECVD reactor shown in Figure 8.9b consists of a cylindrical glass or aluminum chamber sealed with aluminum endplates. Inside are two parallel aluminum electrodes. An rf voltage is applied to the upper electrode, whereas the lower electrode is grounded. The rf voltage causes a plasma discharge between the electrodes. Wafers are placed on the lower electrode, which is heated between 100°C and 400°C by resistance heaters. The reaction gases flow through the discharge from outlets located along the circumference of the lower electrode. The main advantage of this reactor is its low deposition temperature. However, its capacity is limited, especially for large-diameter wafers, and the wafers may become contaminated if loosely adhering deposits fall on them.

8.3.1 Silicon Dioxide

CVD silicon dioxide cannot replace thermally grown oxides, because the best electrical properties are obtained with thermally grown films. CVD oxides are used instead to complement thermal oxides. A layer of undoped silicon dioxide is used to insulate multilevel metallization, to mask ion implantation and diffusion, and to increase the thickness of thermally grown field oxides. Phosphorus-doped silicon dioxide is used both as an insulator between metal layers and as a final passivation layer over devices. Oxides doped with phosphorus, arsenic, or boron are used occasionally as diffusion sources.

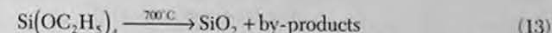
Deposition Methods

Silicon dioxide films can be deposited by several methods. For low-temperature deposition (300–500°C), the films are formed by reacting silane, dopant, and oxygen. The chemical reactions for phosphorus-doped oxides are



The deposition process can be performed either at atmospheric pressure in a CVD reactor or at reduced pressure in an LPCVD reactor (Fig. 8.9a). The low deposition temperature of the silane–oxygen reaction makes it a suitable process when films must be deposited over a layer of aluminum.

For intermediate-temperature deposition (500–800°C), silicon dioxide can be formed by decomposing tetraethylorthosilicate, $\text{Si}(\text{OC}_2\text{H}_5)_4$, in an LPCVD reactor. The compound, abbreviated TEOS, is vaporized from a liquid source. The TEOS compound decomposes as follows:

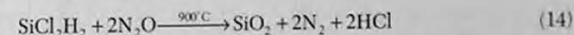


forming both SiO_2 and a mixture of organic and organosilicon by-products. Although the higher temperature required for the reaction prevents its use over aluminum, it is suitable for polysilicon gates requiring a uniform insulating layer with good step coverage. The good step coverage is a result of enhanced surface mobility at higher temperatures. The oxides can be doped by adding small amounts of the dopant hydrides (phosphines, arsine, or diborane), similar to the process in epitaxial growth.

The deposition rate as a function of temperature varies as $\exp(-E_a/kT)$, where E_a is the activation energy. The E_a of the silane–oxygen reaction is quite low: about 0.6 eV for undoped oxides and almost zero for phosphorus-doped oxide. In contrast, E_a for the TEOS reaction is much higher: about 1.9 eV for undoped oxide and 1.4 eV when phosphorus doping compounds are present. The dependence of the deposition rate on TEOS partial pressure is proportional to $(1 - e^{-P/P_0})$, where P is the TEOS partial pressure and P_0 is about 30 Pa. At low TEOS partial pressures, the deposition rate is determined by the rate of the surface reaction. At high partial pressures, the surface becomes nearly saturated with adsorbed TEOS, and the deposition rate becomes essentially independent of TEOS pressure.⁸

Recently, atmospheric-pressure and low-temperature CVD processes using TEOS and ozone (O_3) have been proposed,⁹ as shown in Figure 8.10. This CVD technology produces oxide films with high conformality and low viscosity under low deposition temperature. The shrinkage of oxide film during annealing is also a function of ozone concentration, as shown in Figure 8.11. Because of their porosity, O_3 –TEOS CVD oxides are often accompanied by plasma-assisted oxides to permit planarization in ULSI processing.

For high-temperature deposition (900°C), silicon dioxide is formed by reacting dichlorosilane, SiCl_2H_2 , with nitrous oxide at reduced pressure:



This deposition gives excellent film uniformity and is sometimes used to deposit insulating layers over polysilicon.

Properties of Silicon Dioxide

Table 8.1 lists deposition methods and properties of silicon dioxide films.⁸ In general, there is a direct correlation between deposition temperature and film quality. At higher

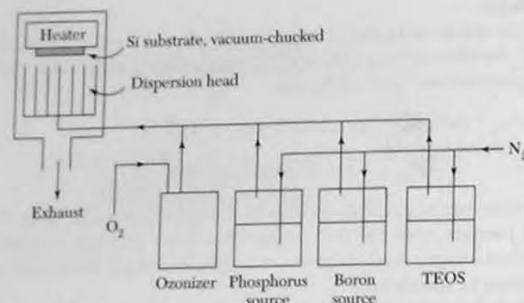


Figure 8.10 Experimental apparatus for the O_3 -TEOS chemical vapor deposition system.

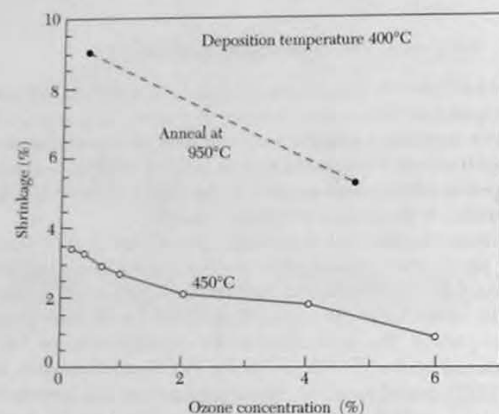


Figure 8.11 Dependence of the shrinkage of the O_3 -TEOS CVD film on ozone concentration using annealing. (Courtesy of SAMCO Company, Japan.)

TABLE 8.1 Properties of Silicon Dioxide Films

Property	Thermally Grown at 1000°C	$SiH_4 + O_2$ at 450°C	TEOS at 700°C	$SiCl_2H_2 + N_2O$ at 900°C
Composition	SiO_2	$SiO_2 (H)$	SiO_2	$SiO_2(Cl)$
Density (g/cm^3)	2.2	2.1	2.2	2.2
Refractive index	1.46	1.44	1.46	1.46
Dielectric strength ($10^6 V/cm$)	>10	8	10	10
Etch rate ($\text{\AA}/min$) (100:1 $H_2O:HF$)	30	60	30	30
Etch rate ($\text{\AA}/min$) (buffered HF)	440	1200	450	450
Step coverage	—	Nonconformal	Conformal	Conformal

temperatures, deposited oxide films are structurally similar to silicon dioxide that has been thermally grown.

The lower densities occur in films deposited below 500°C. Heating deposited silicon dioxide at temperatures between 600°C and 1000°C causes densification, during which the oxide thickness decreases, whereas the density increases to 2.2 g/cm^3 . The refractive index of silicon dioxide is 1.46 at a wavelength of 0.6328 μm . Oxides with lower indices are porous, such as the oxide resulting from silane-oxygen deposition, which has a refractive index of 1.44. The porous nature of the oxide also is responsible for the lower dielectric strength, which is the applied electric field that will cause a high current to flow in the oxide film. The etch rates of oxides in a hydrofluoric acid solution depend on deposition temperature, annealing history, and dopant concentration. Usually, higher-quality oxides are etched at lower rates.

Step Coverage

Step coverage relates the surface topography of a deposited film to the various steps on the semiconductor substrate. In the illustration of ideal, or *conformal*, step coverage shown in Figure 8.12a, film thickness is uniform along all surfaces of the step. The uniformity of the film thickness, regardless of topography, is due to the rapid migration of reactants after adsorption on the step surfaces.¹⁰

Figure 8.12b shows an example of nonconformal step coverage, which results when the reactants adsorb and react without significant surface migration. In this instance, the

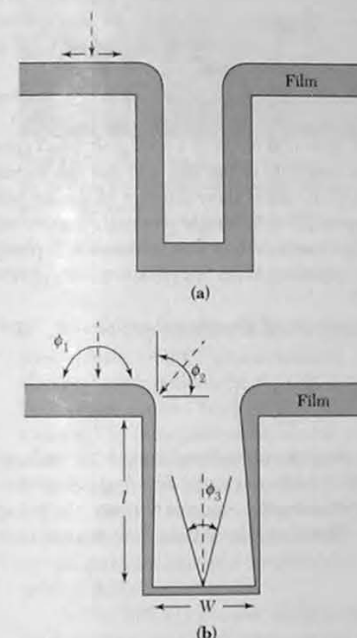


Figure 8.12 Step coverage of deposited films.¹⁰ (a) Conformal step coverage. (b) Nonconformal step coverage.

deposition rate is proportional to the arrival angle of the gas molecules. Reactants arriving along the top horizontal surface come from many different angles, and ϕ_1 , the arrival angle, varies in two dimensions, from 0° to 180° , whereas reactants arriving at the top of a vertical wall have an arrival angle (ϕ_2) that varies from 0° to 90° . Thus, the film thickness on the top surface is double that of a wall surface. Further down the wall, ϕ_3 is related to the width of the opening, and the film thickness is proportional to

$$\phi_3 \equiv \arctan \frac{W}{l} \quad (15)$$

where l is the distance from the top surface and W is the width of the opening. This type of step coverage is thin along the vertical walls, with a possible crack at the bottom of the step caused by self-shadowing.

Silicon dioxide formed by TEOS decomposition at reduced pressure gives a nearly conformal coverage due to rapid surface migration. Similarly, the high-temperature dichlorosilane-nitrous oxide reaction also results in conformal coverage. However, during silane-oxygen deposition, no surface migration takes place, and the step coverage is determined by the arrival angle. Most evaporated or sputtered materials have a step coverage similar to that in Figure 8.12b.

P-Glass Flow

A smooth topography is usually required for the deposited silicon dioxide used as an insulator between metal layers. If the oxide used to cover the lower metal layer is concave, circuit failure may result from an opening that may occur in the upper metal layer during deposition. Because phosphorus-doped silicon dioxide (P-glass) deposited at low temperatures becomes soft and flows upon heating, it provides a smooth surface and is often used to insulate adjacent metal layers. This process is called *P-glass flow*.

Figure 8.13 shows four cross sections of scanning electron microscope photographs of P-glass covering a polysilicon step.¹⁰ All samples are heated in steam at 1100°C for 20 minutes. Figure 8.13a shows a sample of glass that contains a negligibly small amount of phosphorus and does not flow. Note the concavity of the film and that the corresponding angle (θ) is about 120° . Figures 8.13b, c, and d show samples of P-glass with progressively higher phosphorus contents, up to 7.2 wt% (weight percent). In these samples, the decreasing step angles of the P-glass layer indicate how flow increases with phosphorus concentration. P-glass flow depends on annealing time, temperature, phosphorus concentration, and the annealing ambient.¹⁰

The angle θ as a function of weight percent of phosphorus as shown in Figure 8.13 can be approximated by

$$\theta \equiv 120^\circ \left(\frac{10 - \text{wt}\%}{10} \right) \quad (16)$$

If we want an angle smaller than 45° , we require a phosphorus concentration larger than 6 wt%. However, at concentrations above 8 wt%, the metal film (e.g., aluminum) may be corroded by the acid products formed during the reaction between the phosphorus in the oxide and atmospheric moisture. Therefore, the P-glass flow process uses phosphorus concentrations of 6 to 8 wt%.

8.3.2 Silicon Nitride

It is difficult to grow silicon nitride by thermal nitridation (e.g., with ammonia, NH_3) because of its slow growth rate and high growth temperature. However, silicon nitride

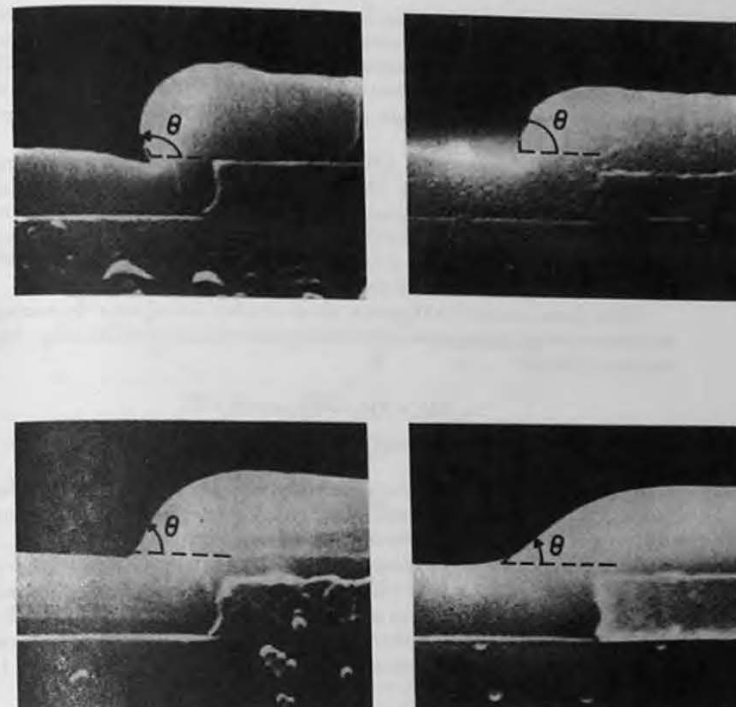
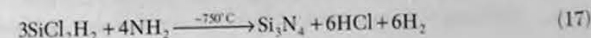


Figure 8.13 Scanning electron micrographs ($10,000\times$) of samples annealed in steam at 1100°C for 20 minutes for the following weight percent of phosphorus: (a) 0 wt%, (b) 2.2 wt%, (c) 4.6 wt%, and (d) 7.2 wt%.

films can be deposited by an intermediate-temperature (750°C) LPCVD process or a low-temperature (300°C) plasma-assisted CVD process.^{11,12} The LPCVD films are of stoichiometric composition (Si_3N_4) with high density ($2.9\text{--}3.1\text{ g/cm}^3$). These films can be used to passivate devices because they serve as good barriers to the diffusion of water and sodium. The films also can be used as masks for the selective oxidation of silicon because silicon nitride oxidizes very slowly and prevents the underlying silicon from oxidizing. The films deposited by plasma-assisted CVD are not stoichiometric and have a lower density ($2.4\text{--}2.8\text{ g/cm}^3$). Because of the low deposition temperature, silicon nitride films can be deposited over fabricated devices and serve as their final passivation. Plasma-deposited nitride provides excellent scratch protection, serves as a moisture barrier, and prevents sodium diffusion.

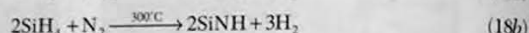
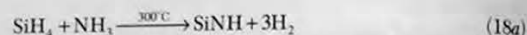
In the LPCVD process, dichlorosilane and ammonia react at reduced pressure to deposit silicon nitride at temperatures between 700°C and 800°C . The reaction is



Good film uniformity and high wafer throughput (i.e., wafers processed per hour) are advantages of the reduced-pressure process. As in the case of oxide deposition, silicon nitride deposition is controlled by temperature, pressure, and reactant concentration. The activation energy for deposition is about 1.8 eV. The deposition rate increases with increasing total pressure or dichlorosilane partial pressure and decreases with an increasing ammonia-to-dichlorosilane ratio.

Silicon nitride deposited by LPCVD is an amorphous dielectric containing up to 8 atomic percent (at%) hydrogen. The etch rate in buffered HF is less than 1 nm/min. The film has a very high tensile stress of approximately 10^{10} dynes/cm², which is nearly 10 times that of TEOS-deposited SiO₂. Films thicker than 200 nm may crack because of the very high stress. The resistivity of silicon nitride at room temperature is about 10^{16} Ω-cm. Its dielectric constant is 7, and its dielectric strength is 10^7 V/cm.

In the plasma-assisted CVD process, silicon nitride is formed either by reacting silane and ammonia in an argon plasma or by reacting silane in a nitrogen discharge. The reactions are as follows:



The products depend strongly on deposition conditions. The radial-flow, parallel-plate reactor (Fig. 8.9b) is used to deposit the films. The deposition rate generally increases with increasing temperature, power input, and reactant gas pressure.

Large concentrations of hydrogen are contained in plasma-deposited films. The plasma nitride (also referred to as SiN) used in semiconductor processing generally contains 20 to 25 at% hydrogen. Films with low tensile stress ($\sim 2 \times 10^9$ dynes/cm²) can be prepared by plasma deposition. Film resistivities range from 10^5 to 10^{21} Ω-cm, depending on silicon-to-nitrogen ratio, whereas dielectric strengths are between 1×10^6 and 6×10^6 V/cm.

8.3.3 Low-Dielectric-Constant Materials

As device sizes continue to shrink down to the deep submicron region, they require a multilevel interconnection architecture to minimize the time delay due to parasitic resistance (R) and capacitance (C). The gain in device speed at the gate level is offset by the propagation delay at the metal interconnects because of the increased RC time constant, as shown in Figure 8.14. For example, in devices with gate length of 250 nm or less, up to 50% of the time delay is due to the RC delay of long interconnections.¹³ Therefore, the device interconnection network becomes a limiting factor in determining chip performance metrics such as device speed, cross talk, and power consumption of ULSI circuits.

Reducing the RC time constant of ULSI circuits requires interconnection materials with low resistivity and interlayer films with low capacitance. Note that $C = \epsilon A/d$, where ϵ is the dielectric permittivity, A is the area, and d is the thickness of the dielectric film. Regarding the low-capacitance issue, it is not easy to lower the parasitic capacitance by increasing thickness of the interlayer dielectric (which makes gap filling more difficult) or decreasing wiring height and area (which results in the increase of interconnect resistance). Therefore, materials with low dielectric constant (low k) are required. The dielectric permittivity is equal to the product of k and ϵ_0 , where k and ϵ_0 are the dielectric constant and permittivity of free space, respectively.

The properties of the interlayer dielectric film and how it is formed have to meet the following requirements: low dielectric constant, low residual stress, high planarization

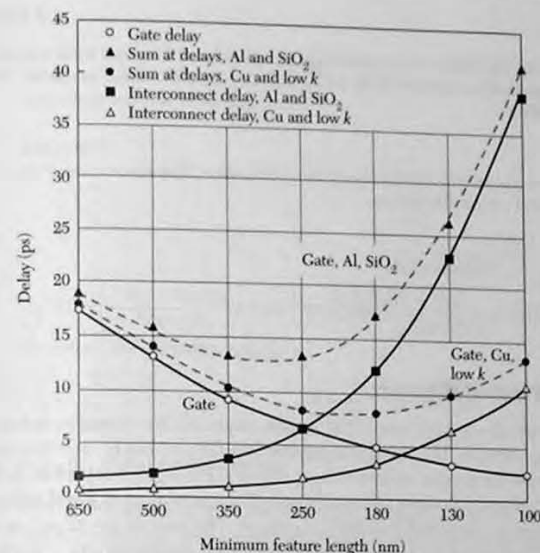


Figure 8.14 Calculated gate and interconnect delay versus technology generation. The dielectric constant for the low- k material is 2.0. Both Al and Cu interconnects are 0.8 μm thick and 43 μm long.

capability, high capability for gap filling, low deposition temperature, simplicity of process, and ease of integration. A substantial number of low- k materials have been synthesized for the intermetal dielectric in ULSI circuits. Some of the promising low- k materials are shown in Table 8.2. These materials can be either inorganic or organic and can be deposited by either CVD or spin-on techniques.¹³

TABLE 8.2 Low- k Materials

Determinant	Materials	Dielectric Constant
Vapor-phase deposition polymers	Fluorosilicate glass (FSG)	3.5–4.0
	Parylene N	2.6
	Parylene F	2.4–2.5
	Black diamond (C-doped oxide)	2.7–3.0
	Fluorinated hydrocarbon	2.0–2.4
	Teflon-AF	1.93
Spin-on polymers	HSQ/MSQ	2.8–3.0
	Polyimide	2.7–2.9
	SiLK (aromatic hydrocarbon polymer)	2.7
	PAE [poly(arylene ethers)]	2.6
	Fluorinated amorphous carbon	2.1
	Xerogels (porous silica)	1.1–2.0

EXAMPLE 3

Estimate the intrinsic RC value of two parallel Al wires $0.5 \mu\text{m} \times 0.5 \mu\text{m}$ in cross section, 1 mm in length, and separated by a polyimide ($k \sim 2.7$) dielectric layer that is $0.5 \mu\text{m}$ thick. The resistivity of Al is $2.7 \mu\Omega\text{-cm}$.

SOLUTION Let t_w be the cross-sectional dimension of the wires. Resistance is equal to ρl divided by the cross-sectional area of the wire, so

$$RC = \left(\rho \frac{l}{t_w} \right) \times \left(\epsilon \frac{A}{\text{Spacing width}} \right) \\ = \left(2.7 \times 10^{-6} \times \frac{10^{-1}}{0.25 \times 10^{-3}} \right) \times \left(8.85 \times 10^{-14} \times 2.7 \times \frac{0.5 \times 10^{-4} \times 10^{-1}}{0.5 \times 10^{-4}} \right) = 2.57 \text{ ps}$$

8.3.4 High-Dielectric-Constant Materials

High- k materials are also required for ULSI circuits, especially for dynamic random access memory (DRAM) circuits. The storage capacitor in a DRAM has to maintain a certain value of capacitance for proper operation (e.g., 40 fF). For a given capacitance, a minimum d is usually selected to meet the conditions of the maximum allowed leakage current and the minimum required breakdown voltage. The area of the capacitor can be increased by using stacked or trench structures. These structures are considered in Chapter 9. However, for a planar structure, area is reduced with increasing DRAM density. Therefore, the dielectric constant of the film must be increased.

Several high- k materials have been proposed, such as barium strontium titanate (BST) and lead zirconium titanate (PZT). These materials are shown in Table 8.3. In addition, there are titanates doped with one or more acceptors, such as alkaline earth metals, or doped with one or more donors, such as rare earth elements. Tantalum oxide (Ta_2O_5) has a dielectric constant in the range of 20 to 30. As a reference, the dielectric constant of Si_3N_4 is in the range of 6 to 7, and that for SiO_2 is 3.9. A Ta_2O_5 film can be deposited by a CVD process using gaseous TaCl_5 and O_2 as the starting materials.

TABLE 8.3 High- k Materials

	Materials	Dielectric Constant
Binary	Ta_2O_5	25
	TiO_2	40
	Y_2O_3	17
	Si_3N_4	7
Peraelectric perovskite	SrTiO_3 (STO)	140
	$(\text{Ba}_{1-x}\text{Sr}_x)\text{TiO}_3$ (BST)	300–500
	$\text{Ba}(\text{Ti}_{1-x}\text{Zr}_x)\text{O}_3$ (BZT)	300
	$(\text{Pb}_{1-x}\text{La}_x)(\text{Zr}_{1-y}\text{Ti}_y)\text{O}_3$ (PLZT)	800–1000
	$\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (PMN)	1000–2000
Ferroelectric perovskite	$\text{Pb}(\text{Zr}_{0.47}\text{Ti}_{0.53})\text{O}_3$ (PZT)	>1000

EXAMPLE 4

A DRAM capacitor has the following parameters: $C = 40$ fF, cell size (A) = $1.28 \mu\text{m}^2$, and $k = 3.9$ for silicon dioxide. If we replace SiO_2 with Ta_2O_5 ($k = 25$) without changing thickness, what is the equivalent cell area of the capacitor?

SOLUTION

$$C = \frac{\epsilon_r A}{d}$$

so

$$\frac{3.9 \times 1.28}{d} = \frac{25 \times A}{d}$$

Therefore, the equivalent cell size is

$$A = \frac{3.9}{25} \times 1.28 = 0.2 \mu\text{m}^2$$

8.4 POLYSILICON DEPOSITION

Using polysilicon as the gate electrode in MOS devices is a significant development in MOS technology. One important reason is that polysilicon surpasses aluminum for electrode reliability. Figure 8.15 shows the maximum time to breakdown for capacitors with both polysilicon and aluminum electrodes.¹⁴ Polysilicon is clearly superior, especially for

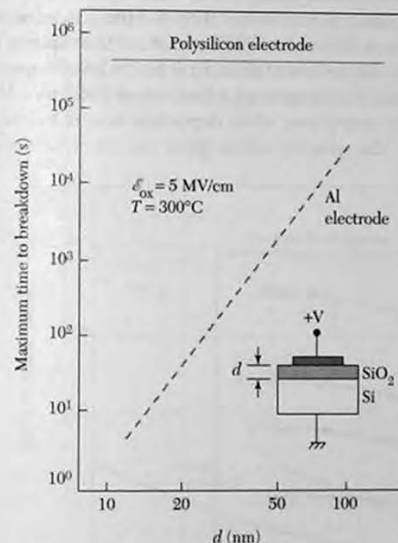
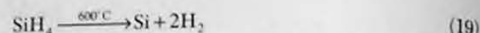


Figure 8.15 Maximum time to breakdown versus oxide thickness for a polysilicon electrode and an aluminum electrode.¹⁴

thinner gate oxides. The inferior time to breakdown of aluminum electrodes is due to the migration of aluminum atoms into the thin oxide under an electrical field. Polysilicon is also used as a diffusion source to create shallow junctions and to ensure ohmic contact to crystalline silicon. Additional uses include the manufacture of conductors and high-value resistors.

A low-pressure reactor (Fig. 8.9a) operated between 600°C and 650°C is used to deposit polysilicon by pyrolyzing silane according to the reaction



Of the two most common low-pressure processes, one operates at a pressure of 25 to 130 Pa using 100% silane, whereas the other involves a diluted mixture of 20% to 30% silane in nitrogen at the same total pressure. Both processes can deposit polysilicon on hundreds of wafers per run with good uniformity (i.e., thicknesses within $\pm 5\%$).

Figure 8.16 shows the deposition rate at four deposition temperatures. At low silane partial pressure, the deposition rate is proportional to the silane pressure.⁸ At higher silane concentrations, saturation of the deposition rate occurs. Deposition at reduced pressure is generally limited to temperatures between 600°C and 650°C. In this temperature range, the deposition rate varies as $\exp(-E_a/kT)$, where E_a is 1.7 eV, which is essentially independent of the total pressure in the reactor. At higher temperatures, gas-phase reactions that result in a rough, loosely adhering deposit become significant, and silane depletion occurs, causing poor uniformity. At temperatures much lower than 600°C, the deposition rate is too slow to be practical.

Process parameters that affect the polysilicon structure are deposition temperature, dopants, and the heat cycle applied following the deposition step. A columnar structure results when polysilicon is deposited at a temperature of 600°C to 650°C. This structure consists of polycrystalline grains ranging in size from 0.03 to 0.3 μm at a preferred orientation of (110). When phosphorus is diffused at 950°C, the structure changes to crystallite, and grain size increases to a size between 0.5 and 1.0 μm . When temperature is increased to 1050°C during oxidation, the grains reach a final size of 1 to 3 μm . Although the initially deposited film appears amorphous when deposition occurs below 600°C, growth characteristics similar to the polycrystalline-grain columnar structure are observed after doping and heating.

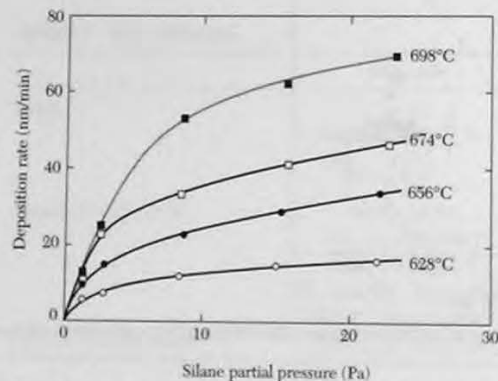


Figure 8.16 Effect of silane concentration on the polysilicon deposition rate.⁸

Polysilicon can be doped by diffusion, ion implantation, or the addition of dopant gases during deposition, referred to as *in situ doping*. The implantation method is most commonly used because of its lower processing temperatures. Figure 8.17 shows the sheet resistance of single-crystal silicon and of 500-nm polysilicon doped with phosphorus and antimony using ion implantation.¹⁵ The ion implantation process was considered in Chapter 7. Implant dose, annealing temperature, and annealing time all influence the sheet resistance of implanted polysilicon. Carrier traps at the grain boundaries cause a very high resistance in the lightly implanted polysilicon. As Figure 8.17 illustrates, resistance drops rapidly, approaching that of implanted single-crystal silicon, as the carrier traps become saturated with dopants.

► 8.5 METALLIZATION

8.5.1 Physical Vapor Deposition

The most common methods of physical vapor deposition (PVD) of metals are evaporation, e-beam evaporation, plasma spray deposition, and sputtering. Metals and metal compounds such as Ti, Al, Cu, TiN, and TaN can be deposited by PVD. Evaporation occurs when a source material is heated above its melting point in an evacuated chamber. The evaporated atoms then travel at high velocity in straight-line trajectories. The source can be melted by resistance heating, by rf heating, or with a focused electron beam. Evaporation and e-beam evaporation were used extensively in earlier generations of integrated circuits, but they have been replaced by sputtering for ULSI circuits.

In ion beam sputtering, a source of ions is accelerated toward the target and impinges on its surface. Figure 8.18a shows a standard sputtering system. The sputtered material deposits on a wafer that is placed facing the target. The ion current and energy can be independently adjusted. Since the target and wafer are placed in a chamber that has lower pressure, more target material and less contamination are transferred to the wafer.

One method to increase the ion density and, hence, the sputter-deposition rate is to use a third electrode that provides more electrons for ionization. Another method is to use

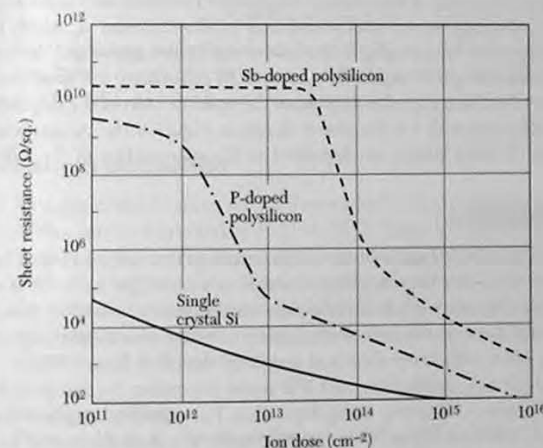


Figure 8.17 Sheet resistance versus ion dose into 500-nm polysilicon at 30 keV.¹⁵

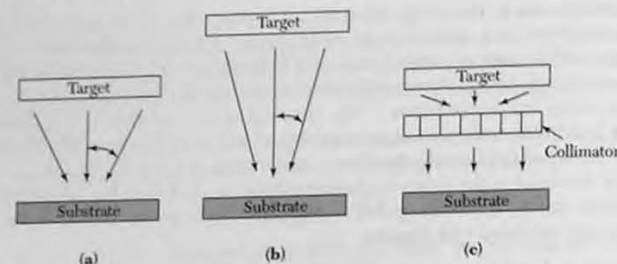


Figure 8.18 (a) Standard sputtering. (b) Long-throw sputtering. (c) Sputtering with a collimator.

a magnetic field, such as *electron cyclotron resonance* (ECR), to capture and spiral electrons, increasing their ionizing efficiency in the vicinity of the sputtering target. This technique, referred to as *magnetron sputtering*, has found widespread applications for the deposition of aluminum and its alloys at a rate that can approach 1 $\mu\text{m}/\text{min}$.

Long-throw sputtering is another technique used to control the angular distribution. Figure 8.18b shows a long-throw sputtering system. In standard sputtering configurations, there are two primary reasons for a wide angular distribution of incident flux at the surface: the use of a small target-to-substrate separation, d_t ; and scattering of the flux by the working gas as the flux travels from the target to the substrate. These two factors are linked because a small d_t is needed to achieve good throughput, uniformity, and film properties when there is substantial gas scattering. A solution to this problem is to sputter at very low pressures, a capability that has been developed using a variety of systems that can sustain the magnetron plasma under more rarefied conditions. These systems allow for sputtering at working pressures of less than 0.1 Pa. At these pressures, gas scattering is less important, and the target-to-substrate distance can be greatly increased. From a simple geometrical argument, this allows the angular distribution to be greatly narrowed, which permits more deposition at the bottom of high-aspect features such as contact holes.

Contact holes with large aspect ratio are difficult to fill with material, mainly because scattering events cause the top opening of the hole to seal before appreciable material has deposited on its floor. This problem can be overcome by collimating the sputtered atoms by placing an array of collimating tubes just above the wafer to restrict the depositing flux to normal $\pm 5^\circ$. Sputtering with a collimator is shown in Figure 8.18c. Atoms whose trajectory is more than 5° from normal are deposited on the inner surface of the collimators.

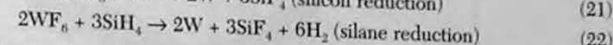
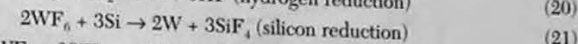
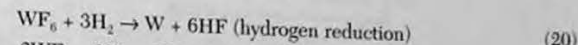
8.5.2 Chemical Vapor Deposition

CVD is attractive for metallization because it offers coatings that are conformal, has good step coverage, and can coat a large number of wafers at a time. The basic CVD setup is the same as that used for deposition of dielectrics and polysilicon (see Fig. 8.9a). Low-pressure CVD is capable of producing conformal step coverage over a wide range of topographical profiles, often with lower electrical resistivity than that from PVD.

One of the major new applications of CVD metal deposition for integrated circuit production is in the area of refractory metal deposition. For example, tungsten's low electrical resistivity (5.3 $\mu\Omega\text{-cm}$) and refractory nature make it a desirable metal for use in integrated circuit fabrication.

CVD Tungsten

Tungsten is used both as a contact plug and as a first-level metal. Tungsten can be deposited by using WF_6 as the source gas, since it is a liquid that boils at room temperature. WF_6 can be reduced by silicon, hydrogen, or silane. The basic chemistry for CVD W is as follows:

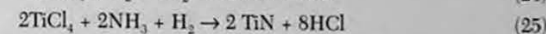
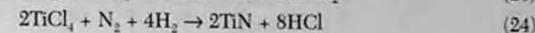
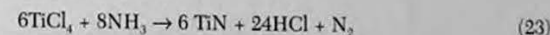


On a Si contact, the selective process starts from a silicon reduction process. This process provides a nucleation layer of W grown on Si but not on SiO_2 . The hydrogen reduction process can deposit W rapidly on the nucleation layer, forming the plug. The hydrogen reduction process provides excellent conformal coverage of the topography. This process, however, does not have perfect selectivity, and the HF gas by-product of the reaction is responsible for the encroachment of the oxide, as well as for the rough surface of deposited W films.

The silane reduction process gives a high deposition rate and much smaller W grain size than that obtained with the hydrogen reduction process. In addition, the problems of encroachment and a rough W surface are eliminated because there is no HF by-product generation. Usually, a silane reduction process is used as the first step in blanket W deposition to serve as a nucleation layer and to reduce junction damage. After the silane reduction, hydrogen reduction is used to grow the blanket W layer.

CVD TiN

TiN is widely used as a diffusion barrier metal layer in metallization and can be deposited by sputtering from a compound target or by CVD. CVD TiN can provide better step coverage than PVD methods in deep submicron technology. CVD TiN can be deposited¹⁶⁻¹⁸ using TiCl_4 with NH_3 , H_2/N_2 , or NH_3/H_2 :



The deposition temperature is about 400°C to 700°C for NH_3 reduction and is higher than 700°C for the N_2/H_2 reaction. The higher the deposition temperature, the better the TiN film, and the less Cl incorporated in TiN (~5%).

8.5.3 Aluminum Metallization

Aluminum and its alloys are used extensively for metallization in integrated circuits. The Al film can be deposited by PVD or CVD. Since aluminum and its alloys have low resistivities (2.7 $\mu\Omega\text{-cm}$ for Al and up to 3.5 $\mu\Omega\text{-cm}$ for its alloys), these metals satisfy the low-resistance requirements. Aluminum also adheres well to silicon dioxide. However, the use of aluminum in integrated circuits with shallow junctions often creates problems, such as spiking and electromigration. This section considers the problems of aluminum metallization and their solutions.

Junction Spiking

Figure 8.19 shows the phase diagram of the Al-Si system at 1 atm.¹⁹ The phase diagram relates these two components as a function of temperature. The Al-Si system exhibits

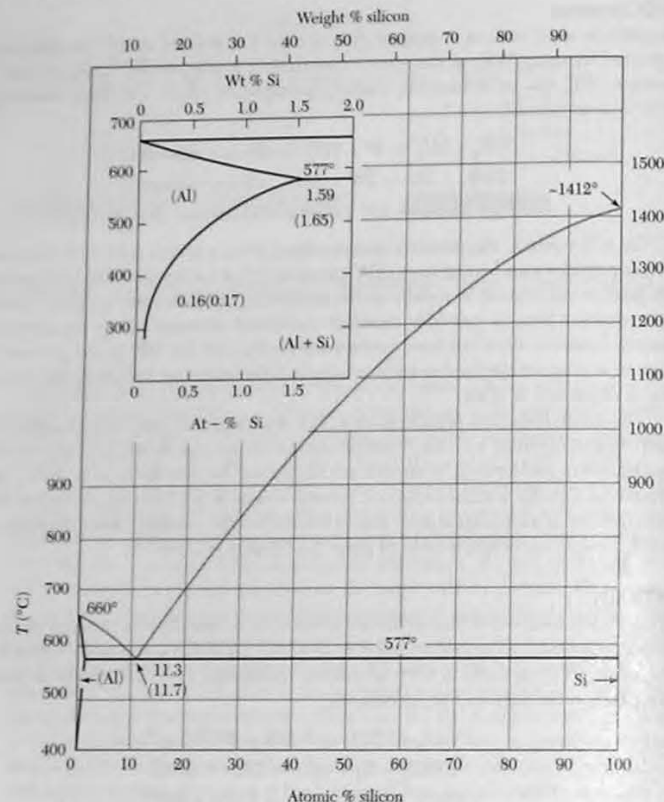


Figure 8.19 Phase diagram of the aluminum-silicon system.¹⁹

eutectic characteristics; that is, the addition of either component lowers the system's melting point below that of either metal. Here, the minimum melting temperature, called *eutectic temperature*, is 577°C, corresponding to a 11.3% Si and 88.7% Al composition. The melting points of pure aluminum and pure silicon are 660°C and 1412°C, respectively. Because of the eutectic characteristics, during aluminum deposition the temperature on the silicon substrate must be limited to less than 577°C.

The inset of Figure 8.19 shows the solid solubility of silicon in aluminum. For example, the solubility of silicon in aluminum is 0.25 wt% at 400°C, 0.5 wt% at 450°C, and 0.8 wt% at 500°C. Therefore, wherever aluminum contacts silicon, the silicon will dissolve into the aluminum during annealing. The amount of silicon dissolved will depend not only on the solubility at the annealing temperature, but also on the volume of aluminum to be saturated with silicon. Consider a long aluminum metal line in contact with an area ZL of silicon, as shown in Figure 8.20. After an annealing time t , the silicon will diffuse a distance of approximately $\sqrt{Dt}$ along the aluminum line from the edge of the contact, where D is the diffusion coefficient given by $4 \times 10^{-2} \exp(-0.92/kT)$ for silicon

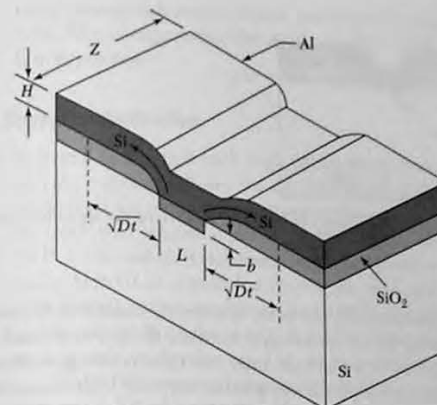


Figure 8.20 Diffusion of silicon in aluminum metallization.²⁰

diffusion in deposited aluminum films. Assuming that this length of aluminum is completely saturated with silicon, the volume of silicon consumed is then

$$\text{Vol} \cong 2\sqrt{Dt}(HZ)S\left(\frac{\rho_{\text{Al}}}{\rho_{\text{Si}}}\right) \quad (26)$$

where ρ_{Al} and ρ_{Si} are the densities of aluminum and silicon, respectively, and S is the solubility of silicon in aluminum at the annealing temperature.²⁰ If the consumption takes place uniformly over the contact area A (where $A = ZL$ for uniform dissolution), the depth to which silicon would be consumed is

$$b \cong 2\sqrt{Dt}\left(\frac{HZ}{A}\right)S\left(\frac{\rho_{\text{Al}}}{\rho_{\text{Si}}}\right) \quad (27)$$

EXAMPLE 5

Let $T = 500^\circ\text{C}$, $t = 30$ min, $ZL = 16 \mu\text{m}^2$, $Z = 5 \mu\text{m}$, and $H = 1 \mu\text{m}$. Find the depth b , assuming uniform dissolution.

SOLUTION The diffusion coefficient of silicon in aluminum at 500°C is about $2 \times 10^{-8} \text{ cm}^2/\text{s}$; thus, $\sqrt{Dt}$ is 60 μm . The density ratio is $2.7/2.33 = 1.16$. At 500°C, S is 0.8 wt%. From Eq. 27, we have

$$b = 2 \times 60 \left(\frac{1 \times 5}{16} \right) 0.8\% \times 1.16 = 0.35 \mu\text{m}$$

Aluminum will fill a depth of $b = 0.35 \mu\text{m}$ from which silicon is consumed. If at the contact point there is a shallow junction whose depth is less than b , the diffusion of silicon into aluminum can short-circuit the junction. ◀

In a practical situation, the dissolution of silicon does not take place uniformly, but rather at only a few points. The effective area in Eq. 27 is less than the actual contact area; hence, b is much larger. Figure 8.21 illustrates the actual situation in the p - n junction area of aluminum penetrating the silicon at only the few points where spikes are formed. One way to minimize aluminum spiking is to add silicon to the aluminum by co-evaporation

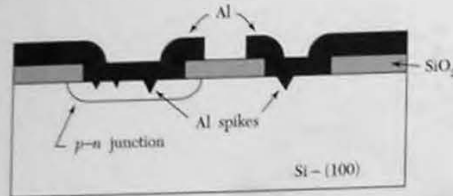


Figure 8.21 Schematic view of aluminum films contacting silicon. Note the aluminum spiking in the silicon.

until the amount of silicon contained by the alloy satisfies the solubility requirement. Another method is to introduce a barrier metal layer between the aluminum and the silicon substrate (Fig. 8.22). This barrier metal layer must meet the following requirements: It must form low contact resistance with silicon, it must not react with aluminum, and its deposition and formation must be compatible with the overall process. Barrier metals such as titanium nitride (TiN) have been evaluated and found to be stable for contact annealing temperatures of up to 550°C for 30 minutes.

Electromigration

As devices become smaller, the corresponding current density becomes larger. High current densities can cause device failure due to electromigration. The term *electromigration* refers to the transport of mass (i.e., atoms) in metals under the influence of current. It occurs by the transfer of momentum from the electrons to the positive metal ions. When a high current passes through thin metal conductors in integrated circuits, metal ions in some regions will pile up, and voids will form in other regions. This pileup can short-circuit adjacent conductors, whereas the voids can result in an open circuit.

The mean time to failure (MTF) of a conductor due to electromigration can be related to the current density (J) and the activation energy by

$$\text{MTF} \sim \frac{1}{J^2} \exp\left(\frac{E_a}{kT}\right) \quad (28)$$

Experimentally, a value of $E_a \approx 0.5$ eV is obtained for deposited aluminum. This indicates that low-temperature grain-boundary diffusion is the primary vehicle of material transport, since an $E_a \approx 1.4$ eV would characterize the self-diffusion of single-crystal aluminum. The electromigration resistance of aluminum conductors can be increased by

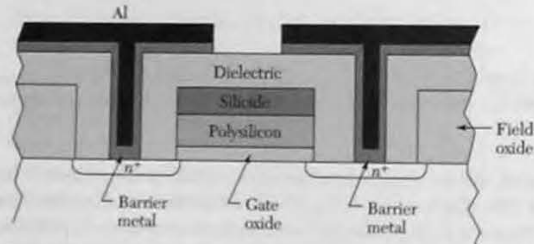


Figure 8.22 Cross-sectional view of a MOSFET with a barrier metal between the aluminum and silicon and a composite gate electrode of silicide and polysilicon.

using several techniques. These techniques include alloying with copper (e.g., Al with 0.5% Cu), encapsulating the conductor in a dielectric, or incorporating oxygen during film deposition.

8.5.4 Copper Metallization

It is well known that both high-conductivity wiring and low-dielectric-constant insulators are required to lower the RC time delay of the interconnect network. Copper is the obvious choice for a new interconnection metallization because it has higher conductivity and higher electromigration resistance than aluminum. Copper can be deposited by PVD, CVD, and electrochemical methods. However, the use of Cu as an alternative material to Al in ULSI circuits has drawbacks, such as its tendency to corrode under standard chip manufacturing conditions, its lack of a feasible dry etching method or a stable self-passivating oxide similar to Al_2O_3 on Al, and its poor adhesion to dielectric materials, such as SiO_2 and low- k polymers. This section discusses copper metallization techniques.

Several different techniques for fabrication of multilevel Cu interconnects have been reported.^{21, 22} The first method is a conventional method to pattern the metal lines, followed by dielectric deposition. The second method is to pattern the dielectric layer first and fill copper metal into trenches. This step is followed by *chemical mechanical polishing*, discussed later, to remove the excess metal on the top surface of the dielectric and leave Cu material in the holes and trenches. This method is also known as a *damascene* process.

Damascene Technology

The approach for fabricating a copper/low- k dielectric interconnect structure is by the damascene or dual damascene process. Figure 8.23 shows the dual damascene sequence for an advanced Cu interconnection structure. For a typical damascene structure, trenches for metal lines are defined and etched in the interlayer dielectric (ILD), followed by metal deposition of TaN/Cu. The TaN layer serves as a diffusion barrier layer and prevents copper from penetrating the low- k dielectric. The excess copper metal on the surface is removed to obtain a planar structure with metal inlays in the dielectric.

For the dual damascene process, the vias and trenches in the dielectric are defined using two lithography and reactive ion etching (RIE) steps before depositing the Cu metal (Figs. 8.23a–c). Then a Cu chemical mechanical polishing process is used to remove the metal on the top surface, leaving the planarized wiring and via imbedded in the insulator.²³ One special benefit of the dual damascene technique is that the via plug is now of the same material as the metal line and the risk of via electromigration failure is reduced.

EXAMPLE 6

If we replace Al with Cu wire associated with some low- k dielectric ($k = 2.6$) instead of a SiO_2 layer, what percentage of reduction in the RC time constant will be achieved? The resistivity of Al is $2.7 \mu\Omega\text{-cm}$, and the resistivity of Cu is $1.7 \mu\Omega\text{-cm}$.

SOLUTION

$$\frac{1.7}{2.7} \times \frac{2.6}{3.9} \times 100\% = 42\%$$

Chemical Mechanical Polishing

In recent years, the development of chemical mechanical polishing (CMP) has become increasingly important for multilevel interconnection because it is the only technology

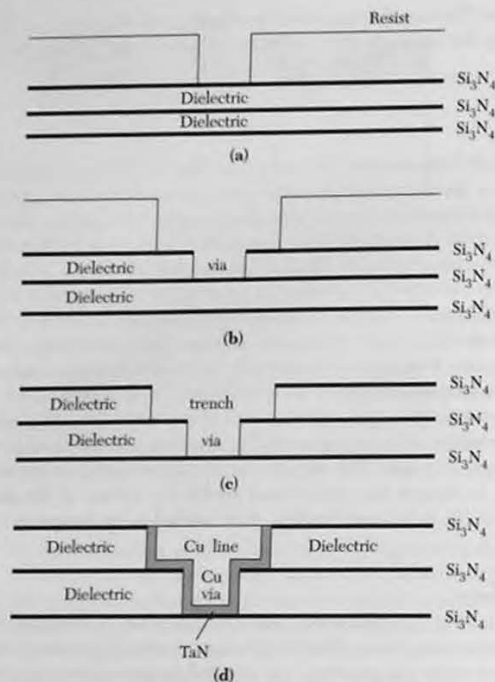


Figure 8.23 Process sequence used to fabricate a Cu line-stud structure using dual damascene. (a) Resist stencil applied. (b) Reactive ion etching dielectric and resist patterning. (c) Trench and via definition. (d) Cu depositions followed by chemical mechanical polishing.

that allows global planarization (i.e., makes a flat surface across the whole wafer). It offers many advantages over other types of technologies, including better global planarization over large or small structures, reduced defect density, and the avoidance of plasma damage. Three CMP approaches are summarized in Table 8.4.

The CMP process consists of moving the sample surface against a pad that carries slurry between the sample surface and the pad. Abrasive particles in the slurry cause mechanical damage on the sample surface, loosening the material for enhanced chemical attack or fracturing off the pieces of surface into a slurry where they dissolve or are swept away. The process is tailored to provide an enhanced material removal rate from high points on surfaces, thus affecting the planarization because most chemical actions are isotropic. Mechanical grinding alone may theoretically achieve the desired planarization,

TABLE 8.4 Three Methods of Chemical Mechanical Polishing (CMP)

Method	Wafer Facing	Platen Movement	Slurry Feeding
Rotary CMP	Down	Rotary against rotating wafer carrier	Dripping to pad surface
Orbital CMP	Down	Orbital against rotating wafer carrier	Through the pad surface
Linear CMP	Down	Linear against rotating wafer carrier	Dripping to pad surface

but is not desirable because of extensive associated damage to the material surfaces. There are three main parts of the process: (a) the surface to be polished; (b) the pad, which is the key medium enabling the transfer of mechanical action to the surface being polished; and (c) the slurry, which provides both chemical and mechanical effects. Figure 8.24 shows the CMP setup.²⁴

EXAMPLE 7

The oxide removal rate and the removal rate of a layer underneath the oxide (called a stop layer) are $1r$ and $0.1r$, respectively. To remove $1\text{ }\mu\text{m}$ of oxide and a $0.01\text{ }\mu\text{m}$ stop layer, the total removal time is 5.5 minutes. Find the oxide removal rate.

SOLUTION

$$\frac{1}{1r} + \frac{0.01}{0.1r} = 5.5$$

$$r = 0.2\text{ }\mu\text{m/min}$$

8.5.5 Silicide

Silicon forms many stable metallic and semiconducting compounds with metals. Several metal silicides show low resistivity and high thermal stability, making them suitable for ULSI application. Silicides such as TiSi₂ and CoSi₂ have reasonably low resistivities and are generally compatible with integrated circuit processing. Silicides become important metallization materials as devices become smaller. One important application of silicide is for the MOSFET gate electrode, either alone or with doped polysilicon (polycide) above the gate oxide. Table 8.5 shows a comparison of titanium silicide and cobalt silicide.

Metal silicides have been used to reduce the contact resistance of the source and drain, the gate electrodes, and interconnections. Self-aligned metal silicide technology

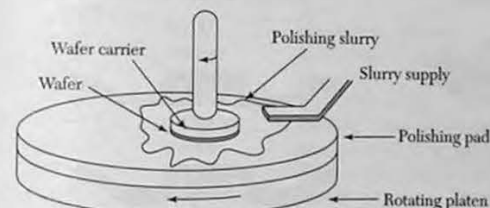


Figure 8.24 Schematic of a CMP polisher.

TABLE 8.5 A Comparison of TiSi₂ and CoSi₂ Films

Properties	TiSi ₂	CoSi ₂
Resistivity	13–16	22–28
Silicide/metal ratio	2.37	3.56
Silicide/Si ratio	1.04	0.97
Reactive to native oxide	Yes	No
Silicidation temperature (°C)	800–850	550–900
Film stress (dyne/cm ²)	1.5×10^{10}	1.2×10^{10}

(salicide) has been proven to be a highly attractive technique for improving the performance of submicron devices and circuits. The self-aligned process uses the silicide gate electrode as the mask to form the source and drain electrodes of a MOSFET (e.g., by ion implantation, considered in Chapter 7). This process can minimize the overlap of these electrodes and thus reduce parasitic capacitances.

Figure 8.25 shows the polycide and salicide processes. A typical polycide formation sequence is shown in Figure 8.25a. For sputter deposition, a high-temperature, high-purity compound target is used to ensure the quality of the silicide. The most commonly used silicides for the polycide process are WSi_2 , TaSi_2 , and MoSi_2 . They are all refractory, thermally stable, and resistant to processing chemicals. A self-aligned silicide process is illustrated in Figure 8.25b. In the process, the polysilicon gate is patterned without any silicide, and a sidewall spacer (silicon oxide or silicon nitride) is formed to prevent shorting the gate to the source and drain during the silicidation process. A metal layer, either Ti or Co, is blanket-sputtered on the entire structure, followed by silicide sintering. Silicide is formed, in principle, only where the metal is in contact with Si. A wet chemical wash then rinses off the unreacted metal, leaving only the silicide. This technique eliminates the need to pattern the composite polycide gate structure and adds silicide to the source/drain area to reduce the contact resistance.

Silicides are promising materials for ULSI circuits because of their low resistivity and excellent thermal stability. Cobalt silicide has been widely investigated because of its low resistivity and high-temperature thermal stability. However, cobalt is sensitive to native oxides as well as an oxygen-containing environment, and a large amount of silicon is consumed during silicidation.

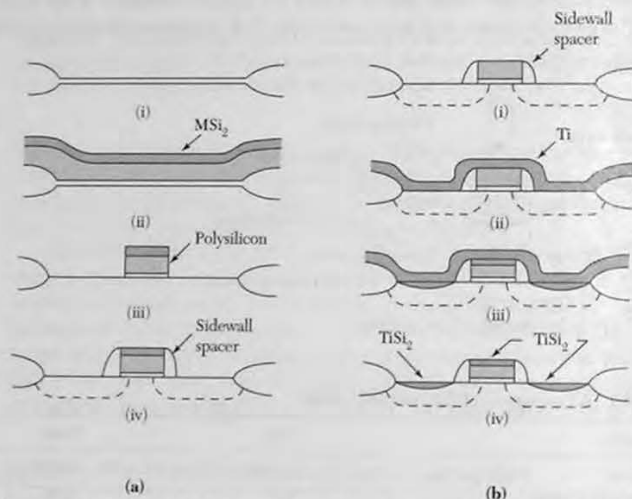


Figure 8.25 Polycide and salicide processes. (a) Polycide structure: (i) gate oxide, (ii) polysilicon and silicide deposition, (iii) pattern polycide, and (iv) lightly doped drain (LDD) implant, sidewall formation, and source/drain implant. (b) Salicide structure: (i) gate patterning (polysilicon only), LDD, sidewall, and source/drain implant; (ii) metal (Ti, Co) deposition; (iii) anneal to form salicide; and (iv) selective (wet) etch to remove unreacted metal.

EXAMPLE 8

Calculate the thickness of cobalt silicide for a desired sheet resistance of $0.6 \Omega/\square$. The resistivity is $18 \mu\Omega\text{-cm}$.

SOLUTION The resistivity is equal to the product of the sheet resistance and the film thickness:

$$\rho = R_s \times t$$

Then

$$t = \frac{\rho}{R_s} = \frac{18 \times 10^{-6}}{0.6} = 3 \times 10^{-5} \text{ cm} = 300 \text{ nm}$$

8.6 DEPOSITION SIMULATION

SUPREM may be used to simulate the deposition process. Like etch simulation, deposition modeling is very straightforward. Simulation is executed using the DEPOSITION command, which deposits a given amount of user-specified material on top of the current structure. The material deposited may be either undoped or uniformly doped. If single-crystal silicon is deposited, then the crystal orientation must also be specified. If polysilicon is deposited, the temperature must be specified for SUPREM to determine the proper polysilicon grain size.

EXAMPLE 9

Suppose we want to simulate the deposition of 800Å CVD silicon nitride on top of a dry oxide layer approximately 400Å thick. If the *p*-type silicon substrate is doped with boron at a level of 10^{15} cm^{-3} , use SUPREM to determine the final oxide and nitride layer thicknesses, as well as the boron doping profile in the oxide and nitride layers.

SOLUTION The SUPREM input listing is as follows:

TITLE	Deposition Example
COMMENT	Initialize silicon substrate
INITIALIZE	<100> Silicon Boron Concentration=1e15
COMMENT	Grow 400Å oxide
DIFFUSION	Time=40 Temperature=1000 DryO2
COMMENT	Deposit 800Å CVD nitride
DEPOSITION	Nitride Thickness=0.08
PRINT	Layers
PLOT	Chemical Boron Net
STOP	End Deposition Example

After simulation is complete, the results are shown in Figure 8.26, which indicates final oxide and nitride layer thicknesses of 379 and 800Å, respectively, and depicts the boron incorporation in the oxide layer.

8.7 SUMMARY

Modern semiconductor device fabrication requires the use of thin films. In the epitaxial growth process, the substrate wafer is the seed. High-quality, single-crystal films can be grown at a temperature 30% to 50% lower than the melting point. The common techniques for epitaxial growth are chemical vapor deposition, metalorganic CVD, and molecular beam

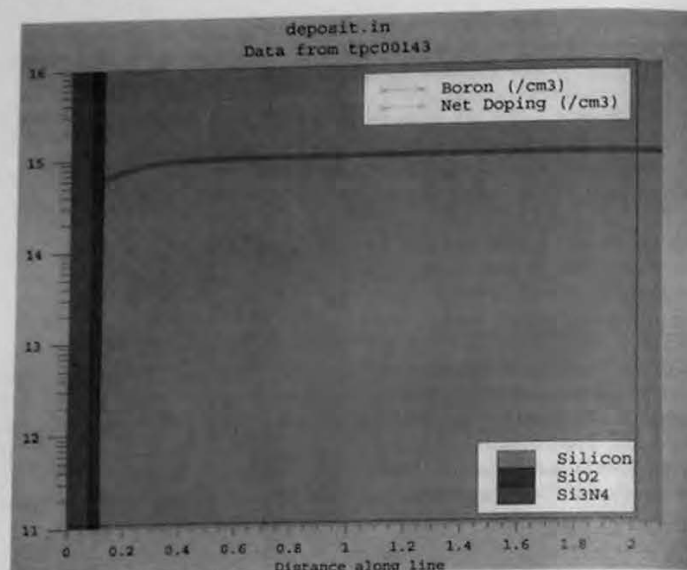


Figure 8.26 Plot of boron incorporation into the oxide layer, using SUPREM.

epitaxy. CVD and MOCVD are chemical deposition processes. Gases and dopants are transported in vapor form to the substrate, where a chemical reaction occurs that results in the deposition of the epitaxial layer. Inorganic compounds are used for CVD, whereas metalorganic compounds are used for MOCVD. MBE, on the other hand, is a physical deposition process. It is done by the evaporation of a species in an ultrahigh-vacuum system. Because it is a low-temperature process that has a low growth rate, MBE can grow single-crystal, multilayer structures with dimensions on the order of atomic layers.

In addition to conventional homoepitaxy, such as n -type silicon on an n -type silicon substrate, this chapter also considered heteroepitaxy, which includes lattice-matched and strained-layer epitaxy. For strained-layer epitaxy, there is a critical layer thickness above which edge dislocations will nucleate to relieve the strain energy.

Besides the edge dislocations in an epitaxial layer, defects from the substrate, defects from the interface, precipitates, and low-angle grain boundaries and twins exist. These defects degrade device performance. Various means have been presented to minimize, or even eliminate, these defects so that a defect-free semiconductor layer can be grown either homoepitaxially or heteroepitaxially.

Aside from epitaxial layers, there are four other important types of films: thermal oxides, dielectric layers, polycrystalline silicon, and metal films. The major issues related to film formation are low-temperature processing, step coverage, selective deposition, uniformity, film quality, planarization, throughput, and large wafer capacity.

Thermal oxidation offers the best quality for the Si-SiO₂ interface and has the lowest interface trap density (see Chapter 3). Therefore, it is used to form gate and field oxides. LPCVD of dielectrics and polysilicon offers conformal step coverage. In contrast, PVD and atmospheric-pressure CVD generally result in nonconformal step coverage. CMP offers global planarization and reduces defect density. Conformal step coverage and planarization are also required for precise pattern transfer at the deep-submicron lithography level.

To minimize the RC time delay due to parasitic resistance and capacitance, the silicide process for ohmic contacts, copper metallization for interconnects, and low-dielectric-constant materials for interlayer films are extensively used to meet the requirements of the multilevel interconnect structures of ULSI circuits. In addition, this chapter discussed high-dielectric-constant materials to improve the gate insulator performance and to increase the capacitance per unit area for DRAM.

► REFERENCES

1. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.
2. R. Reif, T. I. Kamins, and K. C. Saraswat, "A Model for Dopant Incorporation into Growing Silicon Epitaxial Films," *J. Electrochem. Soc.*, **126**, 644, 653 (1979).
3. R. D. Dupuis, "Metalorganic Chemical Vapor Deposition of III-V Semiconductors," *Science*, **226**, 623 (1984).
4. M. A. Herman and H. Sitter, *Molecular Beam Epitaxy*, Springer-Verlag, Berlin, 1996.
5. A. Roth, *Vacuum Technology*, North-Holland, Amsterdam, 1976.
6. M. Ohring, *The Materials Science of Thin Films*, Academic Press, New York, 1992.
7. J. C. Bean, "The Growth of Novel Silicon Materials," *Physics Today*, **39**, 10, 36 (1986).
8. For a discussion on film deposition, see, for example, A. C. Adams, "Dielectric and Polysilicon Film Deposition," in S. M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, 1983.
9. K. Eujino, et al., "Doped Silicon Oxide Deposition by Atmospheric Pressure and Low Temperature Chemical Vapor Deposition Using Tetraethoxysilane and Ozone," *J. Electrochem. Soc.*, **138**, 3019 (1991).
10. A. C. Adams and C. D. Capio, "Planarization of Phosphorus-Doped Silicon Dioxide," *J. Electrochem. Soc.*, **127**, 2222 (1980).
11. T. Yamamoto, et al., "An Advanced 2.5nm Oxidized Nitride Gate Dielectric for Highly Reliable 0.25 μ m MOSFETs," *Symp. VLSI Technol. Dig. Tech. Pap.*, p. 45 (1997).
12. K. Kumar, et al., "Optimization of Some 3 nm Gate Dielectrics Grown by Rapid Thermal Oxidation in a Nitric Oxide Ambient," *Appl. Phys. Lett.*, **70**, 384 (1997).
13. T. Homma, "Low Dielectric Constant Materials and Methods for Interlayer Dielectric Films in Ultralarge-Scale Integrated Circuit Multilevel Interconnects," *Mater. Sci. Eng.*, **23**, 243 (1998).
14. H. N. Yu, et al., "1 μ m MOSFET VLSI Technology. Part I—An Overview," *IEEE Trans. Electron Devices*, **ED-26**, 318 (1979).
15. J. M. Andrews, "Electrical Conduction in Implanted Polycrystalline Silicon," *J. Electron. Mater.*, **8**, 3, 227 (1979).
16. M. J. Buiting, A. F. Otterloo, and A. H. Montree, "Kinetic Aspects of the LPCVD of Titanium Nitride from Titanium Tetrachloride and Ammonia," *J. Electrochem. Soc.*, **138**, 500 (1991).
17. R. Tobe, et al., "Plasma-Enhanced CVD of TiN and Ti Using Low-Pressure and High-Density Helicon Plasma," *Thin Solid Film*, **281-282**, 155 (1996).
18. J. Hu, et al., "Electrical Properties of Ti/TiN Films Prepared by Chemical Vapor Deposition and Their Applications in Submicron Structures as Contact and Barrier Materials," *Thin Solid Film*, **308**, 589 (1997).
19. M. Hansen and A. Anderko, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.
20. D. Pramanik and A. N. Saxena, "VLSI Metallization Using Aluminum and Its Alloys," *Solid State Tech.*, **26**(1), 127 (1983); **26**(3), 131 (1983).
21. C. L. Hu and J. M. E. Harper, "Copper Interconnections and Reliability," *Mater. Chem. Phys.*, **52**, 5 (1998).
22. P. C. Andricacos, et al., "Damascene Copper Electroplating for Chip Interconnects," *193rd Meet. Electrochem. Soc.*, p. 3 (1998).
23. J. M. Steigerwald, et al., *Chemical Mechanical Planarization of Microelectronic Materials*, Wiley, New York, 1997.
24. L. M. Cook, et al., "Theoretical and Practical Aspects of Dielectric and Metal CMP," *Semicond. Int.*, p. 141 (1995).

► PROBLEMS

Asterisks denote difficult problems.

SECTION 8.1: EPITAXIAL GROWTH TECHNIQUES

- *1. Find the average molecular velocity of air at 300 K (the molecular weight for air is 29).
2. The distance between source and wafer in a deposition chamber is 15 cm. Estimate the pressure at which this distance become 10% of the mean free path of source molecules.
- *3. Find the number of atoms per unit area, N_s , needed to form a monolayer under close-packing condition (i.e., each atom is in contact with its six neighboring atoms), assuming the diameter d of the atom is 4.68 Å.
- *4. Assume an effusion oven geometry of $A = 5 \text{ cm}^2$ and $L = 12 \text{ cm}$. (a) Calculate the arrival rate of gallium and the MBE growth rate for the effusion oven filled with gallium arsenide at 970°C. (b) For a tin effusion oven operated at 700°C under the same geometry, calculate the doping concentration (assuming tin atoms are fully incorporated in the gallium arsenide grown at the aforementioned rate). The molecular weight for tin is 118.69, and the pressure at 700°C for tin is $2.66 \times 10^{-6} \text{ Pa}$.

SECTION 8.2: STRUCTURES AND DEFECTS IN EPITAXIAL LAYERS

5. Find the maximum percentage of In (i.e., the x value for $\text{Ga}_{1-x}\text{In}_x\text{As}$) film grown on GaAs substrate without formation of misfit dislocation, if the final film thickness is 10 nm.
6. The lattice misfit, f , of a film is defined as $f = [a_0(s) - a_0(f)]/a_0(f) = \Delta a_0/a_0$ where $a_0(s)$ and $a_0(f)$ are the unstrained lattice constants of the substrate and the film, respectively. Find the f values for InAs–GaAs and Ge–Si systems.

SECTION 8.3: DIELECTRIC DEPOSITION

7. (a) In a plasma-deposited silicon nitride that contains 20 at% hydrogen and has a silicon-to-nitrogen ratio (Si/N) of 1.2, find x and y in the empirical formula of $\text{Si}_x\text{H}_y\text{N}_z$. (b) If the variation of film resistivity with Si/N ratio is given by $5 \times 10^{25} \exp(-33.3\gamma)$ for $2 > \gamma > 0.8$, where γ is the ratio, find the resistivity of the film in (a).
8. The dielectric constants of SiO_2 , Si_3N_4 , and Ta_2O_5 are about 3.9, 7.6, and 25, respectively. What is the capacitance ratio for the capacitors with the Ta_2O_5 and oxide/nitride/oxide dielectrics for the same dielectric thickness, provided the oxide/nitride/oxide has thickness ratio 1:1:1 for the oxide to the nitride?
9. In Problem 8, if BST with a dielectric constant of 500 is chosen to replace Ta_2O_5 , calculate the area reduction ratio to maintain the same capacitance if the two films have the same thickness.
10. In Problem 8, calculate the equivalent thickness of the Ta_2O_5 in terms of SiO_2 thickness if both have the same capacitance. Assume the actual thickness of Ta_2O_5 is 3 μm .
11. In a silane–oxygen reaction to deposit undoped SiO_2 film, the deposition rate is 15 nm/min at 425°C. What temperature is required to double the deposition rate?
12. The P-glass flow process requires temperatures above 1000°C. As device dimensions become smaller in ULSI, we must use lower temperatures. Suggest methods to obtain a smooth topography at < 900°C for deposited silicon dioxide that can be used as an insulator between metal layers.

SECTION 8.4: POLYSILICON DEPOSITION

13. Why is silane more often used for polysilicon deposition than silicon chloride?
14. Explain why the deposition temperature for polysilicon films is moderately low, usually between 600°C and 650°C.

SECTION 8.5: METALLIZATION

15. An e-beam evaporation system is used to deposit aluminum to form MOS capacitors. If the flatband voltage of the capacitance is shifted by 0.5 V because of e-beam radiation, find the number of fixed oxide charges (the silicon dioxide thickness is 50 nm). How can these charges be removed?
16. A metal line ($L = 20 \mu\text{m}$, $W = 0.25 \mu\text{m}$) has a sheet resistance of $5 \Omega/\square$. Calculate the resistance of the metal line.
17. Calculate the thicknesses of the TiSi_2 and CoSi_2 , where the initial Ti and Co film thicknesses are 30 nm.
18. Compare the advantages and disadvantages of TiSi_2 and CoSi_2 for salicide applications.
19. A dielectric material is placed between the two parallel metal lines. The length $L = 1 \text{ cm}$, width $W = 0.28 \mu\text{m}$, thickness $T = 0.3 \mu\text{m}$, and spacing $S = 0.36 \mu\text{m}$. (a) Calculate the RC time delay. The metal is Al with a resistivity of $2.67 \mu\Omega\text{-cm}$, and the dielectric is oxide with dielectric constant 3.9. (b) Calculate the RC time delay. The metal is Cu with a resistivity of $1.7 \mu\Omega\text{-cm}$, and the dielectric is organic polymer with dielectric constant 2.8. (c) Compare the results in (a) and (b). How much can we decrease the RC time delay?
20. Repeat Problems 19 (a) and (b) if the fringing factor for the capacitors is 3. The fringing factor is due to the spreading of the electric-field lines beyond the length and the width of the metal lines.
- *21. To avoid electromigration problems, the maximum allowed current density in an aluminum runner is about $5 \times 10^5 \text{ A/cm}^2$. If the runner is 2 mm long, 1 μm wide, and nominally 1 μm thick, and if 20% of the runner length passes over steps and is only 0.5 μm thick there, find the total resistance of the runner if the resistivity is $3 \times 10^{-6} \Omega\text{-cm}$. Find the maximum voltage that can be applied across the runner.
- *22. To use Cu for wiring, one must overcome several obstacles: the diffusion of Cu through SiO_2 , adhesion of Cu to SiO_2 , and corrosion of Cu. One way to overcome these obstacles is to use a cladding/adhesion layer (e.g., Ta or TiN) to protect the Cu wires. Consider a clad Cu wire with a square cross section of $0.5 \mu\text{m} \times 0.5 \mu\text{m}$ and compare it with a layered TiN/Al/TiN wire of the same size, with the top and bottom TiN layers 40 nm and 60 nm thick, respectively. What is the maximum thickness of the cladding layer if the resistances of the clad Cu wire and the TiN/Al/TiN wire are the same?

Process Integration

Microwave, photonic, and power applications generally employ discrete devices. For example, an IMPATT diode is used as a microwave generator, an injection laser as an optical source, and a thyristor as a high-power switch. However, most electronic systems are built on the *integrated circuit*, which is an ensemble of both active (e.g., transistor) and passive (e.g., resistor, capacitor, and inductor) devices formed on and within a single-crystal semiconductor substrate and interconnected by a metallization pattern.¹ ICs have enormous advantages over discrete devices connected by wire bonding. The advantages include (a) reduction of the interconnection parasitics, because an IC with multilevel metallization can substantially reduce the overall wiring length; (b) full utilization of a semiconductor wafer's area, because devices can be closely packed within an IC chip; and (c) drastic reduction in processing cost, because wire bonding is a time-consuming and error-prone operation.

This chapter discusses combinations of the basic processes described in the previous chapters to fabricate active and passive components in an IC. Because the key element of an IC is the transistor, specific processing sequences are developed to optimize its performance. The chapter considers three major IC technologies associated with the three transistor families: the bipolar transistor, the MOSFET, and the MESFET. In addition, it discusses the fabrication of microelectromechanical systems by micromachining techniques. Specifically, it covers the following topics:

- The design and fabrication of IC resistors, capacitors, and inductors
- The processing sequence for standard bipolar transistor and advanced bipolar devices
- The processing sequence for MOSFETs, with special emphasis on CMOS and memory devices
- The processing sequence for high-performance MESFETs and monolithic microwave ICs
- The major challenges for future microelectronics, including ultrashallow junction, ultrathin oxide, new interconnection materials, low power dissipation, and isolation
- Microelectromechanical systems formed by orientation-dependent etching, sacrificial etching, or LIGA (lithography, electroplating, and molding) processes
- The simulation of IC fabrication processes using SUPREM

Figure 9.1 illustrates the interrelationship between the major process steps used for IC fabrication. Polished wafers with a specific resistivity and orientation are used as the starting material. The film formation steps include thermally grown oxide films (Chapter 3) and deposited polysilicon, dielectric, and metal films (Chapter 8). Film formation is often

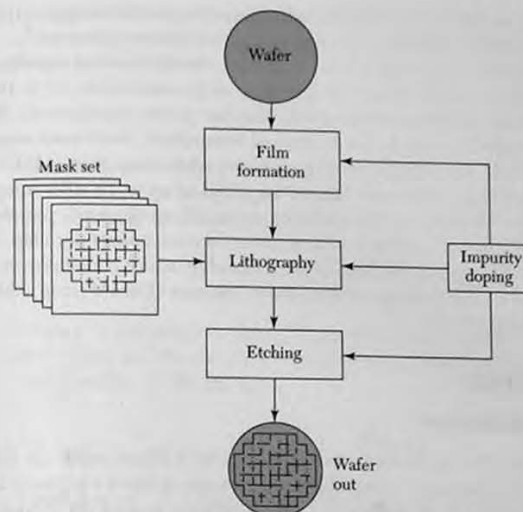


Figure 9.1 Schematic flow diagram of integrated circuit fabrication.

followed by lithography (Chapter 4) or impurity doping (Chapters 6 and 7). Lithography is generally followed by etching (Chapter 5), which in turn is often followed by another impurity doping or film formation. The final IC is made by sequentially transferring the patterns from each mask, level by level, onto the surface of the semiconductor wafer.

After processing, each wafer contains hundreds of identical rectangular chips (or *dice*), typically between 1 and 20 mm on each side, as shown in Figure 9.2a. The chips are separated by sawing or laser cutting; Figure 9.2b shows a separated chip. Schematic top views of a single MOSFET and a single bipolar transistor are shown in Figure 9.2c to give some perspective of the relative size of a component in an IC chip. Prior to chip separation, each chip is electrically tested (see Chapter 10). Defective chips are usually marked with

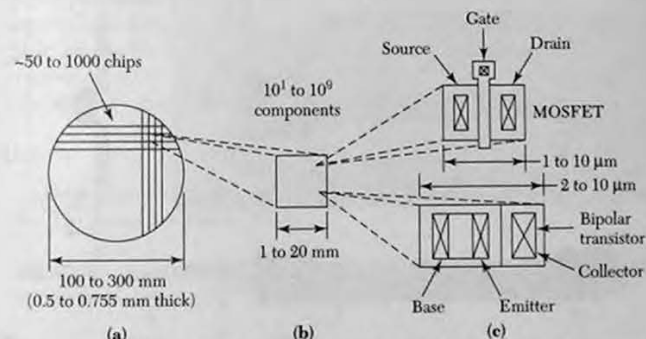


Figure 9.2 Size comparison of a wafer to individual components. (a) Semiconductor wafer. (b) Chip. (c) MOSFET and bipolar transistor.

a dab of black ink. Good chips are selected and packaged to provide an appropriate thermal, electrical, and interconnection environment for electronic applications.²

IC chips may contain a few components (transistors, diodes, resistors, capacitors, etc.) or as many as a billion or more. Since the invention of the monolithic IC in 1959, the number of components on a state-of-the-art IC chip has grown exponentially. We usually refer to the complexity of an IC as its level of integration. *Small-scale integration* (SSI) refers to up to 100 components per chip, *medium-scale integration* (MSI) to up to 1000 components per chip, *large-scale integration* (LSI) to up to 100,000 components per chip, *very large-scale integration* (VLSI) to up to 10^7 components per chip, and *ultralarge-scale integration* (ULSI) to larger numbers of components per chip. Section 9.3 describes two ULSI chips: a 32-bit microprocessor chip, which contains over 42 million components, and a 1-Gb dynamic random access memory (DRAM) chip, which contains over 2 billion components.

9.1 PASSIVE COMPONENTS

9.1.1 The Integrated Circuit Resistor

To form an IC resistor, we can deposit a resistive layer on a silicon substrate, then pattern the layer by photolithography and etching. We can also define a window in a silicon dioxide layer grown thermally on a silicon substrate and then implant (or diffuse) impurities of the opposite conductivity type into the wafer. Figure 9.3 shows the top and cross-sectional views of two resistors formed by the latter approach: One has a meander shape, and the other has a bar shape.

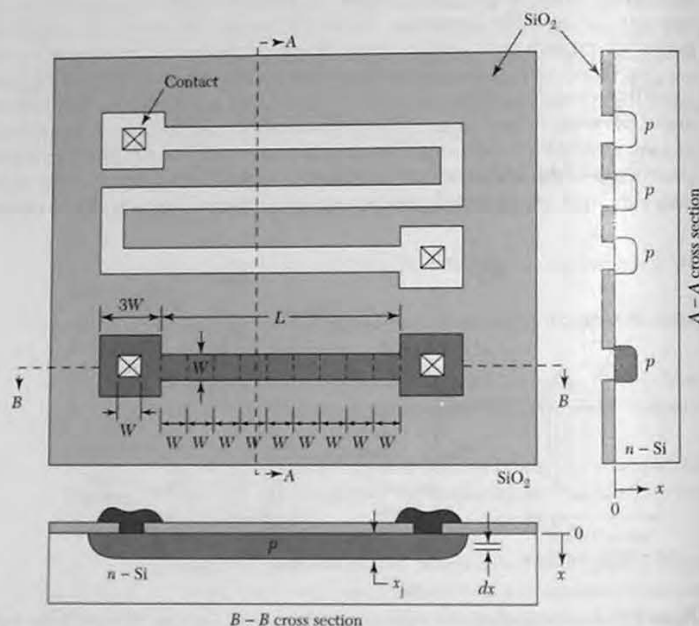


Figure 9.3 Integrated circuit resistors. All narrow lines in the large square area have the same width, W , and all contacts are the same size.

Consider the bar-shaped resistor first. The differential conductance dG of a thin layer of the p -type material that is of thickness dx parallel to the surface and at a depth x (as shown by the B - B cross section) is

$$dG = q\mu_p p(x) \frac{W}{L} dx \quad (1)$$

where W is the width of the bar, L is the length of the bar (we neglect the end contact areas for the time being), μ_p is the mobility of a hole, and $p(x)$ is the doping concentration. The total conductance of the entire implanted region of the bar is given by

$$G = \int_0^{x_j} dG = q \frac{W}{L} \int_0^{x_j} \mu_p p(x) dx \quad (2)$$

where x_j is the junction depth. If the value of μ_p (which is a function of the hole concentration), and the distribution of $p(x)$ are known, the total conductance can be evaluated from Eq. 2. We can write

$$G \equiv g \frac{W}{L} \quad (3)$$

where $g \equiv q \int_0^{x_j} \mu_p p(x) dx$ is the conductance of a square resistor pattern; that is, $G = g$ when $L = W$.

The resistance is therefore given by

$$R \equiv \frac{1}{G} = \frac{L}{W} \left(\frac{1}{g} \right) \quad (4)$$

where $1/g$ usually is defined by the symbol $R_{\square}$ and is called the sheet resistance. The sheet resistance has units of ohms but is conventionally specified in units of ohms per square ($\Omega/\square$).

Many resistors in an integrated circuit are fabricated simultaneously by defining different geometric patterns in the mask, such as those shown in Figure 9.3. Since the same processing cycle is used for all these resistors, it is convenient to separate the resistance into two parts: the sheet resistance $R_{\square}$, determined by the implantation (or diffusion) process; and the ratio L/W , determined by the pattern dimensions. Once the value of $R_{\square}$ is known, the resistance is given by the ratio L/W , or the number of squares (each square has an area of $W \times W$) in the resistor pattern. The end contact areas will introduce additional resistance to the IC resistors. For the type shown in Figure 9.3, each end contact corresponds to approximately 0.65 squares. For the meander-shape resistor, the electric-field lines at the bends are not spaced uniformly across the width of the resistor but are crowded toward the inside corner. A square at the bend does not contribute exactly 1 square, but rather 0.65 squares.

EXAMPLE 1

Find the value of a resistor 90 μm long and 10 μm wide, such as the bar-shaped resistor in Figure 9.3. The sheet resistance is 1 $\text{k}\Omega/\square$.

SOLUTION The resistor contains 9 squares. The two end contacts correspond to 1.3 $\square$. The value of the resistor is $(9 + 1.3) \times 1 \text{ k}\Omega/\square = 10.3 \text{ k}\Omega$.

9.1.2 The Integrated Circuit Capacitor

Basically, two types of capacitors are used in integrated circuits: MOS capacitors and p - n junctions. The MOS capacitor can be fabricated by using a heavily doped region (such

as an emitter region) as one plate, the top metal electrode as the other plate, and the intervening oxide layer as the dielectric. The top and cross-sectional views of a MOS capacitor are shown in Figure 9.4a. To form a MOS capacitor, a thick oxide layer is thermally grown on a silicon substrate. Next, a window is lithographically defined and etched in the oxide. Diffusion or ion implantation is used to form a p^+ region in the window area, whereas the surrounding thick oxide serves as a mask. A thin oxide layer is then thermally grown in the window area, followed by a metallization step. The capacitance per unit area is given by

$$C = \frac{\epsilon_{ox}}{d} \quad (5)$$

where ϵ_{ox} is the dielectric permittivity of silicon dioxide (the dielectric constant ϵ_{ox}/ϵ_0 is 3.9) and d is the oxide thickness. To increase the capacitance further, insulators with higher dielectric constants are being studied, such as Si_3N_4 and Ta_2O_5 , which have dielectric constants of 7 and 25, respectively. The MOS capacitance is essentially independent of the applied voltage, because the lower plate of the capacitor is made of heavily doped material. This also reduces the associated series resistance.

A p - n junction is sometimes used as a capacitor in an integrated circuit. The top and cross-sectional views of an n^+ - p junction capacitor are shown in Figure 9.4b. The detailed fabrication process is considered in Section 9.2, because this structure forms part of a bipolar transistor. As a capacitor, the device is usually reverse biased; that is, the p region is reverse biased with respect to the n^+ region. The capacitance is not a constant, but varies as $(V_R + V_{bi})^{-1/2}$, where V_R is the applied reverse voltage and V_{bi} is the built-in potential of the junction. The series resistance is considerably higher than that of a MOS capacitor because the p region has higher resistivity than does the p^+ region.

EXAMPLE 2

What is the stored charge and the number of electrons on a MOS capacitor with an area of $4 \mu\text{m}^2$ for a dielectric of (a) 10-nm thick SiO_2 and (b) 5-nm thick Ta_2O_5 ? The applied voltage is 5 V for both cases.

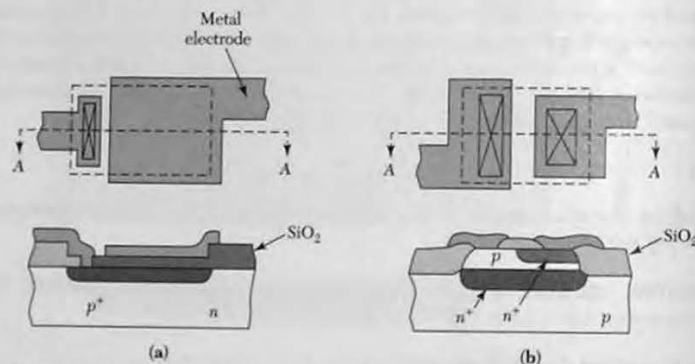


Figure 9.4 (a) Integrated MOS capacitor. (b) Integrated p - n junction capacitor.

SOLUTION

$$\begin{aligned} (a) \quad Q &= \epsilon_{ox} \times A \times \frac{V_a}{d} \\ &= 3.9 \times 8.85 \times 10^{-14} \text{ F/cm} \times 4 \times 10^{-8} \text{ cm}^2 \times \frac{5 \text{ V}}{10^{-6} \text{ cm}} \\ &= 6.9 \times 10^{-14} \text{ C} \\ \text{Number of electrons} &= 6.9 \times 10^{-14} \text{ C} / q = 4.3 \times 10^5 \end{aligned}$$

(b) Changing the dielectric constant from 3.9 to 25 and the thickness from 10 nm to 5 nm, we obtain $Q_a = 8.85 \times 10^{-13} \text{ C}$, and number of electrons $= 8.85 \times 10^{-13} \text{ C} / q = 5.53 \times 10^6$.

9.1.3 The Integrated Circuit Inductor

IC inductors have been widely used in III-V-based monolithic microwave integrated circuits (MMICs).³ With the increased speed of silicon devices and advancement in multilevel interconnection technology, IC inductors have started to receive more and more attention in silicon-based radio frequency and high-frequency applications. Many kinds of inductors can be fabricated using IC processes. The most popular method is the thin-film spiral inductor. Figures 9.5a and b show the top view and the cross section of a silicon-based, two-level-metal spiral inductor. To form a spiral inductor, a thick oxide is thermally grown or deposited on a silicon substrate. The first metal is then deposited and defined as one end of the inductor. Next, another dielectric is deposited onto metal 1. A via hole is defined lithographically and etched in the oxide. Metal 2 is deposited and the via hole is filled. The spiral pattern can be defined and etched on metal 2 as the second end of the inductor.

To evaluate the inductor, an important figure of merit is the quality factor, Q , which is defined as $L/\omega R$, where L , R , and ω are the inductance, resistance, and frequency, respectively. The higher the Q value, the lower the loss from resistance, and hence the better the performance. Figure 9.5c shows the equivalent circuit model for an IC inductor. R_i is the inherent resistivity of the metal, C_{p1} and C_{p2} are the coupling capacitances between the metal lines and the substrate, and R_{sub1} and R_{sub2} are the resistances of the silicon substrate associated with the metal lines, respectively. The Q value increases linearly with frequency initially and then drops at higher frequencies because of parasitic resistances and capacitances.

Some approaches exist for improving the Q value. The first is to use low-dielectric-constant materials (< 3.9) to reduce C_p . The other is to use a thick-film metal or low-resistivity metals (e.g., Cu, Au to replace Al) to reduce R_i . The third approach uses an insulating substrate (e.g., silicon on sapphire, silicon on glass, or quartz) to reduce R_{sub} .

To obtain the exact value of a thin-film inductor, complicated simulation tools must be employed for both circuit simulation and inductor optimization. The model for thin-film inductors must take into account the resistance of the metal, the capacitance of the oxide, line-to-line capacitance, the resistance of the substrate, the capacitance to the substrate, and the inductance and mutual inductance of the metal lines. Hence, it is more difficult to calculate the integrated inductance than to calculate integrated capacitance or resistance. However, a simple equation to estimate the square planar spiral inductor is given as³

$$L \approx \mu_0 n^2 r \approx 1.2 \times 10^{-6} n^2 r \quad (6)$$

where μ_0 is the permeability in vacuum ($4\pi \times 10^{-7} \text{ H/m}$), L is in henries, n is the number of turns, and r is the radius of the spiral in meters.

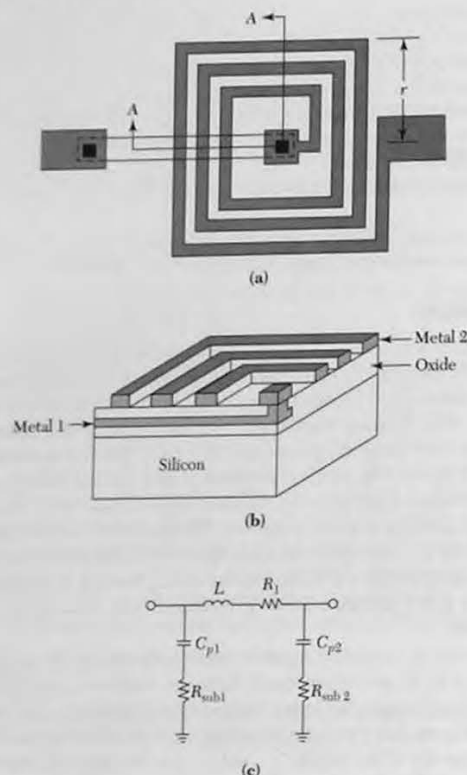


Figure 9.5 (a) Schematic view of a spiral inductor on a silicon substrate. (b) Perspective view along A-A'. (c) An equivalent circuit model for an integrated inductor.

EXAMPLE 3

For an integrated inductor with an inductance of 10 nH, what is the required radius if the number of turns is 20?

SOLUTION According to Eq. 6,

$$r = \frac{10 \times 10^{-9}}{1.2 \times 10^{-6} \times 20^2} = 2.08 \times 10^{-5} \text{ m} = 20.8 \text{ } \mu\text{m}$$

► 9.2 BIPOLAR TECHNOLOGY

For IC applications, especially for VLSI and ULSI, the size of bipolar transistors must be reduced to meet high density requirements. Figure 9.6 illustrates the reduction in the size of the bipolar transistor in recent years.⁴ The main differences in a bipolar transistor in an IC, compared with a discrete transistor, are that all electrode contacts are

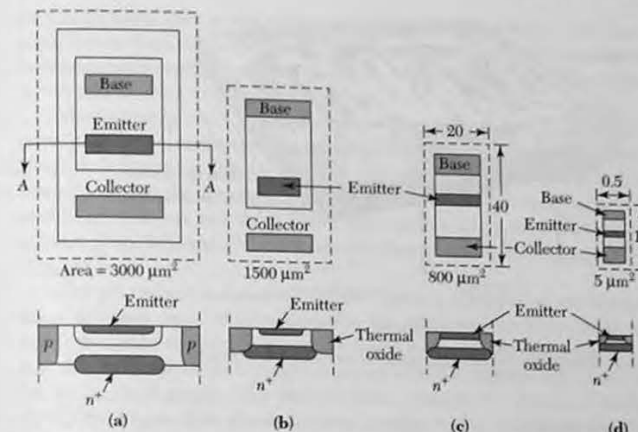


Figure 9.6 Reduction of the horizontal and vertical dimensions of a bipolar transistor. (a) Junction isolation. (b) Oxide isolation. (c and d) Scaled oxide isolation.⁴

located on the top surface of the IC wafer, and each transistor must be electrically isolated to prevent interactions between devices. Prior to 1970, both the lateral and vertical isolations were provided by p - n junctions (Fig. 9.6a), and the lateral p isolation region was always reverse biased with respect to the n -type collector. In 1971, thermal oxide was used for lateral isolation, resulting in a substantial reduction in device size (Fig. 9.6b) because the base and collector contacts abut the isolation region. In the mid-1970s, the emitter extended to the walls of the oxide, resulting in an additional reduction in area (Fig. 9.6c). At the present time, all lateral and vertical dimensions have been scaled down, and emitter stripe widths have dimensions in the submicron region (Fig. 9.6d).

9.2.1 The Basic Fabrication Process

The majority of bipolar transistors used in ICs are of the n - p - n type because the higher mobility of minority carriers (electrons) in the base region results in higher-speed performance than can be obtained with p - n - p types. Figure 9.7 shows a perspective view of an n - p - n bipolar transistor, in which lateral isolation is provided by oxide walls and vertical isolation is provided by the n^+ - p junction. The lateral oxide isolation approach reduces not only the device size, but also the parasitic capacitance because of the smaller dielectric constant of silicon dioxide (3.9, compared with 11.9 for silicon). This section now considers the major process steps that are used to fabricate the device shown in Figure 9.7.

For an n - p - n bipolar transistor, the starting material is a p -type, lightly doped ($\sim 10^{15} \text{ cm}^{-3}$), $\langle 111 \rangle$ - or $\langle 100 \rangle$ -oriented, polished silicon wafer. Because the junctions are formed inside the semiconductor, the choice of crystal orientation is not as critical as for MOS devices (see Section 9.3). The first step is to form a buried layer. The main purpose of this layer is to minimize the series resistance of the collector. A thick oxide (0.5–1 μm) is thermally grown on the wafer, and a window is then opened in the oxide. A precisely controlled amount of low-energy arsenic ions ($\sim 30 \text{ keV}$, $\sim 10^{15} \text{ cm}^{-2}$) is implanted into the window region to serve as a predeposit (Fig. 9.8a). Next, a high-temperature ($\sim 1100^\circ\text{C}$) drive-in step forms the n^+ buried layer, which has a typical sheet resistance of approximately $20 \text{ } \Omega/\square$.

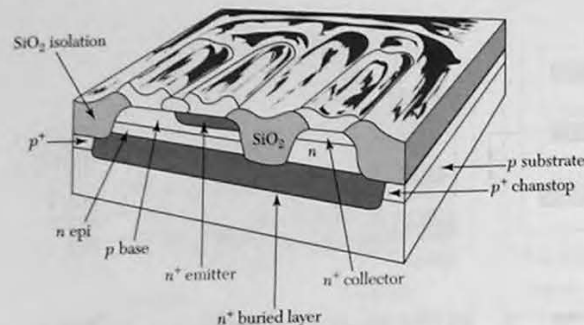


Figure 9.7 Perspective view of an oxide-isolated bipolar transistor.

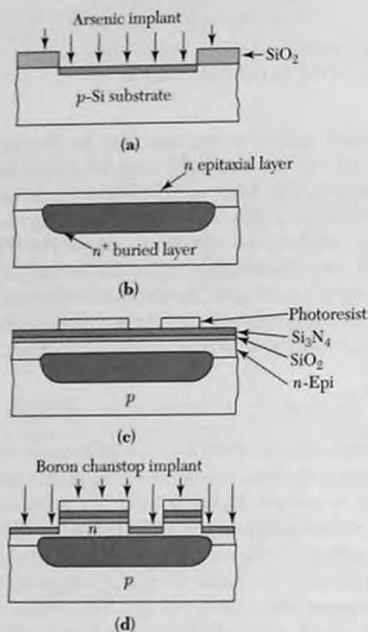


Figure 9.8 Cross-sectional views of bipolar transistor fabrication. (a) Buried-layer implantation. (b) Epitaxial layer. (c) Photoresist mask. (d) Chanstop implant.

The second step is to deposit an n -type epitaxial layer. The oxide is removed and the wafer is placed in an epitaxial reactor for epitaxial growth. The thickness and the doping concentration of the epitaxial layer are determined by the ultimate use of the device. Analog circuits (with their higher voltages for amplification) require thicker layers ($\sim 10 \mu\text{m}$) and lower dopings ($\sim 5 \times 10^{15} \text{ cm}^{-3}$), whereas digital circuits (with their lower voltages for switching) require thinner layers ($\sim 3 \mu\text{m}$) and higher dopings ($\sim 2 \times 10^{16} \text{ cm}^{-3}$). Figure 9.8b shows

a cross-sectional view of the device after the epitaxial process. Note that there is some outdiffusion from the buried layer into the epitaxial layer. To minimize outdiffusion, a low-temperature epitaxial process should be employed, and low-diffusivity impurities (e.g., As) should be used in the buried layer.

The third step is to form the lateral oxide isolation region. A thin oxide pad ($\sim 50 \text{ nm}$) is thermally grown on the epitaxial layer, followed by a silicon-nitride deposition ($\sim 100 \text{ nm}$). If nitride is deposited directly onto the silicon without the thin oxide pad, the nitride may cause damage to the silicon surface during subsequent high-temperature steps. Next, the nitride-oxide layers and about half of the epitaxial layer are etched using a photoresist as mask (Figs. 9.8c and d). Boron ions are then implanted into the exposed silicon areas (Fig. 9.8d).

The photoresist is removed, and the wafer is placed in an oxidation furnace. Since the nitride layer has a very low oxidation rate, thick oxides will be grown only in the areas not protected by the nitride layer. The isolation oxide is usually grown to a thickness such that the top of the oxide becomes coplanar with the original silicon surface to minimize the surface topography. This oxide isolation process is called *local oxidation of silicon* (LOCOS). Figure 9.9a shows the cross section of the isolation oxide after the removal of the nitride layer. Because of segregation effects, most of the implanted boron ions are pushed underneath the isolation oxide to form a p^+ layer. This is called the p^+ channel stop (or *chanstop*) because the high concentration of p -type semiconductor will prevent

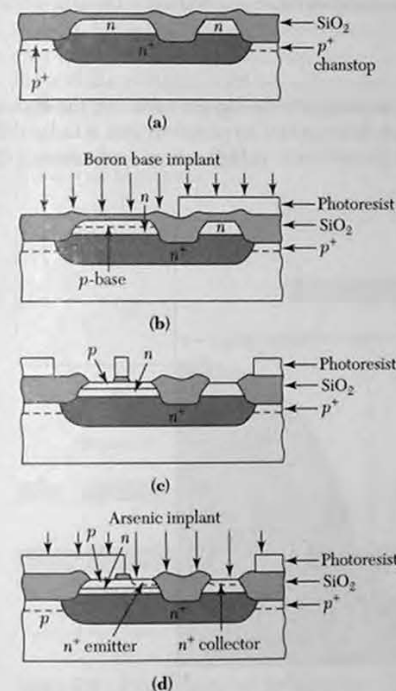


Figure 9.9 Cross-sectional views of bipolar transistor fabrication. (a) Oxide isolation. (b) Base implant. (c) Removal of thin oxide. (d) Emitter and collector implant.

surface inversion and eliminate possible high-conductivity paths (or channels) among neighboring buried layers.

The fourth step is to form the base region. A photoresist is used as a mask to protect the right half of the device. Then, boron ions ($\sim 10^{12} \text{ cm}^{-2}$) are implanted to form the base regions, as shown in Figure 9.9b. Another lithographic process removes all the thin pad oxide except a small area near the center of the base region (Fig. 9.9c).

The fifth step is to form the emitter region. As shown in Figure 9.9d, the base contact area is protected by a photoresist mask. Then, a low-energy, high-arsenic-dose ($\sim 10^{16} \text{ cm}^{-2}$) implantation forms the n^+ emitter and the n^+ collector contact regions. The photoresist is removed, and a final metallization step forms the contacts to the base, emitter, and collector, as shown in Figure 9.7.

In this basic bipolar process, there are six film formation operations, six lithographic operations, four ion implantations, and four etching operations. Each operation must be precisely controlled and monitored. Failure of any one of the operations generally will render the wafer useless.

The doping profiles of the completed transistor along a coordinate perpendicular to the surface and passing through the emitter, base, and collector are shown in Figure 9.10. The emitter profile is abrupt because of the concentration-dependent diffusivity of arsenic. The base doping profile beneath the emitter can be approximated by a Gaussian distribution for limited source diffusion. The collector doping is given by the epitaxial doping level ($\sim 2 \times 10^{16} \text{ cm}^{-3}$) for a representative switching transistor. However, at larger depths, the collector doping concentration increases because of outdiffusion from the buried layer.

9.2.2 Dielectric Isolation

In the isolation scheme described previously for the bipolar transistor, the device is isolated from other devices by the oxide layer around its periphery and is isolated from its common substrate by a n^+ p junction (buried layer). In high-voltage applications, a different

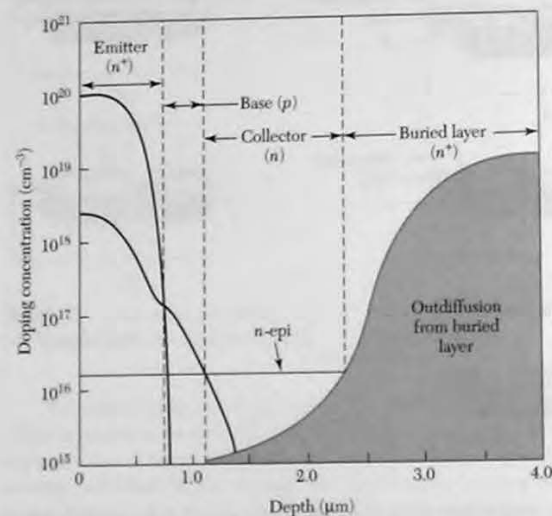


Figure 9.10 n - p - n transistor doping profiles.

approach, called *dielectric isolation*, is used to form insulating tubs to isolate a number of pockets of single-crystal semiconductors. In this approach, the device is isolated from both its common substrate and its surrounding neighbors by a dielectric layer.

A process sequence for dielectric isolation is shown in Figure 9.11. An oxide layer is formed inside a $\langle 100 \rangle$ -oriented n -type silicon substrate using high-energy oxygen ion implantation (Fig. 9.11a). Next, the wafer undergoes a high-temperature annealing process so that the implanted oxygen will react with silicon to form the oxide layer. The damage resulting from implantation is also annealed out in this process (Fig. 9.11b). After this, we obtain an n silicon layer that is fully isolated on an oxide (namely, a silicon-on-insulator, or SOI, layer). This process is called SIMOX (separation by implanted oxygen). Since the top silicon is so thin, the isolation region is easily formed by the LOCOS process illustrated in Figure 9.8c or by etching a trench (Fig. 9.11c) and refilling it with oxide (Fig. 9.11d). The other processes are almost the same as those in Figures 9.8c through 9.9 to form the p -type base, n^+ emitter, and n collector.

The main advantage of this technique is its high breakdown voltage between the emitter and the collector, which can be in excess of several hundred volts. This technique is also compatible with modern CMOS integration (Section 9.3.3). This CMOS-compatible process is very useful for mixed high-voltage and high-density ICs.

9.2.3 Self-Aligned Double-Polysilicon Bipolar Structures

The process shown in Figure 9.9c needs another lithographic step to define an oxide region to separate the base and emitter contact regions. This gives rise to a large inactive device area within the isolated boundary, which increases not only the parasitic capacitances, but also the resistance that degrades transistor performance. The most effective way to reduce these effects is by using a *self-aligned* structure.

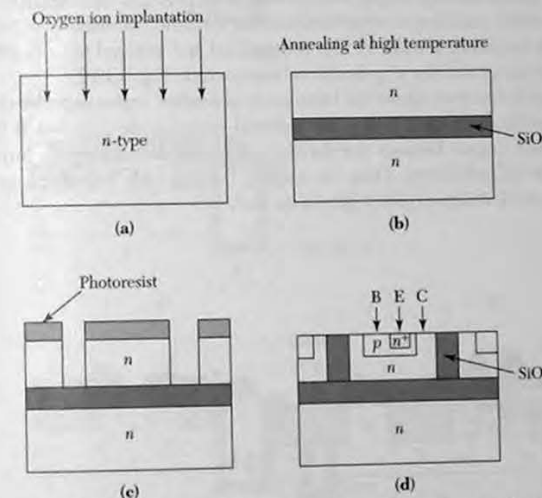


Figure 9.11 Process sequence for dielectric isolation bipolar device using silicon-on-insulator for high-voltage application. (a) Oxygen ion implantation. (b) Annealing at high temperature to form the isolation dielectric. (c) Trench isolation formed by a dry etching process. (d) Base, emitter, and collector formation.

The most widely used self-aligned structure is the double-polysilicon structure with advanced isolation provided by a trench refilled with polysilicon,⁵ shown in Figure 9.12. Figure 9.13 shows the detailed sequence of steps for the self-aligned double-polysilicon (n - p - n) bipolar structure.⁶ The transistor is built on an n -type epitaxial layer. A trench approximately 5.0 μm in depth is etched by reactive ion etching through the n^+ subcollector region into the p^- substrate region. A thin layer of thermal oxide is then grown and serves as the screen oxide for the channel stop implant of boron at the bottom of the trench. The trench is then filled with undoped polysilicon and capped by a thick planar field oxide.

The first polysilicon layer is deposited and heavily doped with boron. The p^+ polysilicon (called poly 1) will be used as a solid-phase diffusion source to form the extrinsic base region and the base electrode. This layer is covered with a chemical vapor deposition oxide and nitride (Fig. 9.13a). The emitter mask is used to pattern the emitter regions, and a dry etch process is used to produce an opening in the CVD oxide and poly 1 (Fig. 9.13b). A thermal oxide is then grown over the etched structure, and a relatively thick oxide (approximately 0.1–0.4 μm) is grown on the vertical sidewalls of the heavily doped poly. The thickness of this oxide determines the spacing between the edges of the base and emitter contacts. The extrinsic p^+ base regions are also formed during the thermal oxide growth step as a result of the outdiffusion of boron from the poly 1 into the substrate (Fig. 9.13c). Because boron diffuses laterally as well as vertically, the extrinsic base region will be able to make contact with the intrinsic base region that is formed next, under the emitter contact.

Following the oxide growth step, the intrinsic base region is formed using ion implantation of boron (Fig. 9.13d). This serves to self-align the intrinsic and extrinsic base regions. After the contact is cleaned to remove any oxide layer, the second polysilicon layer is deposited and implanted with As or P. The n^+ polysilicon (called poly 2) is used as a solid-phase diffusion source to form the emitter region and the emitter electrode. A shallow emitter region is then formed through dopant outdiffusion from poly 2. A rapid thermal anneal for the base and emitter outdiffusion steps facilitates the formation of shallow emitter–base and collector–base junctions. Finally, Pt film is deposited and sintered to form PtSi over the n^+ polysilicon emitter and the p^+ polysilicon base contact (Fig. 9.13e).

This self-aligned structure allows the fabrication of emitter regions smaller than the minimum lithographic dimension. When the sidewall-spacer oxide is grown, it fills the contact hole to some degree because the thermal oxide occupies a larger volume than the original volume of polysilicon. Thus, an opening 0.8 μm wide will shrink to about 0.4 μm if sidewall oxide 0.2 μm thick is grown on each side.

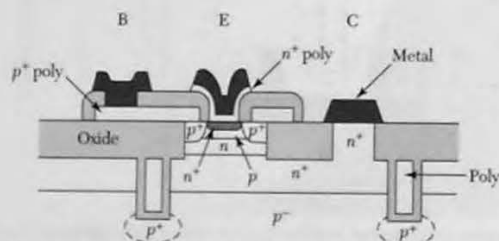


Figure 9.12 Cross section of a self-aligned double-polysilicon bipolar transistor with advanced trench isolation.⁵

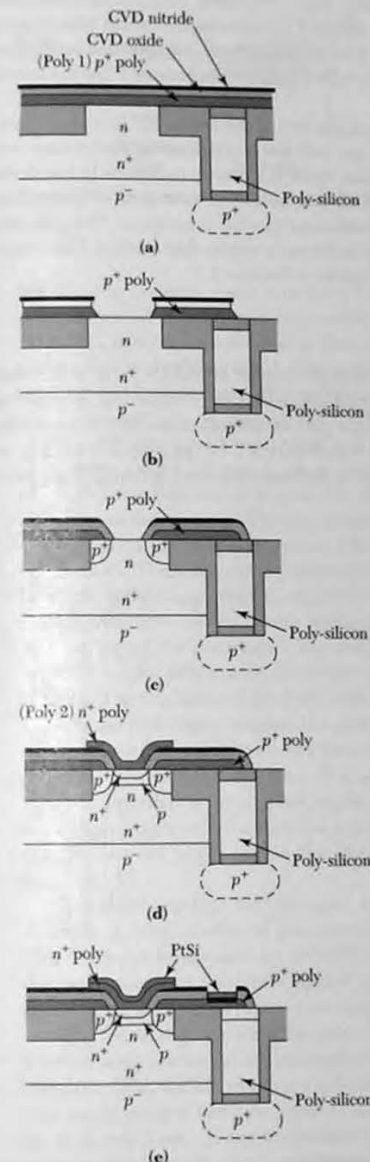


Figure 9.13 Process sequence for fabricating double-polysilicon self-aligned n - p - n transistors.⁶

9.3 MOSFET TECHNOLOGY

At present, the MOSFET is the dominant device used in ULSI circuits because it can be scaled to smaller dimensions than other types of devices. The dominant technology for MOSFET is CMOS (complementary MOSFET) technology, in which both n -channel and p -channel devices (called NMOS and PMOS, respectively) are provided on the same chip. CMOS technology is particularly attractive for ULSI circuits because it has the lowest power consumption of all IC technology.

Figure 9.14 shows the reduction in the size of the MOSFET in recent years. In the early 1970s, the gate length was $7.5\text{ }\mu\text{m}$ and the corresponding device area was about $6000\text{ }\mu\text{m}^2$. As the device is scaled down, there is a drastic reduction in the device area. For a MOSFET with a gate length of $0.5\text{ }\mu\text{m}$, the device area shrinks to less than 1% of the early MOSFET. Device miniaturization is expected to continue. The gate length will probably decrease to less than $0.10\text{ }\mu\text{m}$ in the early twenty-first century. The future trends of MOSFET devices are considered briefly in Section 9.7.

9.3.1 The Basic Fabrication Process

Figure 9.15 shows a perspective view of an n -channel MOSFET prior to final metallization.⁷ The top layer is a phosphorus-doped silicon dioxide (P-glass) that is used as an insulator between the polysilicon gate and the gate metallization and also as a gettering layer for mobile ions. Compare Figure 9.15 with Figure 9.7 for the bipolar transistor and note that a MOSFET is considerably simpler in its basic structure. Although both devices use

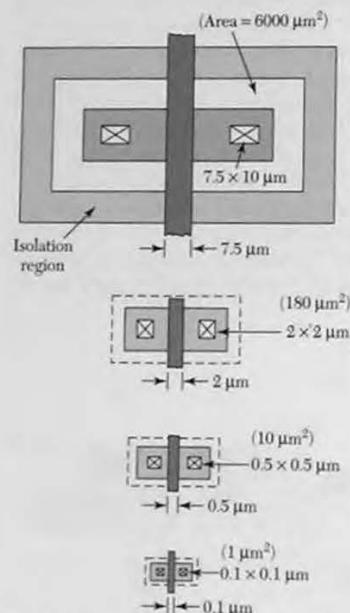


Figure 9.14 Reduction in the area of the MOSFET as the gate length (minimum feature length) is reduced.

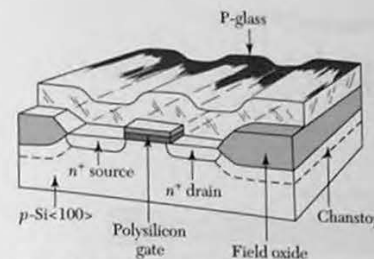


Figure 9.15 Perspective view of an n -channel MOSFET.⁷

lateral oxide isolation, there is no need for vertical isolation in the MOSFET, whereas a buried-layer n^+p junction is required in the bipolar transistor. The doping profile in a MOSFET is not as complicated as that in a bipolar transistor, and control of the dopant distribution is also less critical. This section now considers the major process steps that are used to fabricate the device in Figure 9.15.

To process an n -channel MOSFET (NMOS), the starting material is a p -type, lightly doped ($\sim 10^{15}\text{ cm}^{-3}$), $\langle 100 \rangle$ -oriented, polished silicon wafer. The $\langle 100 \rangle$ orientation is preferred over $\langle 111 \rangle$ because it has an interface trap density that is about one-tenth that of $\langle 111 \rangle$. The first step is to form the oxide isolation region using LOCOS technology. The process sequence for this step is similar to that for the bipolar transistor. A thin pad oxide ($\sim 35\text{ nm}$) is thermally grown, followed by a silicon nitride ($\sim 150\text{ nm}$) deposition (Fig. 9.16a).⁷ The active device area is defined by a photoresist mask and a boron channel stop layer and is then implanted through the composite nitride-oxide layer (Fig. 9.16b). The nitride layer not covered by the photoresist mask is subsequently removed by etching. After stripping the photoresist, the wafer is placed in an oxidation furnace to grow an oxide (called the *field oxide*), where the nitride layer is removed, and to drive in the boron implant. The thickness of the field oxide is typically 0.5 to $1\text{ }\mu\text{m}$.

The second step is to grow the gate oxide and to adjust the threshold voltage. The composite nitride-oxide layer over the active device area is removed, and a thin gate oxide layer (less than 10 nm) is grown. For an enhancement-mode n -channel device, boron ions are implanted in the channel region, as shown in Figure 9.16c, to increase the threshold voltage to a predetermined value (e.g., $+0.5\text{ V}$). For a depletion-mode n -channel device, arsenic ions are implanted in the channel region to decrease the threshold voltage (e.g., -0.5 V).

The third step is to form the gate. A polysilicon is deposited and is heavily doped by diffusion or implantation of phosphorus to a typical sheet resistance of 20 to $30\text{ }\Omega/\square$. This resistance is adequate for MOSFETs with gate lengths larger than $3\text{ }\mu\text{m}$. For smaller devices, polycide (a composite layer of metal silicide and polysilicon, such as W-polycide) can be used as the gate material to reduce the sheet resistance to about $1\text{ }\Omega/\square$.

The fourth step is to form the source and drain. After the gate is patterned (Fig. 9.16d), it serves as a mask for the arsenic implantation ($\sim 30\text{ keV}$, $\sim 5 \times 10^{15}\text{ cm}^{-2}$) to form the source and drain (Fig. 9.17a), which are self-aligned with respect to the gate.⁷ At this stage, the only overlapping of the gate is due to lateral straggling of the implanted ions (for 30-keV As , σ_x is only 5 nm). If low-temperature processes are used for subsequent steps to minimize lateral diffusion, the parasitic gate-drain and gate-source coupling capacitances can be much smaller than the gate-channel capacitance.

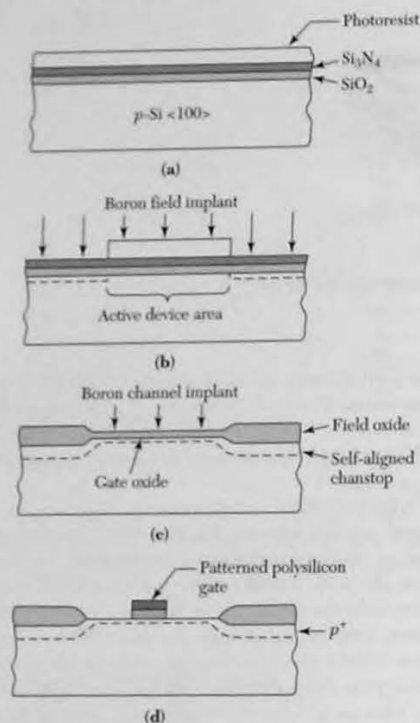


Figure 9.16 Cross-sectional view of NMOS fabrication sequence.⁷ (a) Formation of SiO_2 , Si_3N_4 , and photoresist layers. (b) Boron implant. (c) Field oxide. (d) Gate.

The last step is metallization. A phosphorus-doped oxide (P-glass) is deposited over the entire wafer and is flowed by heating the wafer to give a smooth surface topography (Fig. 9.17b). Contact windows are defined and etched in the P-glass. A metal layer, such as aluminum, is then deposited and patterned. A cross-sectional view of the completed MOSFET is shown in Figure 9.17c, and the corresponding top view is shown in Figure 9.17d. The gate contact is usually made outside the active device area to avoid possible damage to the thin gate oxide.

EXAMPLE 4

What is the maximum gate-to-source voltage that a MOSFET with a 5-nm gate oxide can withstand? Assume that the oxide breaks down at 8 MV/cm and that the substrate voltage is zero.

SOLUTION

$$V = \xi \times d = 8 \times 10^6 \times 5 \times 10^{-7} = 4 \text{ V}$$

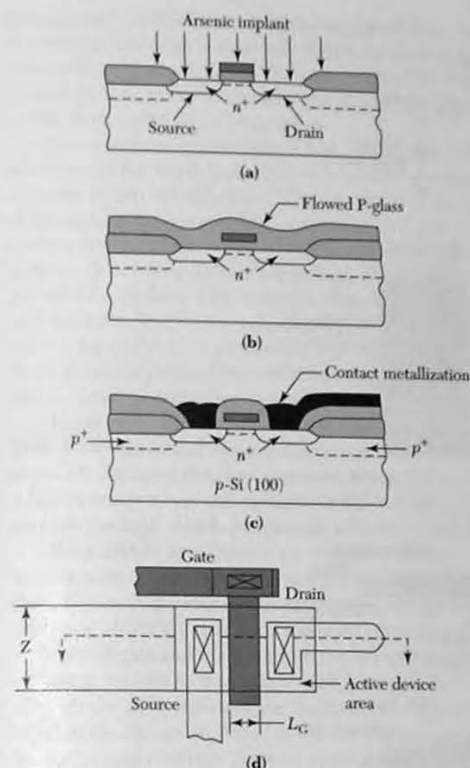


Figure 9.17 NMOS fabrication sequence.⁷ (a) Source and drain. (b) P-glass deposition. (c) Cross section of the MOSFET. (d) Top view of the MOSFET.

9.3.2 Memory Devices

Memories are devices that can store digital information (or data) in terms of *bits* (binary digits). Various memory chips have been designed and fabricated using NMOS technology. For most large memories, the random access memory (RAM) organization is preferred. In a RAM, memory cells are organized in a matrix structure, and data can be accessed (i.e., stored, retrieved, or erased) in random order, independent of its physical location. A static random access memory (SRAM) can retain stored data indefinitely as long as the power supply is on. The SRAM is basically a flip-flop circuit that can store one bit of information. An SRAM cell has four enhancement-mode MOSFETs and two depletion-mode MOSFETs. The depletion-mode MOSFETs can be replaced by resistors formed in undoped polysilicon to minimize power consumption.⁸

To reduce the cell area and power consumption, the dynamic random access memory (DRAM) has been developed. Figure 9.18a shows the circuit diagram of the one-transistor DRAM cell, in which the transistor serves as a switch and one bit of information

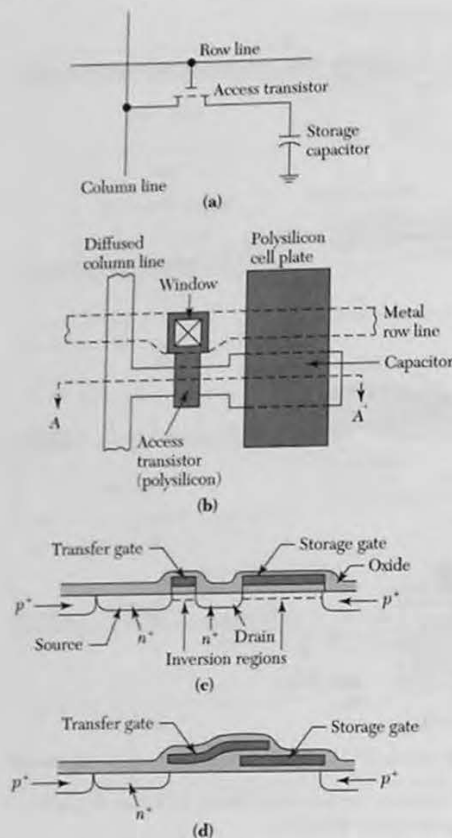


Figure 9.18 Single-transistor DRAM cell with a storage capacitor.⁸ (a) Circuit diagram. (b) Cell layout. (c) Cross section through A–A'. (d) Double-level polysilicon.

can be stored in the storage capacitor. The voltage level on the capacitor determines the state of the cell. For example, +1.5 V may be defined as logic 1, and 0 V defined as logic 0. The stored charge will be removed typically in a few milliseconds, mainly because of the leakage current of the capacitors. Thus, dynamic memories require periodic “refreshing” of the stored charge.

Figure 9.18b shows the layout of a DRAM cell, and Figure 9.18c shows the corresponding cross section through A–A'. The storage capacitor uses the channel region as one plate, the polysilicon gate as the other plate, and the gate oxide as the dielectric. The row line is a metal track to minimize the delay due to parasitic resistance (R) and parasitic capacitance (C), the product of which is the RC delay. The column line is formed by n^+ diffusion. The internal drain region of the MOSFET serves as a conductive link between the inversion layers under the storage gate and the transfer gate. The drain region can be eliminated using the double-level polysilicon approach shown in Figure 9.18d.

The second polysilicon electrode is separated from the first polysilicon capacitor plate by an oxide layer that is thermally grown on the first-level polysilicon before the second electrode has been defined. The charge from the column line can therefore be transmitted directly to the area under the storage gate by the continuity of inversion layers under the transfer and storage gates.

To meet the requirements of high-density DRAMs, the DRAM structure has been extended to the third dimension with stacked or trench capacitors. Figure 9.19a shows a simple trench cell structure.⁹ The advantage of the trench type is that the capacitance of the cell can be increased by increasing the depth of the trench without increasing the surface area of silicon occupied by the cell. The main difficulties of making trench-type cells are the etching of the deep trench, which needs a rounded bottom corner, and the growth of a uniform thin dielectric film on trench walls. Figure 9.19b shows a stacked cell structure. The storage capacitance increases as a result of stacking the storage capacitor on top of the access transistor. The dielectric is formed using the thermal oxidation or CVD nitride methods between the two polysilicon plates. Hence, the stacked cell process is easier than the trench-type process.

Figure 9.20 shows a 1-Gb DRAM chip. This memory chip uses 0.18- μm design rules. Trench capacitors and its peripheral circuits are in CMOS, which is discussed in Section 9.3.3. The memory chip has an area of 390 mm² (14.3 mm $\times$ 27.3 mm) that contains over 2 billion components and operates at 2.5 V. This 1-Gb DRAM is mounted in an 88-pin ceramic package, which can provide adequate heat dissipation.

Both SRAM and DRAM are volatile memories; that is, they lose their stored data when power is switched off. Nonvolatile memories, on the other hand, can retain their data. Figure 9.21a shows a floating-gate nonvolatile memory, which is basically a conventional MOSFET that has a modified gate electrode. The composite gate has a regular (control) gate and a floating gate that is surrounded by insulators. When a large positive voltage is applied to the control gate, charge will be injected from the channel region through the gate oxide into the floating gate. When the applied voltage is removed, the injected charge can be stored in the floating gate for a long time. To remove this charge, a large negative voltage must be applied to the control gate so that the charge will be injected back into the channel region.

Another version of the nonvolatile memory is the metal-nitride-oxide-semiconductor (MNOS) type shown in Figure 9.21b. When a positive gate voltage is applied, electrons can tunnel through the thin oxide layer (~ 2 nm), be captured by the traps at the oxide–nitride interface, and thus become stored charges there. The equivalent circuit for both types of nonvolatile memories can be represented by two capacitors in series for the

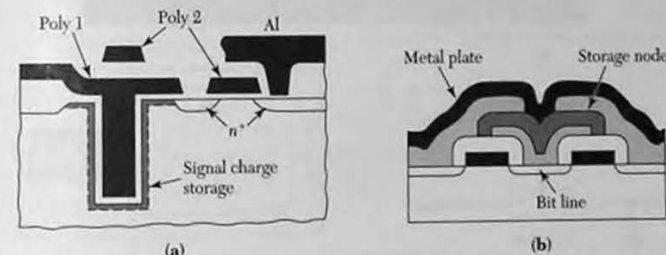


Figure 9.19 (a) DRAM with a trench cell structure.⁹ (b) DRAM with a single-layer stacked-capacitor cell.

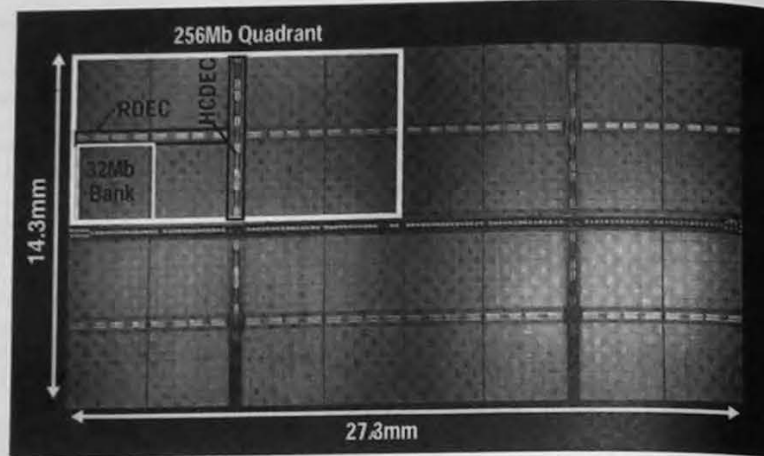


Figure 9.20 A 1-Gb DRAM that contains over 2 billion components. (Photograph courtesy of IBM/Siemens, 1999 IEEE Int. Solid State Circuit Conference.)

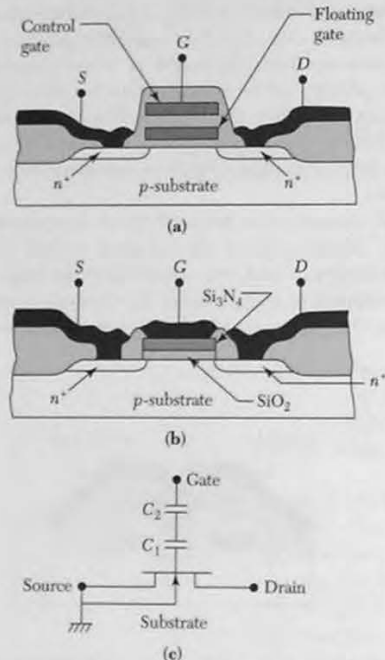


Figure 9.21 Nonvolatile memory devices. (a) Floating-gate nonvolatile memory. (b) MNOS nonvolatile memory. (c) Equivalent circuit of either type of nonvolatile memory.

gate structure, as illustrated in Figure 9.21c. The charge stored in the capacitor (C_1) causes a shift in the threshold voltage, and the device remains at the higher threshold voltage state (logic 1). For a well-designed memory device, the charge retention time can be over 100 years. To erase the memory (i.e., the stored charge) and return the device to a lower threshold voltage state (logic 0), a gate voltage or other means (such as ultraviolet light) can be used.

Nonvolatile semiconductor memory (NVSM) has been extensively used in portable electronics systems, such as cellular phones and digital cameras. Another interesting application is the chip card, also called the IC card. The photo in Figure 9.22 is an IC card. The diagram at the bottom of Figure 9.22 illustrates the nonvolatile memory device that stores the data, which can be read and written through the bus to a central processing unit (CPU). In contrast to the limited volume (1 kB) inside a conventional magnetic tape card, the size of the nonvolatile memory can be increased to 16 kB, 64 kB, or even larger depending on the application (e.g., one could store personal photos or fingerprints). Through the IC card read/write machines, the data can be used in numerous applications, including telecommunications (card telephone, mobile radio), payment transactions (electronic purse, credit card), pay television, transport (electronic ticket, public transport), health care (patient-data card), and access control. The IC card will likely play a central role in the global information and service society of the future.¹⁰

9.3.3 CMOS Technology

Figure 9.23a shows a CMOS inverter. The gate of the upper PMOS device is connected to the gate of the lower NMOS device. Both devices are enhancement-mode MOSFETs

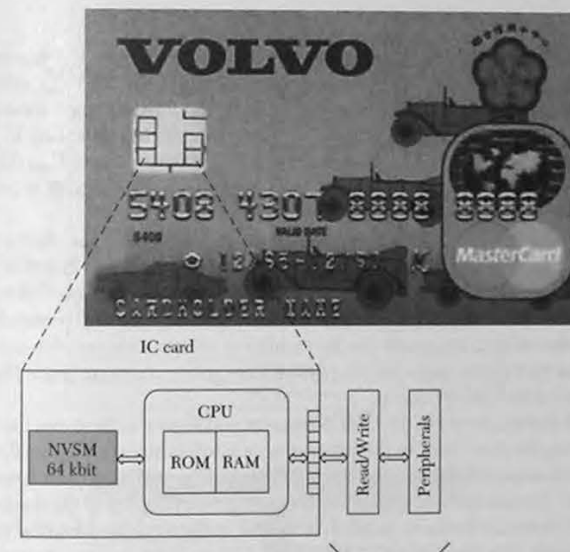


Figure 9.22 An IC card. The data stored in the NVSM can be accessed through the bus of the central processing unit (CPU). There are several metal pads connecting to the read/write machine. (Photograph courtesy of Retone Information System Co., LTD.)

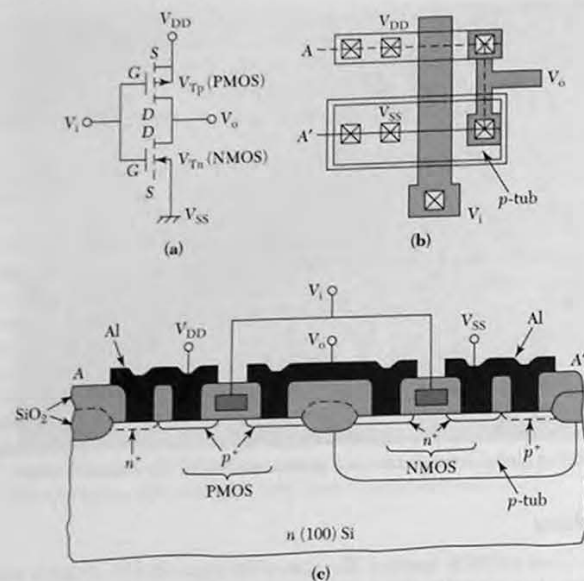


Figure 9.23 Complementary MOS (CMOS) inverter. (a) Circuit diagram. (b) Circuit layout. (c) Cross section along dotted A-A' line of (b).

with the threshold voltage V_{Tp} less than zero for the PMOS device and V_{Tn} greater than zero for the NMOS device (typically the threshold voltage is about $1/4 V_{DD}$). When the input voltage (V_i) is at ground or at small positive values, the PMOS device is turned on (the gate-to-ground potential of PMOS is $-V_{DD}$, which is more negative than V_{Tp}), and the NMOS device is off. Hence, the output voltage (V_o) is very close to V_{DD} (logic 1). When the input is at V_{DD} , the PMOS (with $V_{GS} = 0$) is turned off, and the NMOS is turned on ($V_i = V_{DD} > V_{Tn}$). Therefore, V_o equals zero (logic 0).

The CMOS inverter has a unique feature: In either logic state, one device in the series path from V_{DD} to ground is nonconductive. The current that flows in either steady state is a small leakage current, and only when both devices are on during switching does a significant current flow through the CMOS inverter. Thus, the average power dissipation is small, on the order of nanowatts. As the number of components per chip increases, power dissipation becomes a major limiting factor. Low power consumption is the most attractive feature of the CMOS circuit.

Figure 9.23b shows a layout of the CMOS inverter, and Figure 9.23c shows the device cross section along the A-A' line. In the processing, a p tub (also called a p well) is first implanted and subsequently driven into the n substrate. The p -type dopant concentration must be high enough to overcompensate the background doping of the n substrate. The subsequent processes for the n -channel MOSFET in the p tub are identical to those described previously. For the p -channel MOSFET, $^{11}\text{B}^+$ or $^{49}(\text{BF}_2)^+$ ions are implanted into the n substrate to form the source and drain regions. A channel implant of $^{75}\text{As}^+$ ions may be used to adjust the threshold voltage, and an n^+ chanstop is formed underneath the field oxide around the p -channel device. Because of the p tub and the additional steps

needed to make the p -channel MOSFET, the number of steps to make a CMOS circuit is essentially double that to make an NMOS circuit. Thus, there is a trade-off between the complexity of processing and the reduction in power consumption.

Instead of the p tub just described, an alternate approach is to use an n tub formed in p -type substrate, as shown in Figure 9.24a. In this case, the n -type dopant concentration must be high enough to overcompensate for the background doping of the p substrate (i.e., $N_D > N_A$). In both the p -tub and the n -tub approach, the channel mobility will be degraded because mobility is determined by the total dopant concentration ($N_A + N_D$). An approach using two separated tubs implanted into a lightly doped substrate is shown in Figure 9.24b. This structure is called a *twin tub*.¹ Because no overcompensation is needed in either of the twin tubs, higher channel mobility can be obtained.

All CMOS circuits have the potential for a troublesome problem called *latchup* that is associated with parasitic bipolar transistors. These parasitic devices consist of the n - p - n transistor formed by the NMOS source/drain regions, p tub, and n -type substrate, as well as the p - n - p transistor formed by the PMOS source/drain regions, n -type substrate, and p tub. Under appropriate conditions, the collector of the p - n - p device supplies base current to the n - p - n , and vice versa in a positive feedback arrangement. This latchup current can have serious negative repercussions in a CMOS circuit.

An effective processing technique to eliminate the latchup problem is to use deep-trench isolation, as shown in Figure 9.24c.¹¹ In this technique, a trench with a depth deeper than the well is formed in the silicon by anisotropic reactive sputter etching. An oxide layer is thermally grown on the bottom and walls of the trench, which is then refilled by deposited polysilicon or silicon dioxide. This technique can eliminate latchup because the n -channel and p -channel devices are physically isolated by the refilled trench. The detailed steps for trench isolation and some related CMOS processes are discussed next.

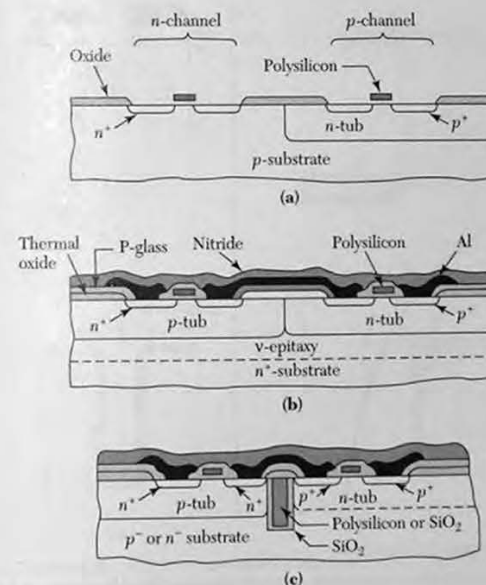


Figure 9.24 Various CMOS structures. (a) n tub. (b) Twin tub.¹ (c) Refilled trench.¹¹

Well Formation Technology

The well of a CMOS can be a single well, a twin well, or a retrograde well. The twin-well process exhibits some disadvantages. For example, it needs high-temperature processing (above 1050°C) and a long diffusion time (longer than 8 hours) to achieve the required well depth of 2 to 3 μm . In this process, the doping concentration is highest at the surface and decreases monotonically with depth.

To reduce the process temperature and time, high-energy implantation is used (i.e., implanting the ion to the desired depth instead of diffusion from the surface). Since the depth is determined by the implantation energy, we can design the well depth with different implantation energy. The profile of the well in this case can have a peak at a certain depth in the silicon substrate. This is called a *retrograde well*. Figure 9.25 shows a comparison of the impurity profiles in the retrograde well and the conventional thermally diffused well.¹² The energy for the *n*- and *p*-type retrograde wells is around 700 keV and 400 keV, respectively. As mentioned previously, the advantage of high-energy implantation is that it can form the well under low-temperature and short-time conditions. Hence, it can reduce the lateral diffusion and increase the device density. The retrograde well offers some additional advantages over the conventional well: (a) Because of high doping near the bottom, the well resistivity is lower than that of the conventional well, and the latchup problem can be minimized; (b) the chanstop can be formed at the same time as the retrograde well implantation, reducing processing steps and time; and (c) higher well doping in the bottom can reduce the chance of punchthrough from the drain to the source.

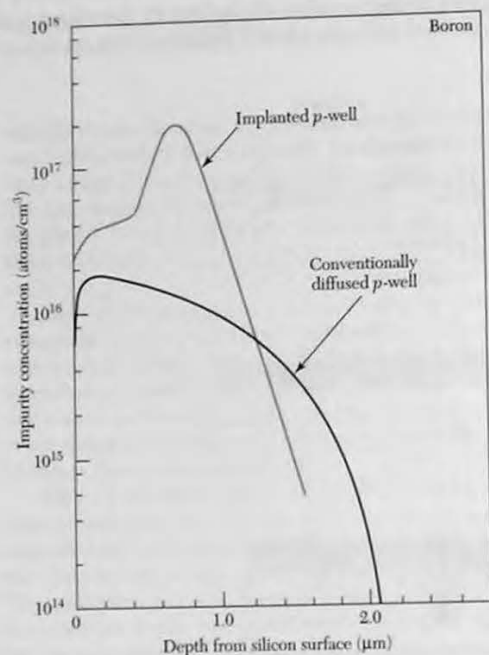


Figure 9.25 Retrograde *p*-well implanted impurity concentration profile. Also shown is a conventionally diffused well.¹²

Advanced Isolation Technology

The conventional isolation process (Section 9.3.1) has some disadvantages that make it unsuitable for deep-submicron (0.25 μm features and smaller) fabrication. The high-temperature oxidation of silicon and long oxidation time result in the encroachment of the chanstop implantation (usually boron for NMOS) to the active region and cause V_T shift. The area of the active region is reduced because of the lateral oxidation. In addition, the field-oxide thickness in submicron isolation spacings is significantly less than the thickness of field oxide grown in wider spacings. Trench isolation technology can avoid these problems and has become the mainstream technology for isolation.

Figure 9.26 shows the process sequence for forming a deep (larger than 3 μm) but narrow (less than 2 μm) trench isolation structure. There are four steps: patterning the area; trench etching and oxide growth; refilling with dielectric materials, such as oxide or undoped polysilicon; and planarization. This deep-trench isolation can be used in both advanced CMOS and bipolar devices and for the trench-type DRAM. Since the isolation material is deposited by CVD, it does not need a long time or a high-temperature process, and it eliminates lateral oxidation and boron encroachment problems.

Another example is shallow-trench (depth less than 1 μm) isolation for CMOS, shown in Figure 9.27. After patterning (Fig. 9.27a), the trench area is etched (Fig. 9.27b) and then refilled with oxide (Fig. 9.27c). Before refilling, a chanstop implantation can be performed. Since the oxide has overfilled the trench, the oxide on the nitride should be removed. Chemical mechanical polishing, which was discussed in Section 8.5.4, is used to remove the oxide on the nitride and to get a flat surface (Fig. 9.27d). Because of its high resistance to polishing, the nitride acts as a stop layer for the CMP process. After

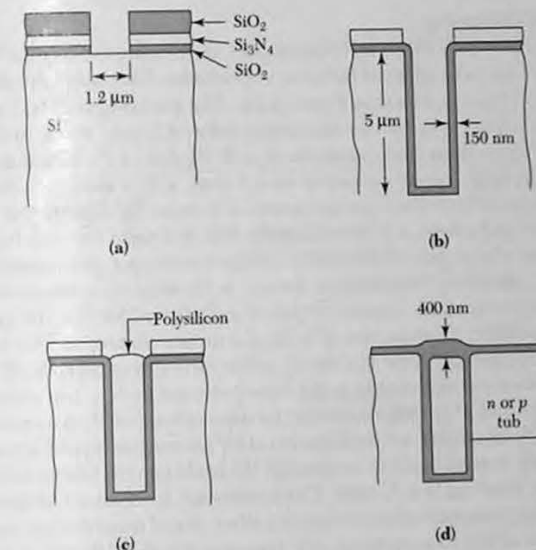


Figure 9.26 Process sequence for forming a deep, narrow-trench isolation structure. (a) Trench mask patterning. (b) Trench etching and oxide growth. (c) Polysilicon deposition to fill the trench. (d) Planarization.

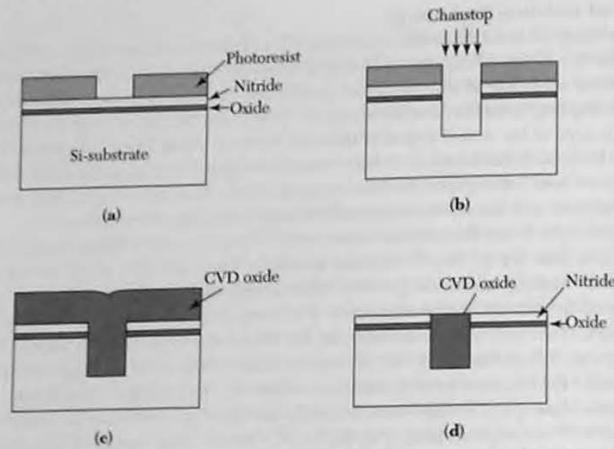


Figure 9.27 Shallow-trench isolation for CMOS. (a) Patterning with photoresist on nitride/oxide films. (b) Dry etching and chanstop implantation. (c) Chemical vapor deposition (CVD) oxide to refill. (d) Planar surface after chemical mechanical polishing.

the polishing, the nitride layer and the oxide layer can be removed by H_3PO_4 and HF, respectively. This initial planarization step is helpful for the subsequent polysilicon patterning and planarizations of the multilevel interconnection processes.

Gate Engineering Technology

If we use n^+ polysilicon for both PMOS and NMOS gates, the threshold voltage for PMOS ($V_{tp} \approx -0.5$ to -1.0 V) has to be adjusted by boron implantation. This makes the channel of the PMOS a buried type, as shown in Figure 9.28a. The buried-type PMOS suffers serious short-channel effects as the device size shrinks below $0.25 \mu\text{m}$. The most noticeable phenomena for short-channel effects are the V_t roll-off, drain-induced barrier lowering (DIBL), and the large leakage current at the off state, so that even with the gate voltage at zero, leakage current flows through source and drain. To alleviate this problem, one can change n^+ polysilicon to p^+ polysilicon for PMOS. Due to the work function difference (1.0 eV from n^+ to p^+ polysilicon), one can obtain a surface p -type channel device without the boron V_t adjustment implantation. Hence, as the technology shrinks to $0.25 \mu\text{m}$ and less, dual-gate structures are required: p^+ polysilicon gate for PMOS, and n^+ polysilicon for NMOS (Fig. 9.28b). A comparison of V_t for the surface channel and the buried channel is shown in Figure 9.29. Note that the V_t of the surface channel rolls off more slowly in the deep-submicron regime than in the buried-channel device. This makes the surface-channel device with p^+ polysilicon suitable for deep-submicron device operation.

To form the p^+ polysilicon gate, ion implantation of BF_2 is commonly used. However, boron penetrates easily from the polysilicon through the oxide into the silicon substrate at high temperatures, resulting in a V_t shift. This penetration is enhanced in the presence of an F atom. There are methods to reduce this effect: use of rapid thermal annealing to reduce the time at high temperatures and, consequently, the diffusion of boron; use of nitrided oxide to suppress the boron penetration, since boron can easily combine with nitrogen and becomes less mobile; and the making of a multilayer of polysilicon to trap the boron atoms at the interface of the two layers.

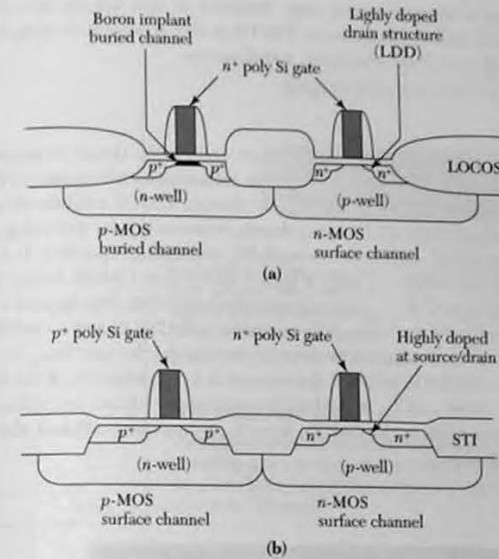


Figure 9.28 (a) A conventional long-channel CMOS structure with a single polysilicon gate (n^+). (b) Advanced CMOS structures with dual polysilicon gates.

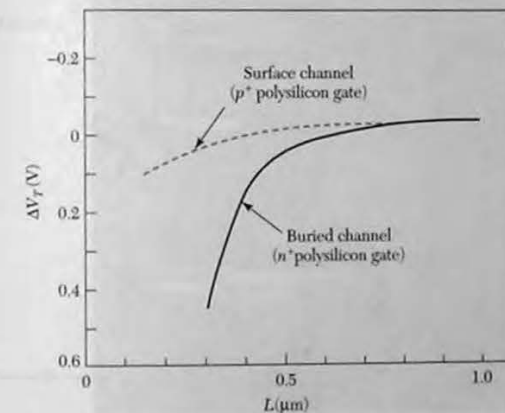


Figure 9.29 The V_t roll-off for a buried-type channel and for a surface-type channel. The V_t drops very quickly as the channel length becomes less than $0.5 \mu\text{m}$.

Figure 9.30 shows a microprocessor chip (Pentium 4) that has an area of about 200 mm^2 and contains 42 million components. This ULSI chip is fabricated using $0.18\text{-}\mu\text{m}$ CMOS technology with a six-level aluminum metallization.

9.3.4 BiCMOS Technology

BiCMOS is a technology that combines both CMOS and bipolar device structures in a single IC. The reason to combine these two different technologies is to create an IC chip that has the advantages of both CMOS and bipolar devices. CMOS exhibits advantages in power dissipation, noise margin, and packing density; whereas bipolar technology shows advantages in switching speed, current drive capability, and analog capability. As a result, for a given design rule, BiCMOS can have a higher speed than CMOS, better performance in analog circuits than CMOS, a lower power dissipation than bipolar, and a higher component density than bipolar. Figure 9.31 compares a BiCMOS and a CMOS logic gate. For a CMOS inverter, the current to drive (or to charge) the next load, C_L , is the drain current I_{DS} . For a BiCMOS inverter, the current is $h_{fe}I_{DS}$, where h_{fe} is the current gain of the bipolar transistor and I_{DS} is the base current of the bipolar transistor and is equal to the drain current of M_2 in the CMOS. Since h_{fe} is much larger than 1, the speed can be substantially enhanced.

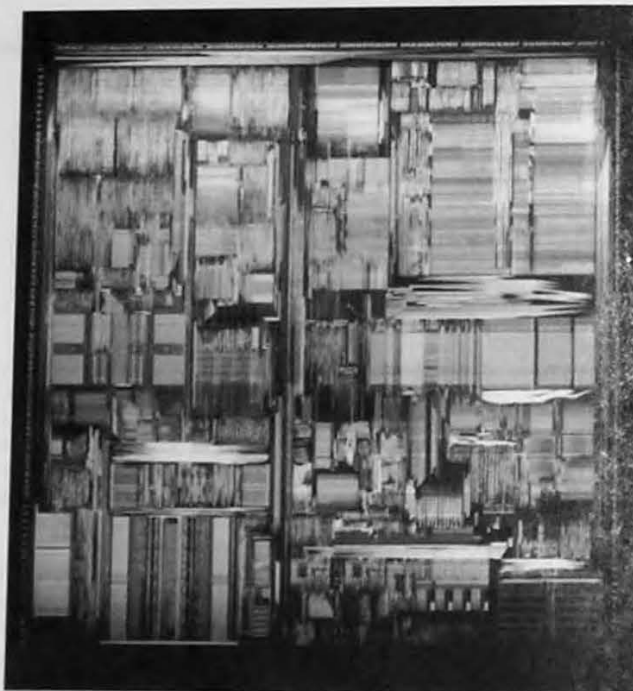


Figure 9.30 Micrograph of a 32-bit Pentium 4 microprocessor chip. (Photograph courtesy of Intel Corporation.)

BiCMOS has been widely used in many applications. In the early days, it was used in SRAM circuits. At the present time, BiCMOS technology has been successfully developed for transceiver, amplifier, and oscillator applications in wireless communication equipment. Most of the BiCMOS processes are based on the CMOS process, with some modifications, such as adding masks for bipolar transistor fabrication. The following example (Fig. 9.32) is for a high-performance BiCMOS process based on the twin-well CMOS process.¹³

The initial material is a p -type silicon substrate. An n^+ buried layer is then formed to reduce collector resistance. The buried p layer is formed by ion implantation to increase the doping level to prevent punchthrough. A lightly doped n -epi layer is grown on the wafer, and a twin-well process for the CMOS is performed. To achieve high performance for the bipolar transistor, four additional masks are needed. They are the buried n^+ mask, the collector deep n^+ mask, the base p mask, and the poly-emitter mask. In other processing steps, the p^+ region for base contact can be formed with the p^+ implant in the source/drain implantation of the PMOS, and the n^+ emitter can be formed with the source/drain implantation of the NMOS. The additional masks and longer processing time compared with a standard CMOS are the main drawbacks of BiCMOS. The additional cost should be justified by the enhanced performance.

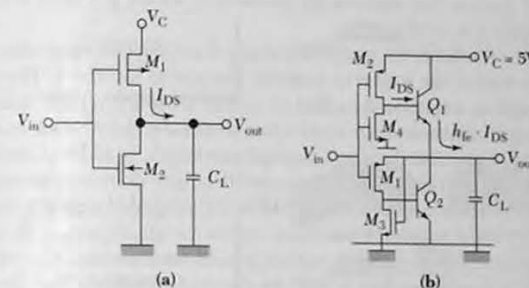


Figure 9.31 (a) CMOS logic gate. (b) Bipolar CMOS (BiCMOS) logic gate.

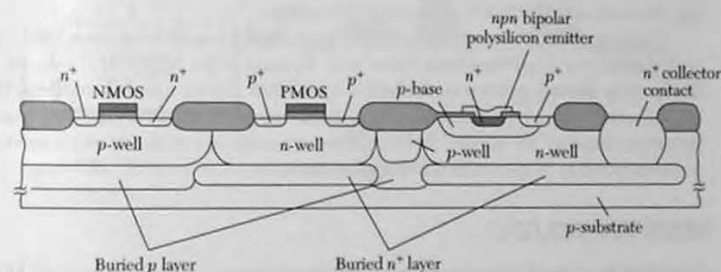


Figure 9.32 Optimized BiCMOS device structure. Key features include self-aligned p and n^+ buried layers for improved packing density, separately optimized n and p wells (twin-well CMOS) formed in an epitaxial layer with intrinsic background doping, and a polysilicon emitter for improved bipolar performance.¹³

9.4 MESFET TECHNOLOGY

Advances in gallium arsenide processing techniques, in conjunction with new fabrication and circuit approaches, have made possible the development of "silicon-like" gallium arsenide IC technology. Gallium arsenide has three inherent advantages compared with silicon: higher electron mobility, which results in lower series resistance for a given device geometry; higher drift velocity at a given electric field, which improves device speed; and the ability to be made semiinsulating, which can provide a lattice-matched dielectric-insulated substrate. However, gallium arsenide also has three disadvantages: a very short minority-carrier lifetime; lack of a stable, passivating native oxide; and crystal defects that are many orders of magnitude higher than in silicon. The short minority-carrier lifetime and the lack of high-quality insulating films have prevented the development of bipolar devices and delayed MOS technology using gallium arsenide. Thus, the emphasis of gallium arsenide IC technology is in the MESFET area, in which our main concerns are the majority carriers' transport and the metal-semiconductor contact.

A typical fabrication sequence¹⁴ for a high-performance MESFET is shown in Figure 9.33. A layer of GaAs is epitaxially grown on a semiinsulating GaAs substrate, followed by an n^+ contact layer (Fig. 9.33a). A mesa etch step is performed for isolation (Fig. 9.33b), and a metal layer is evaporated for the source and drain ohmic contacts (Fig. 9.33c). A channel recess etch is followed by a gate recess etch and gate evaporation (Figs. 9.33d and e). After a liftoff process that removes the photoresist, shown in Figure 9.33e, the MESFET is completed (Fig. 9.33f).

The n^+ contact layer reduces the source and drain ohmic contact resistances. Note that the gate is offset toward the source to minimize the source resistance. The epitaxial layer is thick enough to minimize the effect of surface depletion on the source and drain resistance. The gate electrode has maximal cross-sectional area with a minimal footprint, which provides low gate resistance and minimal gate length. In addition, the length (L_{GD}) is designed to be greater than the depletion width at gate-drain breakdown.

A representative fabrication sequence for a MESFET integrated circuit is shown in Figure 9.34.¹⁵ In this process, n^+ source and drain regions are self-aligned to the gate of each MESFET. A relatively light channel implant is used for the enhancement-mode switching device, and a heavier implant is used for the depletion-mode load device. A gate recess is usually not used for such digital IC fabrication because the uniformity of each depth is difficult to control, leading to unacceptable variation in the threshold voltage. This process sequence can also be used for a *monolithic microwave integrated circuit* (MMIC). Note that gallium arsenide MESFET processing technology is similar to the silicon-based MOSFET processing technology.

Gallium arsenide ICs with complexities up to the large-scale integration level (~10,000 components per chip) have been fabricated. Because of the higher drift velocity (~20% higher than silicon), gallium arsenide ICs have a 20% higher speed than silicon ICs that use the same design rules. However, substantial improvements in crystal quality and processing technology are needed before gallium arsenide can seriously challenge the pre-eminent position of silicon in ULSI applications.

9.5 MEMS TECHNOLOGY

There has been rapidly growing interest in *microelectromechanical systems* (MEMS) since the late 1980s, when a spinning micromotor made from polysilicon was fabricated on a silicon chip.^{16,17} The manufacture of silicon MEMS adopted many of the highly developed technologies of silicon integrated circuits. This approach has enabled MEMS products to

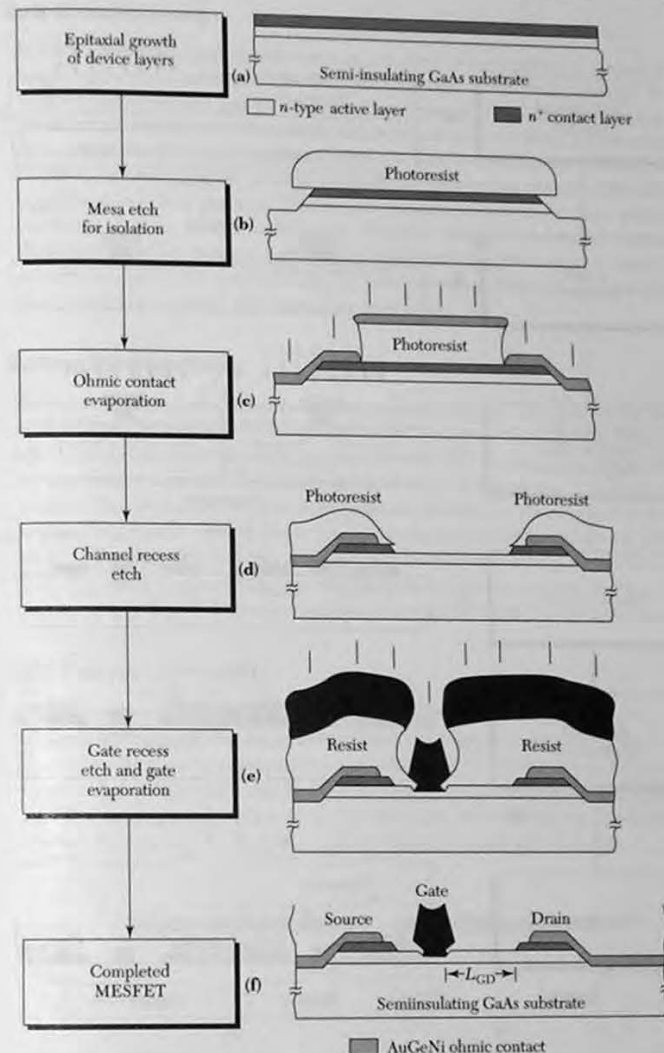


Figure 9.33 Fabrication sequence of a GaAs MESFET.¹⁴

be produced at low cost by using batch fabrication, just as for ICs. In addition to the IC fabrication process, some specialized techniques have been developed for MEMS. This section considers three specialized etching techniques: bulk micromachining, surface micromachining, and the LIGA process.

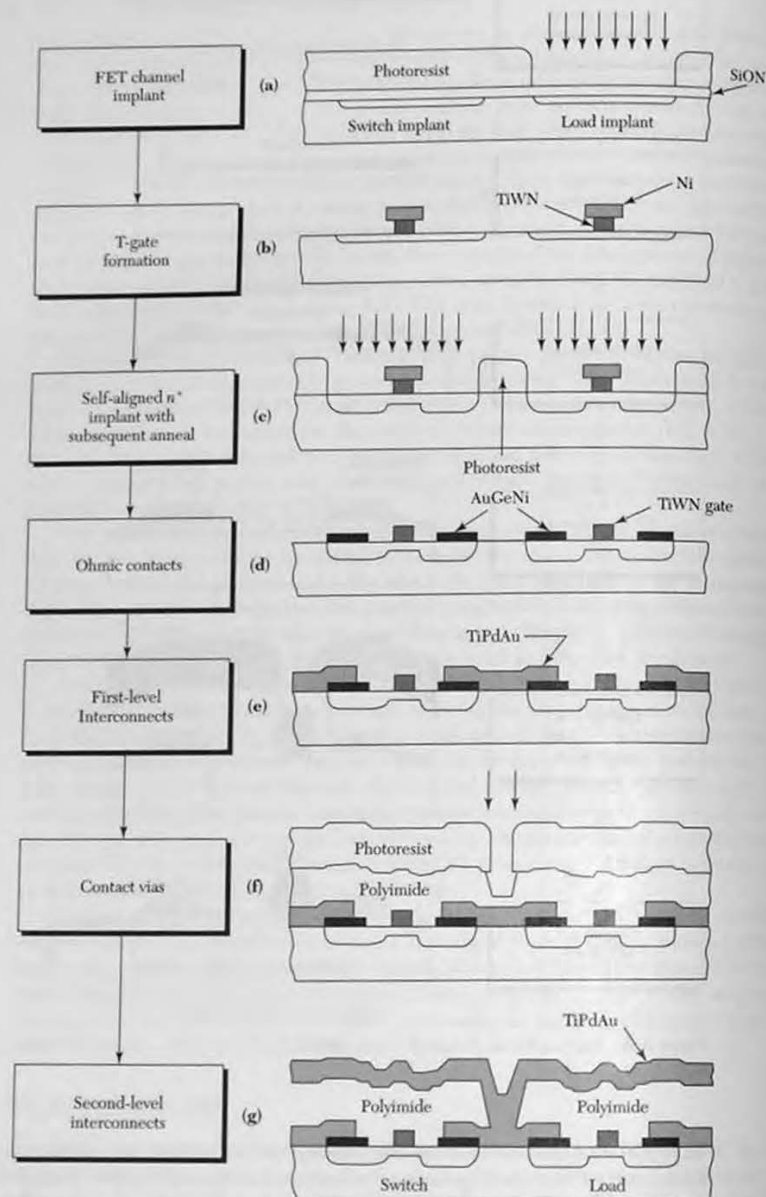


Figure 9.34 Fabrication process for MESFET direct-coupled FET logic (DCFL) with active loads. Note that the n^+ source and drain regions are self-aligned to the gate.¹⁵

9.5.1 Bulk Micromachining

In bulk micromachining, the device (e.g., a sensor or an actuator) is shaped by etching a large single-crystal substrate. The films are patterned on the bulk substrate to define the isolation and transducer functions. Orientation-dependent wet chemical etching techniques provide high-resolution etch and tight dimensional control. Often, a bulk-micromachined device uses two-sided processing, creating a self-isolated structure with one side exposed to the measured variables, such as mechanical or chemical signals, while the other side is enclosed in a clean package. Two-sided structures are very robust for operation in environments hostile to microelectronic devices. Simple mechanical devices such as diaphragm pressure sensors, membranes, and cantilever-beam piezoresistive acceleration sensors are fabricated commercially by this technique. Figure 9.35 illustrates a fabrication process of a simple silicone rubber membrane.¹⁶

9.5.2 Surface Micromachining

Surface-micromachined devices are constructed entirely from thin films. There are several differences and trade-offs between structures made from bulk and thin film materials. Typical dimensions for bulk-micromachined sensors are in the millimeter range, but surface-micromachined devices are of micrometer dimensions. Surface micromachining permits the fabrication of structurally complex devices by stacking and patterning layers or "building blocks" of thin films, whereas multilayered bulk devices are difficult to construct. Free-standing and movable parts can be fabricated using sacrificial layers. Figure 9.36 illustrates how sacrificial etching techniques can be used to create an electrostatic micromotor with well-defined, submicron tolerance between the rotor and the center hub.¹⁷

9.5.3 LIGA Process

LIGA is a German acronym for *lithographic, galvanofarming, abforming*.¹⁸ It consists of three basic processing steps: lithography, electroplating, and molding. The LIGA process is based on x-ray radiation from a synchrotron. The process can produce microstructures with lateral dimensions in the micrometer range and structural heights of several hundred micrometers from a variety of materials. Its potential applications cover microelectronics, sensors, microoptics, micromechanics, and biotechnology.

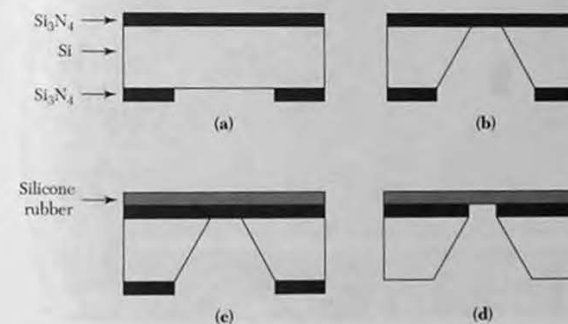


Figure 9.35 Fabrication process of simple silicone rubber membrane. (a) Nitride deposition and patterning. (b) KOH etching. (c) Silicone rubber spin coating. (d) Nitride removal on back side.¹⁹

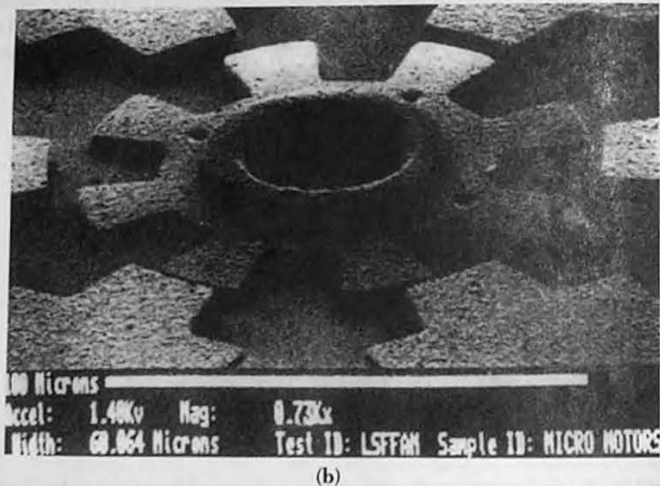
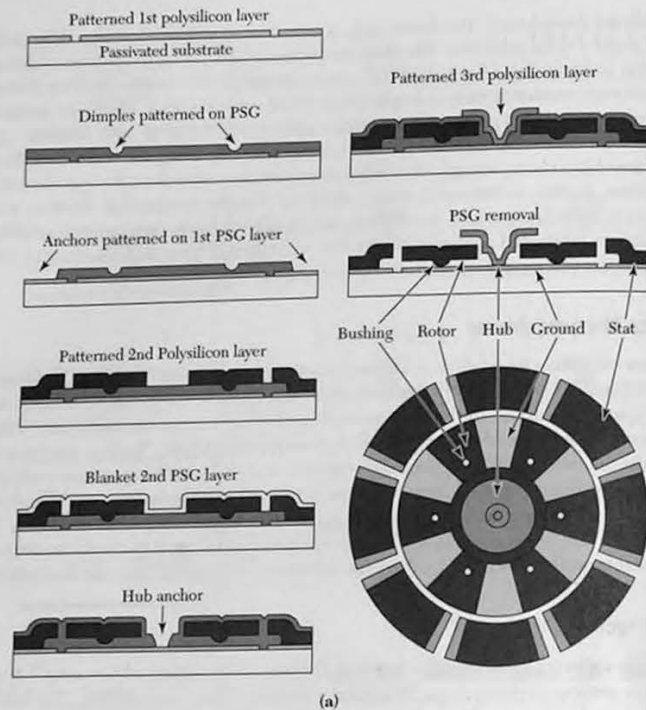


Figure 9.36 (a) Sacrificial process flow for an electrostatic micromotor. PSG, phosphosilicate glass. (b) Photograph of a micromotor.¹⁷

An example of the LIGA process is shown in Figure 9.37. An x-ray resist, ranging from 300 μm to more than 500 μm in thickness, is deposited on a substrate with an electrically conductive surface. Lithographic patterning is done with extended exposure from highly collimated x-ray radiation through an x-ray mask, as shown Figure 9.37a. A flower-shaped trench structure is formed in the thick resist after developer treatment (Fig. 9.37b).

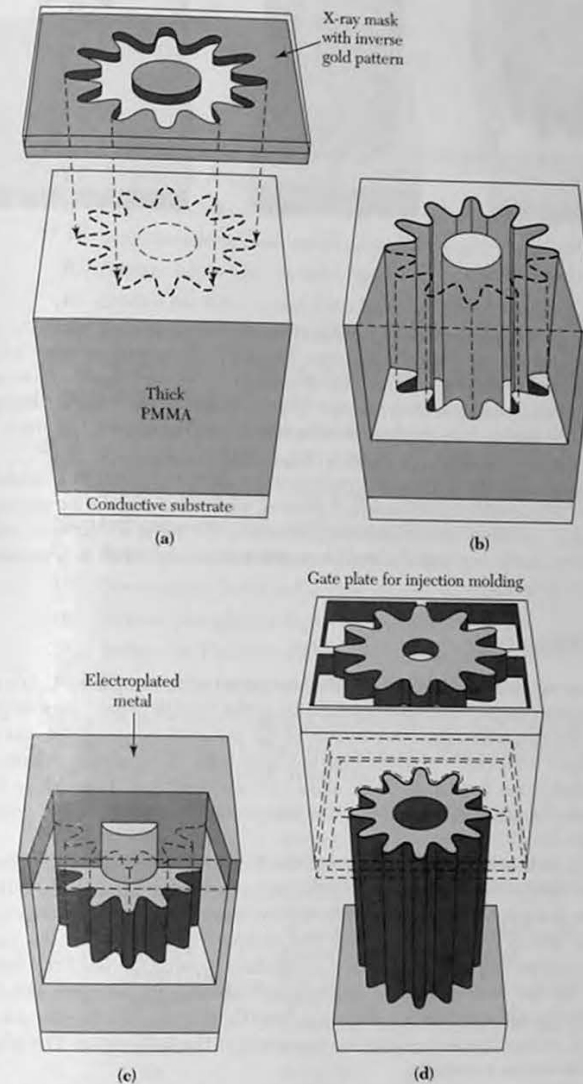


Figure 9.37 The LIGA process.¹⁹

(continued)

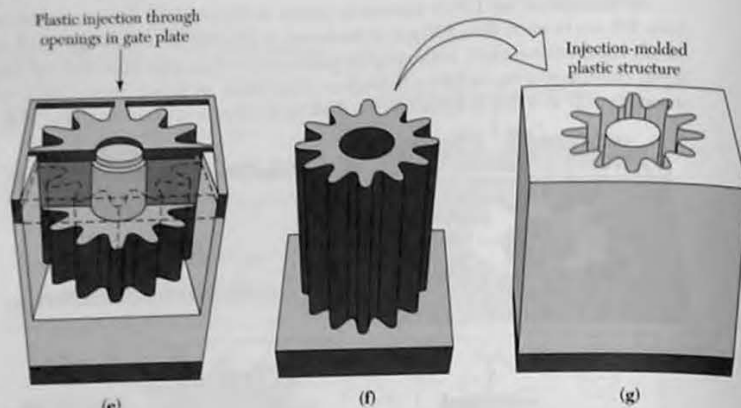


Figure 9.37 (continued)

Metal is then electroplated on the exposed bottom conductive surface, filling the trench space and covering the top surface of the resist (Fig. 9.37e). The metal structure is formed after removing the resist (Fig. 9.37d). This structure can be used repeatedly as a mold insert for injection molding to form multiple plastic replicas of the original plating base (Fig. 9.37e). The plating base replicas, in turn, can be used to electroplate many metal structures as the final products, as shown in Figures 9.37f and g.

The distinct advantage of the LIGA process is the ability to create three-dimensional structures as thick as bulk-micromachined devices, while retaining the same degree of design flexibility as surface micromachining. However, the initial synchrotron radiation process is a very costly step, and the mold separation steps may result in degradation of the original mold insert.

9.6 PROCESS SIMULATION

SUPREM can be extremely useful for the simulation of a complete IC fabrication sequence. As an example,²⁰ consider a simulation of the NMOS polysilicon gate process described in Section 9.3.1. The cross section of the device to be simulated was shown initially in Figure 9.17c and is repeated here in Figure 9.38. Three vertical regions of the device, denoted by cut lines A-A', B-B', and C-C', are simulated. These three simulations represent the center of the device, the source/drain region, and the field region, respectively.

The structure is simulated using a total of five SUPREM input decks. The first deck simulates processing in the active region of the device, up to the point at which the process sequence diverges for the gate and source/drain regions. The second and third decks start with the results from the first deck and complete processing for the gate and source/drain regions, respectively. This is accomplished by using the SAVEFILE statement at the end of the first deck to save the structure and subsequently using the saved structure in the INITIALIZE statements in the second and third decks. The fourth deck is similar to the first, except that it simulates the processing in the field region. The fifth deck completes field region processing.

The complete process sequence is as follows:

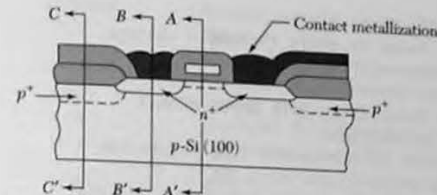


Figure 9.38 NMOS device for SUPREM simulation.

1. Begin with a high-resistivity <100> *p*-type silicon substrate.
2. Grow a 400-Å SiO₂ pad layer.
3. Deposit an 800-Å silicon nitride layer on top of the pad oxide.
4. Strip the nitride from areas outside of the active region.
5. Implant boron into the field regions.
6. Oxidize the field regions for 3 hours at 1000°C in wet O₂.
7. Etch down to the silicon in the active region.
8. Implant boron to set the threshold voltage of the MOSFET.
9. Grow a 400-Å gate oxide.
10. Deposit 0.5 μm of polysilicon.
11. Dope the polysilicon with phosphorus using POCl₃.
12. Etch the polysilicon in the areas outside the gate region.
13. Implant arsenic to form the source/drain regions.
14. Drive in the source/drain As implant for 10 minutes at 1000°C in dry O₂.
15. Open contact holes in the gate, source, and drain regions.
16. Deposit phosphorus-doped SiO₂ (P-glass) over the wafer surface.
17. Reflow the P-glass at 1000°C for 30 minutes.
18. Reopen the contact holes, and deposit aluminum.

Plots of the doping profiles in the gate (section A-A'), source/drain (section B-B'), and field (section C-C') regions are shown in Figures 9.39, 9.40, and 9.41, respectively. The SUPREM input listing is as follows:

```

TITLE      NMOS Polysilicon Gate-Deck 1
COMMENT    Active device region initial processing
COMMENT    Initialize silicon substrate
INITIALIZE <100> Silicon Boron Concentration=1e15
COMMENT    Grow 400A pad oxide
DIFFUSION  Time=40 Temperature=1000 DryO2
COMMENT    Deposit 800A CVD nitride
DEPOSITION Nitride Thickness=0.08
COMMENT    Grow field oxide
DIFFUSION  Time=180 Temperature=1000 WetO2
COMMENT    Etch to silicon surface
ETCH       Oxide all
ETCH       Nitride all
  
```

```

ETCH      Oxide all
COMMENT   Implant boron to shift threshold voltage
IMPLANT   Boron Dose=4e11 Energy=50
COMMENT   Grow gate oxide
DIFFUSION Time=30 Temperature=1050 DryO2 HCl%=3
COMMENT   Deposit polysilicon
DEPOSITION Polysilicon Thickness=0.5 Temperature=600
COMMENT   Dope the polysilicon using POC13
DIFFUSION Time=25 Temperature=1000 Phosphorus solidsol
PRINT     Layers
PLOT      Chemical Boron Phosphor Net
SAVEFILE  Structur Filename=nmosactiveinit.str
STOP      End Deck 1

TITLE     NMOS Polysilicon Gate-Deck 2
COMMENT   Gate region
COMMENT   Initialize silicon substrate
INITIALIZE Structur=nmosactiveinit.str
COMMENT   Implant arsenic for source/drain regions
IMPLANT   Arsenic Dose=5e15 Energy=150
COMMENT   Drive-in arsenic and re-oxidize source/drain regions
DIFFUSION Time=30 Temperature=1000 DryO2
COMMENT   Etch contact holes to gate, source, and drain regions
ETCH      Oxide
COMMENT   Deposit phosphorus-doped SiO2 using CVD
DEPOSITION Oxide Thickness=0.75 C.phosphor=1e21
COMMENT   Reopen contact holes
ETCH      Oxide
COMMENT   Deposit Aluminum
DEPOSITION Aluminum Thickness=1.2
PRINT     Layers
PLOT      Chemical Boron Arsenic Phosphor Net
STOP      End Deck 2

TITLE     NMOS Polysilicon Gate-Deck 3
COMMENT   Source/drain regions
COMMENT   Initialize silicon substrate
INITIALIZE Structur=nmosactiveinit.str
COMMENT   Etch polysilicon and oxide over source/drain regions
ETCH      Polysilicon
ETCH      Oxide
COMMENT   Implant arsenic for source/drain regions
IMPLANT   Arsenic Dose=5e15 Energy=150
COMMENT   Drive-in arsenic and re-oxidize source/drain regions
DIFFUSION Time=30 Temperature=1000 DryO2
COMMENT   Etch contact holes to gate, source, and drain regions
ETCH      Oxide
COMMENT   Deposit phosphorus-doped SiO2 using CVD
DEPOSITION Oxide Thickness=0.75 C.phosphor=1e21
COMMENT   Reflow glass to smooth surface and dope contact holes
DIFFUSION Time=30 Temperature=1000

```

```

COMMENT   Reopen contact holes
ETCH      Oxide
COMMENT   Deposit Aluminum
DEPOSITION Aluminum Thickness=1.2
PRINT     Layers
PLOT      Chemical Boron Arsenic Phosphor Net
STOP      End Deck 3

TITLE     NMOS Polysilicon Gate-Deck 4
COMMENT   Isolation region initial processing
COMMENT   Initialize silicon substrate
INITIALIZE <100> Silicon Boron Concentration=1e15
COMMENT   Grow 400A pad oxide
DIFFUSION Time=40 Temperature=1000 DryO2
COMMENT   Implant boron to increase field doping
IMPLANT   Boron Dose=1e13 Energy=150
COMMENT   Grow field oxide
DIFFUSION Time=180 Temperature=1000 WetO2
COMMENT   Implant boron to shift threshold voltage
IMPLANT   Boron Dose=4e11 Energy=50
COMMENT   Grow gate oxide
DIFFUSION Time=30 Temperature=1050 DryO2 HCl%=3
COMMENT   Deposit polysilicon
DEPOSITION Polysilicon Thickness=0.5 Temperature=600
COMMENT   Dope the polysilicon using POC13
DIFFUSION Time=25 Temperature=1000 Phosphorus solidsol
PRINT     Layers
PLOT      Chemical Boron Phosphor Net
SAVEFILE  Structur Filename=nmosfieldinit.str
STOP      End Deck 4

TITLE     NMOS Polysilicon Gate-Deck 5
COMMENT   Isolation region final processing
COMMENT   Initialize silicon substrate
INITIALIZE Structur=nmosfieldinit.str
COMMENT   Etch polysilicon and oxide over source/drain regions
ETCH      Polysilicon
ETCH      Oxide Thickness=0.07
COMMENT   Implant arsenic for source/drain regions
IMPLANT   Arsenic Dose=5e15 Energy=150
COMMENT   Drive-in arsenic and re-oxidize source/drain regions
DIFFUSION Time=30 Temperature=1000 DryO2
COMMENT   Deposit phosphorus-doped SiO2 using CVD
DEPOSITION Oxide Thickness=0.75 C.phosphor=1e21
COMMENT   Reflow glass to smooth surface and dope contact holes
DIFFUSION Time=30 Temperature=1000
COMMENT   Deposit Aluminum
DEPOSITION Aluminum Thickness=1.2
PRINT     Layers
PLOT      Chemical Boron Arsenic Phosphor Net
STOP      End Deck 5

```

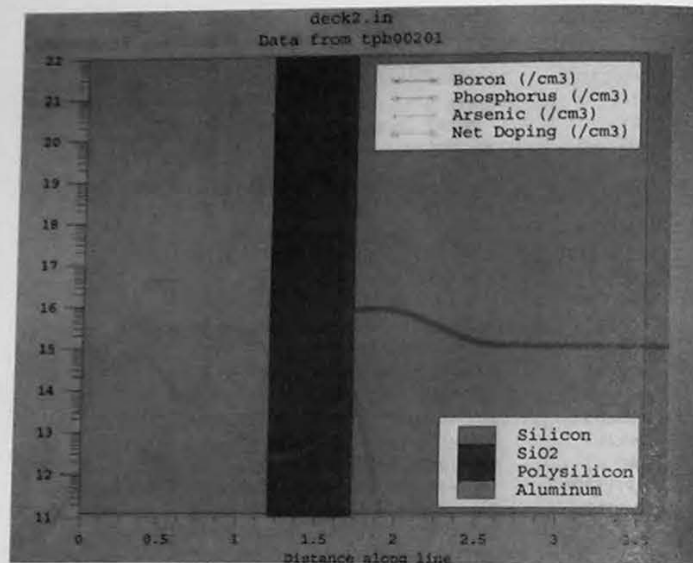


Figure 9.39 Plot of the doping profile in the gate region.

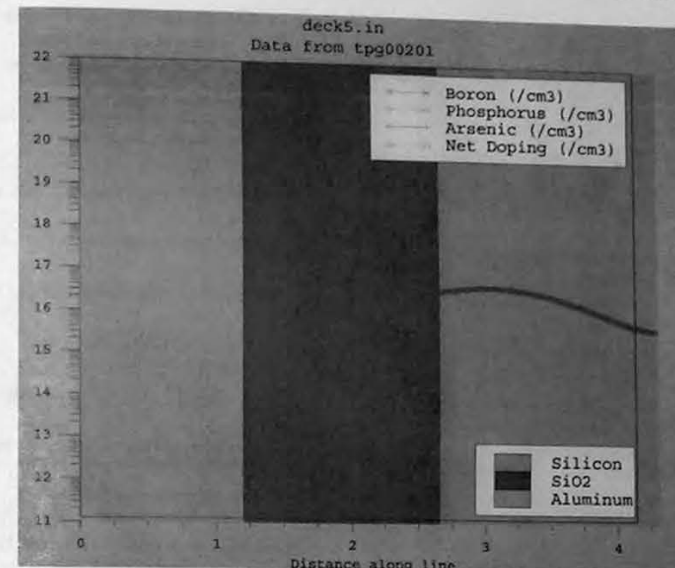


Figure 9.41 Plot of the doping profile in the field region.

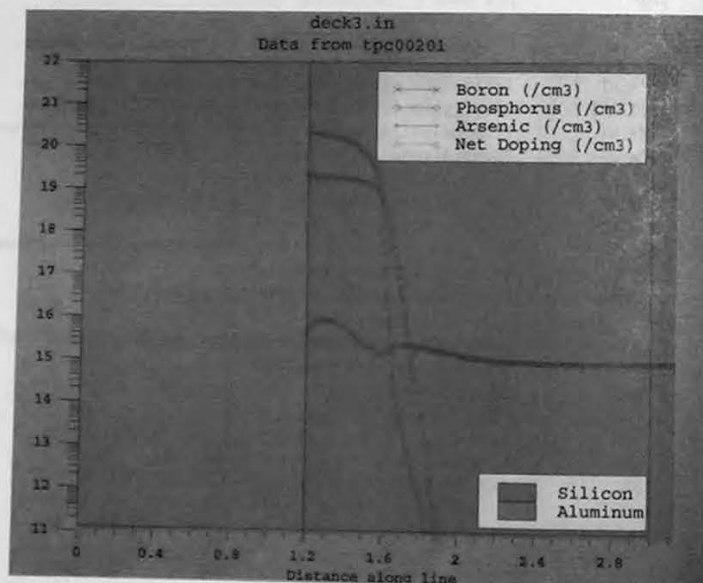


Figure 9.40 Plot of the doping profile in the source/drain region.

► 9.7 SUMMARY

This chapter considered processing technologies for passive components, active devices, ICs, and MEMS. Three major IC technologies based on the bipolar transistor, the MOSFET, and the MESFET were discussed in detail. It appears that the MOSFET will be the dominant technology at least until 2014 because of its superior performance compared with the bipolar transistor. For 100-nm CMOS technology, a good candidate is the combination of an SOI substrate with interconnections using Cu and low- k materials.

MEMS is still an emerging field. MEMS has adopted the lithographic and etching technologies from IC fabrication. Specialized etching techniques have also been developed for MEMS: bulk micromachining using an orientation-dependent etching process, surface micromachining using sacrificial layers, and the LIGA process using x-ray lithography with highly collimated radiation.

► REFERENCES

1. For a detailed discussion on IC process integration, see C. Y. Liu and W. Y. Lee, "Process Integration," in C. Y. Chang and S. M. Sze, Eds., *ULSI Technology*, McGraw-Hill, New York, 1996.
2. T. Tachikawa, "Assembly and Packaging," in C. Y. Chang and S. M. Sze, Eds., *ULSI Technology*, McGraw-Hill, New York, 1996.
3. T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, Cambridge, U.K., 1998, Ch. 2.
4. D. Rise, "Isoplanar-S Scales Down for New Heights in Performance," *Electronics*, **53**, 137 (1979).
5. T. C. Chen, et al., "A Submicrometer High-Performance Bipolar Technology," *IEEE Electron. Device Lett.*, **10**(8), 364 (1989).

6. G. P. Li et al., "An Advanced High-Performance Trench-Isolated Self-Aligned Bipolar Technology," *IEEE Trans. Electron Devices*, **34**(10), 2246 (1987).
7. W. E. Beasley, J. C. C. Tsai, and R. D. Plummer, Eds., *Quick Reference Manual for Semiconductor Engineering*, Wiley, New York, 1985.
8. R. W. Hunt, "Memory Design and Technology," in M. J. Howes and D. V. Morgan, Eds., *Large Scale Integration*, Wiley, New York, 1981.
9. A. K. Sharma, *Semiconductor Memories—Technology, Testing, and Reliability*, IEEE, New York, 1997.
10. U. Hamann, "Chip Cards—The Application Revolution," *IEEE Tech. Dig. Int. Electron Devices Meet.*, p. 15 (1997).
11. R. D. Rung, H. Momose, and Y. Nagakubo, "Deep Trench Isolation CMOS Devices," *IEEE Tech. Dig. Int. Electron. Devices Meet.*, p. 237 (1982).
12. D. M. Bron, M. Ghezzi, and J. M. Pringley, "Trends in Advanced CMOS Process Technology," *Proc. IEEE*, p. 1646 (1986).
13. H. Higuchi, et al., "Performance and Structure of Scaled-Down Bipolar Devices Merge with CMOSFETs," *IEEE Tech. Dig. Int. Electron. Devices Meet.*, p. 694 (1984).
14. M. A. Hollis and R. A. Murphy, "Homogeneous Field-Effect Transistors," in S. M. Sze, Ed., *High-Speed Semiconductor Devices*, Wiley, New York, 1990.
15. H. P. Singh, et al., "GaAs Low Power Integrated Circuits for a High Speed Digital Signal Processor," *IEEE Trans. Electron Devices*, **36**, 240 (1989).
16. C. H. Mastrangelo and W. C. Tang, "Semiconductor Sensor Technology," in S. M. Sze, Ed., *Semiconductor Sensors*, Wiley, New York, 1994.
17. L. S. Fan, Y. C. Tai, and R. S. Muller, "IC-Processed Electrostatic Micromotors," in *IEEE Int. Electron Devices Meet.*, p. 666 (1988).
18. X. Yang, et al., "A MEMS Thermopneumatic Silicone Rubber Membrane Valve," *Sens. Actuators*, **A64**, 101 (1998).
19. W. Ehrfeld, et al., "Fabrication of Microstructures Using the LIGA Process," *Proc. IEEE Micro Robots and Teleoperators Workshop*, Hyannis, MA, Nov. 1987.
20. C. P. Ho and S. E. Hansen, *SUPREM III User's Manual*, Stanford University, 1983.

PROBLEMS

Asterisks denote difficult problems

SECTION 9.1: PASSIVE COMPONENTS

1. For a sheet resistance of $1 \text{ k}\Omega/\square$, find the maximum resistance that can be fabricated on a $2.5 \times 2.5 \text{ mm}$ chip using $2\text{-}\mu\text{m}$ lines with a $4\text{-}\mu\text{m}$ pitch (i.e., distance between the centers of the parallel lines).
2. Design a mask set for a 5-pF MOS capacitor. The oxide thickness is 30 nm . Assume that the minimum window size is $2 \times 10 \text{ }\mu\text{m}$ and the maximum registration errors are $2 \text{ }\mu\text{m}$.
3. Draw a complete step-by-step set of masks for the spiral inductor with three turns on a substrate.
4. Design a 10-nH square spiral inductor in which the total length of the interconnect is $350 \text{ }\mu\text{m}$; the spacing between turns is $2 \text{ }\mu\text{m}$.

SECTION 9.2: BIPOLAR TECHNOLOGY

5. Draw the circuit diagram and device cross section of a clamped transistor.
6. Identify the purpose of the following steps in self-aligned double-polysilicon bipolar structure: (a) undoped polysilicon in trench in Figure 9.13a, (b) the poly 1 in Figure 9.13b, and (c) the poly 2 in Figure 9.13d.

SECTION 9.3: MOSFET TECHNOLOGY

- *7. In NMOS processing, the starting material is a p -type $10\text{-}\Omega\text{-cm}$ $\langle 100 \rangle$ -oriented silicon wafer. The source and drain are formed by arsenic implantation of $10^{15} \text{ ions/cm}^2$ at 30 keV through a gate oxide of 25 nm . (a) Estimate the threshold voltage change of the device. (b) Draw the doping profile along a coordinate perpendicular to the surface and passing through the channel region or the source region.
8. (a) Why is $\langle 100 \rangle$ orientation preferred in NMOS fabrication? (b) What are the disadvantages if too thin a field oxide is used in NMOS devices? (c) What problems occur if a polysilicon gate is used for gate lengths less than $3 \text{ }\mu\text{m}$? Can another material be substituted for polysilicon? (d) How is a self-aligned gate obtained, and what are its advantages? (e) What purpose does P-glass serve?
- *9. For a floating-gate nonvolatile memory, the lower insulator has a dielectric constant of 4 and is 10 nm thick. The insulator above the floating gate has a dielectric constant of 10 and is 100 nm thick. If the current density J in the lower insulators is given by $J = \sigma E$, where $\sigma = 10^{-7} \text{ S/cm}$, and the current in the other insulator is negligibly small, find the threshold voltage shift of the device caused by a voltage of 10 V applied to the control gate for (a) $0.25 \text{ }\mu\text{s}$, and (b) a sufficiently long time that J in the lower insulator becomes negligibly small.
10. Draw a complete step-by-step set of masks for the CMOS inverter shown in Figure 9.23. Pay particular attention to the cross section shown in Figure 9.23c for your scale.
- *11. A $0.5\text{-}\mu\text{m}$ digital CMOS technology has $5\text{-}\mu\text{m}$ -wide transistors. The minimum wire width is $1 \text{ }\mu\text{m}$, and the metallization layer consists of $1\text{-}\mu\text{m}$ -thick aluminum. Assume that μ_n is $400 \text{ cm}^2/\text{V}\cdot\text{s}$, d is 10 nm , V_{DD} is 3.3 V , and the threshold voltage is 0.6 V . Finally, assume that the maximum voltage drop that can be tolerated is 0.1 V when a $1 \text{ }\mu\text{m}^2$ cross section aluminum wire is carrying the maximum current that can be supplied by the NMOS transistor. How long a wire can be allowed? Use a simple square-law, long-channel model to predict the MOS current drive (resistivity of aluminum is $2.7 \times 10^{-8} \text{ }\Omega\text{-cm}$).
12. Plot the cross-sectional views of a twin-tub CMOS structure of the following stages of processing: (a) n -tub implant, (b) p -tub implant, (c) twin-tub drive-in, (d) nonselective p^+ source/drain implant, (e) selective n^+ source/drain implant using photoresist as mask, and (f) P-glass deposition.
13. Why do we use a p^+ polysilicon gate for PMOS?
14. What is the boron penetration problem in p^+ polysilicon PMOS? How would you eliminate it?
15. To obtain a good interfacial property, a buffered layer is usually deposited between the high- k material and substrate. Calculate the effective oxide thickness if the stacked gate dielectric structure is (a) a buffered nitride of 0.5 nm and (b) a Ta_2O_5 of 10 nm .
16. Describe the disadvantages of LOCOS technology and the advantages of shallow-trench isolation technology.

SECTION 9.4: MESFET TECHNOLOGY

17. What is the purpose of the polyimide used in Figure 9.34f?
18. What is the reason that it is difficult to make bipolar transistors and MOSFETs in GaAs?

SECTION 9.6: PROCESS SIMULATION

- *19. Use SUPREM to simulate the bipolar process described in Section 9.2.1. Plot the doping profiles along the following vertical cross sections: (a) from the top of the base contact, (b) from the top of the emitter contact, and (c) from the top of the collector contact.
- *20. Use SUPREM to simulate the CMOS process described by Figure 9.23. Plot the doping profiles along the following vertical cross sections: (a) through the PMOS source/drain regions, (b) through the PMOS gate region, (c) through the NMOS source/drain regions, and (d) through the NMOS gate region.

IC Manufacturing

Manufacturing is defined as the process by which raw materials are converted into finished products. As illustrated in Figure 10.1, a manufacturing operation can be viewed graphically as a system with raw materials and supplies serving as its inputs and finished commercial products serving as outputs. In integrated circuit manufacturing, input materials include semiconductor wafers, insulators, dopants, and metals. The outputs are the ICs themselves. The types of processes that arise in IC manufacturing, which have been the subjects of previous chapters in this text, include oxidation and deposition processes, photolithography, etching, and doping (implantation and/or diffusion).

However, before finished ICs can be put to their intended use in various commercial electronic systems and products (such as computers, cellular phones, and digital cameras), several other key processes must take place. These include electrical testing and packaging. Testing is necessary to yield high-quality products. Quality requires conformance of all products to a set of specifications and the reduction of any variability in the manufacturing process. Maintaining quality often involves the use of statistical process control. A designed experiment is an extremely useful tool for discovering key variables that influence quality characteristics. Statistical experimental design is a powerful approach for systematically varying controllable process conditions and determining their impact on output parameters that measure quality.

A key metric that can be used to evaluate any manufacturing process is cost, and cost is directly impacted by yield. *Yield* refers to the proportion of manufactured products that perform as required by a set of specifications. Yield is inversely proportional to the total manufacturing cost: The higher the yield, the lower the cost. Finally, computer-integrated manufacturing is aimed at optimizing the cost-effectiveness of electronics manufacturing by using the latest developments in computer hardware and software technology to enhance expensive manufacturing methods.

This chapter describes each of these concepts. Specifically, it discusses the following topics:

- Electrical testing and test structures
- Electronics packaging processes
- Statistical process control and experimental design in the context of IC manufacturing

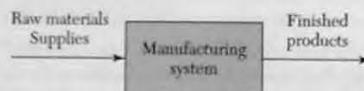


Figure 10.1 Block diagram of a manufacturing system.

- IC yield and various yield models
- Computer-integrated manufacturing systems

Viewed from a systems-level perspective, IC manufacturing intersects with nearly all parts of the production process, including design, fabrication, integration, assembly, testing, and packaging. The end result is an electronic system that meets all specified performance, quality, cost, and reliability requirements.

► 10.1 ELECTRICAL TESTING

Electrical measurements performed on test structures are a major mechanism for assessing integrated circuit yield (see Section 10.5), as well as other indicators of product performance. Such measurements are performed both during and at the conclusion of the fabrication process. In addition, electrical testing of the final product is crucial to ensure quality. These concepts are discussed in more detail in the following subsections.

10.1.1 Test Structures

To assess the impact of the presence of defects on semiconductor wafers caused by particles, contamination, or other sources, specially designed test structures are used. These structures, also known as *process control monitors* (PCMs), include single transistors, single lines of conducting material, MOS capacitors, and interconnect monitors. Product wafers typically contain several PCMs distributed across the surface, either in die sites or in the scribe lines between die (see Fig. 10.2).¹

Process quality can be checked at various stages of manufacturing through in-line measurements on PCM structures. Three typical interconnect test structures are shown in Figure 10.3.¹ Using such test structures, measurements are performed to assess the presence of defects, which can be inferred by the presence of short circuits or open circuits using simple resistance measurements. For example, the meander structure facilitates the detection of open circuits through increased end-to-end resistance of the

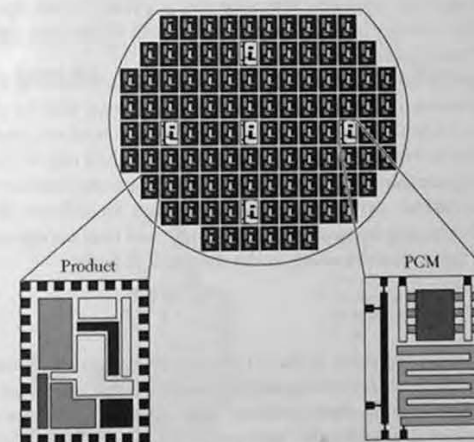


Figure 10.2 Configuration of products and PCMs on a typical semiconductor wafer.¹

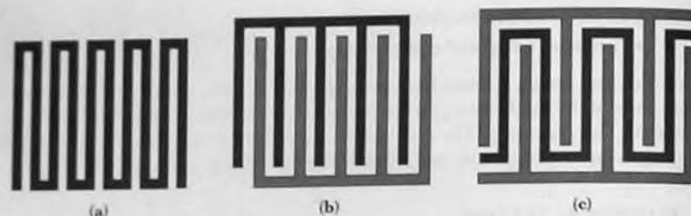


Figure 10.3 Basic test structures for interconnect layers.¹ (a) Meander structure. (b) Double-comb structure. (c) Comb-meander-comb structure.

meander. The double-comb structure can likewise be used to detect shorts, since any extra conducting material bridging the two combs will reduce the resistance between combs significantly. The comb-meander-comb structure combines the capabilities of the other two structures and permits the detection of both shorts and opens. Various combinations of widths of lines and spaces in these test structures allow the collection of statistics on defects of various sizes.

10.1.2 Final Test

Functional testing at the completion of manufacturing is the final arbiter of process quality and yield. The purpose of final testing is to ensure that all products perform to the specifications for which they were designed. For integrated circuits, the test process depends a great deal on whether the chip tested is a logic or memory device. In either case, automated test equipment (ATE) is used to apply a measurement stimulus to the chip and record the results. The major functions of the ATE are input pattern generation, pattern application, and output response detection.²

During each functional test cycle, input vectors are sent through the chip by the ATE in a timed sequence. Output responses are read and compared with expected results. This sequence is repeated for each input pattern. It is often necessary to perform such tests at various supply voltages and operating temperatures to ensure device operation at all potential regimes. The number and sequence of failures in the output signature are indicative of manufacturing process faults.

Test results may be expressed in a variety of ways.² Two examples are shown in Figures 10.4 and 10.5. Figure 10.4 shows a two-dimensional plot called a *shmoo plot* for a hypothetical bipolar product. In a shmoo plot, the outlined shaded region is where the device is intended to operate, while the blank area outside represents the failure region. Another typical test output is the cell map shown in Figure 10.5. Cell maps are very useful in identifying and isolating device failures, particularly in memory arrays. In addition, the patterns generated in the cell map may be compiled, catalogued, and later compared with a library of existing defect types, thereby aiding in the diagnosis of faults.

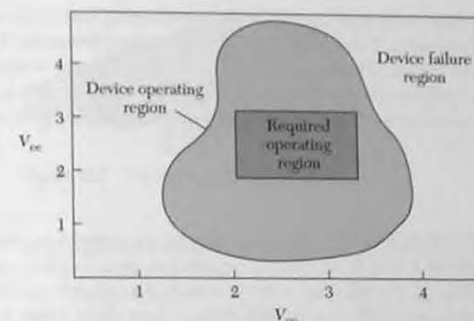


Figure 10.4 Example of two-dimensional voltage shmoo plot for a bipolar IC.²

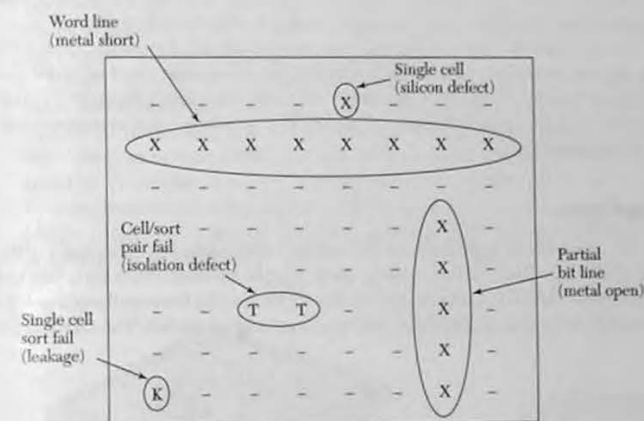


Figure 10.5 Cell map showing examples of failure patterns and defect types.²

► 10.2 PACKAGING

Loosely defined, the term *packaging* refers to the set of technologies and processes that connect ICs with electronic systems. A useful analogy is to consider an electronic product as the human body. Like the body, these products have “brains,” which are analogous to ICs. Electronic packaging provides the “nervous system” as well as the “skeletal system.” The package is responsible for interconnecting, powering, cooling, and protecting the IC.³ This concept is illustrated in Figure 10.6.

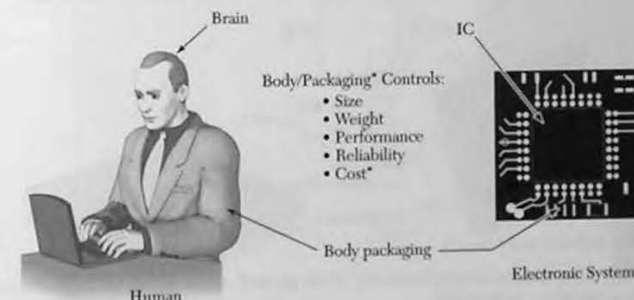


Figure 10.6 Analogy between electronic packaging and the human body.³

Overall, electronic systems consist of several levels of packaging, each with distinctive types of interconnection devices. Figure 10.7 depicts this packaging hierarchy.⁴ Level 0 consists of on-chip interconnections. Chip-to-printed circuit board or chip-to-module connections comprise Level 2, and board-to-board interconnections comprise Level 3. Levels 4 and 5 consist of connections between subassemblies and between systems (such as computer to printer), respectively.

10.2.1 Die Separation

Following functional testing, individual ICs (or dice) must be separated from the substrate. This is essentially the first step in the packaging process. In a common method that has been used for many years, the substrate wafer is mounted on a holder and scribed in both the x and y directions using a diamond scribe. This is done along scribe borders 75 to 250 μm in width that are formed around the periphery of the dice during fabrication. These borders are aligned with the crystal planes of the substrate if possible. After scribing, the wafer is removed from the holder and placed upside-down on a soft support. A roller is then used to apply pressure, fracturing the wafer along the scribe lines. This must be accomplished with minimal damage to the individual die.

More modern die separation processes use a diamond saw rather than a diamond scribe. In this procedure, the wafer is attached to an adhesive sheet of mylar film. The saw is then used to either scribe the wafer or to cut completely through it. After separation, the dice are removed from the mylar. The separated dice are then ready to be placed into packages.

10.2.2 Package Types

There are a number of approaches to the packaging of single ICs. The *dual in-line package*, or DIP (Fig. 10.8), is the package most people envision when they think of integrated circuits. The DIP was developed in the 1960s, quickly became the primary package for ICs, and has long dominated the electronics packaging market. The DIP can be made

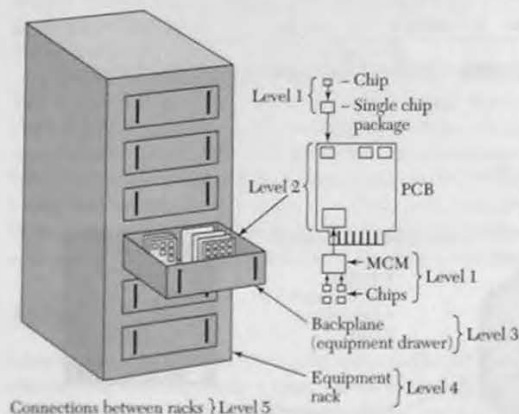


Figure 10.7 Electronic packaging hierarchy.⁴ PCB, printed circuit board; MCM, multichip module.

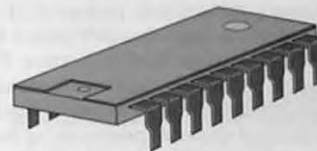


Figure 10.8 Dual in-line package.⁴

of plastic or ceramic, with the latter called the *CerDIP*. The CerDIP consists of a DIP constructed of two pieces of sandwiched ceramic with leads protruding from between the ceramic plates.

In the 1970s and 1980s, *surface mount packages* were developed in response to a need for higher-density interconnect than the DIP approach could provide. In contrast to DIPs, the leads of a surface-mounted package do not penetrate the printed circuit board (PCB) upon which it is mounted. This means that the package can be mounted on both sides of the board, thereby allowing higher density. One example of such a package is the quad flatpack, or QFP (Fig. 10.9), which has leads on all four sides to further increase the number of input/output (I/O) connections.

More recently, the need for greater and greater numbers of I/O connections has led to the development of pin grid array (PGA) and ball grid array (BGA) packages (Figs. 10.10 and 10.11, respectively). PGAs have an I/O density of about 600, and BGAs can have densities greater than 1000, as compared with about 200 for QFPs. BGAs can be identified by the solder bumps on the bottom of the package. With QFPs, as the spacing between leads becomes tighter, the manufacturing yield decreases rapidly. The BGA allows higher density and takes up less space than the QFP, but its manufacturing process is inherently more expensive.

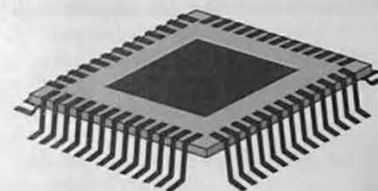


Figure 10.9 Quad flatpack.⁴

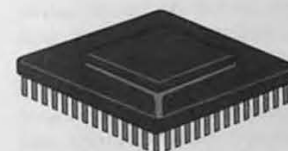


Figure 10.10 Pin grid array package.⁴

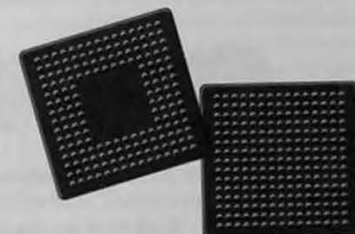


Figure 10.11 Ball grid array package.⁴

The most recent development in packaging is the chip scale package (CSP), which is shown in Figure 10.12. CSPs, defined as packages no larger than 20% greater than the size of the IC die itself, often take the form of miniaturized ball grid arrays. They are designed to be flip-chip mounted (see Section 10.2.3) using conventional equipment and solder reflow. CSPs are typically manufactured in a process that creates external power and signal I/O contacts and encapsulates the finished silicon die prior to dicing the wafer. Essentially, CSPs provide an interconnection framework for ICs so that before dicing, each die has all the functions (e.g., external electrical contacts, encapsulation of the finished silicon) of a conventional fully packaged IC. Two essential features of this approach are that the leads and interposer layer (an added layer on the IC used to provide electrical functionality and mechanical stability) are flexible enough so that the packaged device is compliant to the test fixture for full testing and burn-in, and the package can accommodate the vertical nonplanarity and thermal expansion and contraction of the underlying printed circuit board during assembly and operation.

10.2.3 Attachment Methodologies

An IC must be mounted and bonded to a package, and that package must be attached to a printed circuit board before the IC can be used in an electronic system. Methods of attaching ICs are referred to as *Level 1 packaging*. The technique used to bond a bare die to a package has a significant effect on the ultimate electrical, mechanical, and thermal properties of the electronic system being manufactured. Chip-to-package interconnection is generally accomplished by either wire bonding, tape-automated bonding, or flip-chip bonding (see Fig. 10.13).

Wire Bonding

Wire bonding is the oldest attachment method and is still the dominant technique for chips with fewer than 200 I/O connections. Wire bonding requires connecting gold or aluminum wires between chip bonding pads and contact points on the package. ICs are first attached to the substrate using a thermally conductive adhesive, with their bonding

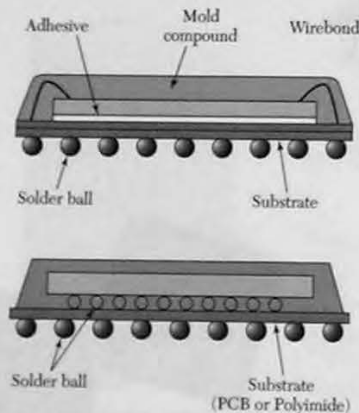


Figure 10.12 Examples of two typical CSPs.³

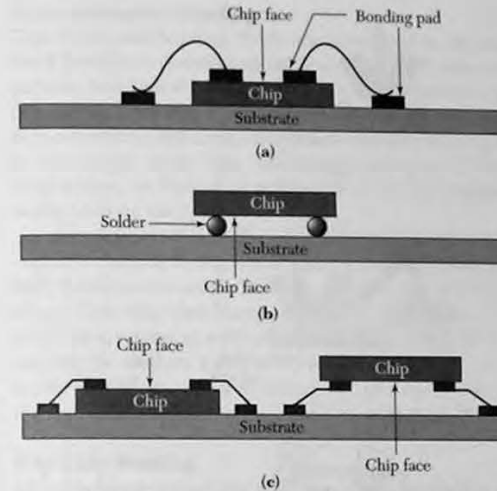


Figure 10.13 Illustration of (a) wire, (b) flip-chip, and (c) tape-automated bonding.⁴

pads facing upward. The Au or Al wires are then attached between the pads and substrate using ultrasonic, thermosonic, or thermocompression bonding.⁵ Although automated, this process is still time-consuming since each wire must be attached individually.

In the *thermocompression* technique (Fig. 10.14), a fine wire (15–75 μm diameter) is fed from a spool through a heated capillary. A small hydrogen torch or electric spark then melts the end of the wire, forming a ball. The ball is then positioned over the chip bonding pad, the capillary is lowered, and the ball deforms into a “nail head” due to pressure and heat from the capillary. (The substrate is maintained at a temperature of 150°C to 200°C, and the bonding interface temperature ranges from 280°C to 350°C). Next, the capillary is raised, and wire is fed from the spool and positioned over the package substrate. The bond to the package is a wedge bond produced by deforming the wire with the edge of the capillary. The capillary is then raised and the wire is broken near the edge of the bond.

Oxidation of Al at high temperatures makes it difficult to form a good ball at the end of the wire; in addition, many epoxies cannot withstand the temperatures needed in thermocompression bonding. *Ultrasonic bonding* represents a lower-temperature alternative that relies on a combination of pressure and rapid mechanical vibration to form bonds (Fig. 10.15). In this approach, the wire is fed from a spool through a hole in the bonding tool, which is then lowered into position as an ultrasonic vibration at 20 to 60 kHz causes the metal to deform and flow (even at room temperature). As the tool is raised after the bond to the package is formed, a clamp pulls and breaks the wire.

Thermosonic bonding is a combination of the other two techniques. The substrate temperature is maintained at about 150°C, and ultrasonic vibration and pressure are used to cause the metal to flow under pressure to form a weld. Thermosonic bonders are quite fast—they are capable of producing 5 to 10 bonds per second.

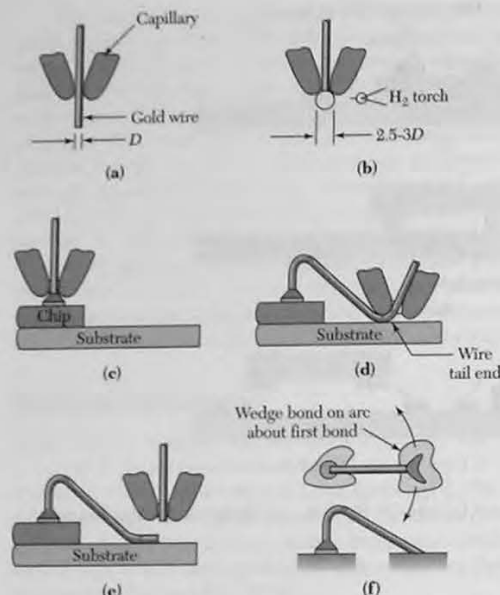


Figure 10.14 Thermocompression bonding process.⁵ (a) Gold wire in a capillary. (b) Ball formation. (c) Bonding. (d) Wire loop and edge bonding. (e) Wire broken at edge. (f) Geometry of ball-wedge bond.

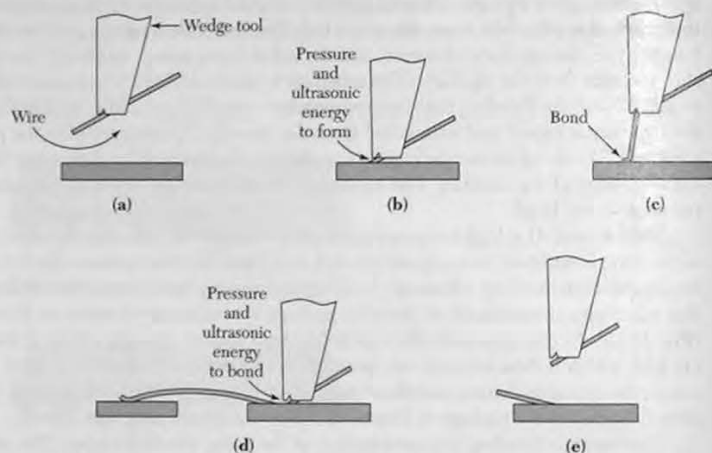


Figure 10.15 Ultrasonic bonding process.⁵ (a) Tool guides wire to the package. (b) Pressure and ultrasonic energy form bond. (c) Tool feeds wire and repositions above the IC. (d) Wire broken at bond.

Tape-Automated Bonding

Tape-automated bonding (TAB) was developed in the early 1970s and is often used to bond packages to printed circuit boards. In TAB, chips are first mounted on a flexible polymer tape (usually polyimide) containing repeated copper interconnection patterns (Fig. 10.16). The copper leads are defined by lithography and etching, and the lead pattern can contain hundreds of connections. After aligning the IC pads to metal interconnection stripes on the tape, attachment takes place by thermocompression (Fig. 10.17). Gold bumps are formed on either side of the die or tape and are used to bond the die to the leads on the tape.

A benefit of TAB is that all bonds are formed simultaneously, which significantly improves manufacturing throughput. However, unless all the leads are coplanar, reliability problems can result. TAB also requires multilayer solder bumps with complex metallurgy. Generally, these bumps use gold or copper as the primary constituent, with titanium or tungsten serving as a diffusion barrier to prevent alloying. In addition, a particular tape can only be used for a chip and package that matches its interconnect pattern, thereby rendering TAB an extremely customized process for which the bonding equipment is relatively expensive.

Flip-Chip Bonding

Flip-chip bonding is a direct interconnection approach in which the IC is mounted upside-down onto a module or printed circuit board. Electrical connections are made via solder bumps (or solderless materials such as epoxies or conductive adhesives) located over the surface of the chip. Since bumps can be located anywhere on the chip, flip-chip bonding ensures that the interconnect distance between the chip and package is minimized. The I/O density is limited only by the minimum distance between adjacent bond pads.

In flip-chip processing, chips are placed face down on the module substrate so that I/O pads on the chip are aligned with those on the substrate (Fig. 10.18). A solder reflow process is then used to simultaneously form all the required connections, thereby drastically improving throughput compared with wire bonding. However, the bump fabrication process itself is fairly complex and capital intensive.

Solderless flip-chip technology involves stencil printing of an organic polymer onto an IC, leaving the bond pads uncoated. A high-conductivity organic polymer paste is then stenciled onto the bond pads to form the solderless bumps, which are then cured. The same organic polymer is stenciled onto the bond pads of the substrate. Alignment is then accomplished, and the final bond is formed by applying pressure and heat to the bumps.

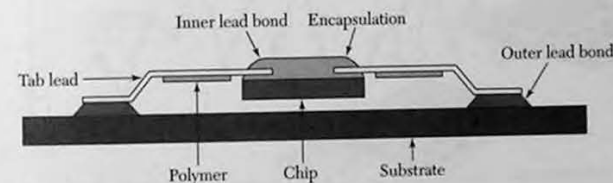


Figure 10.16 Tape-automated bonding.⁴

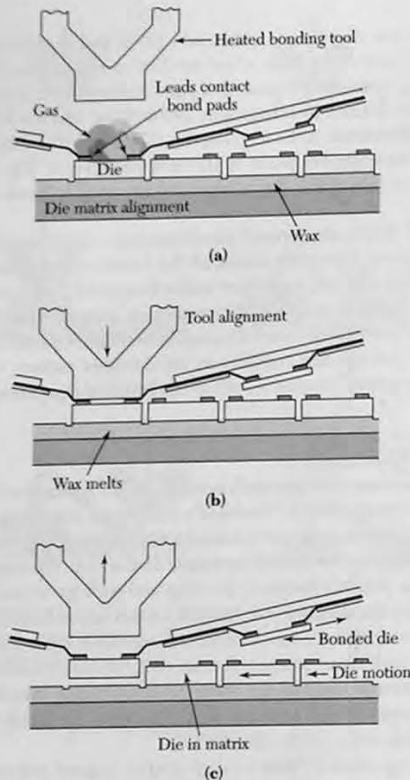


Figure 10.17 TAB procedure.⁵ (a) Leads lowered into position and aligned above bonding pads. (b) Tool descends and performs bonds. (c) Tool and film are raised so a new die can be moved into position.

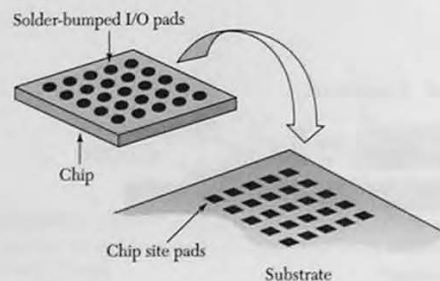


Figure 10.18 Flip-chip bonding.⁴

► 10.3 STATISTICAL PROCESS CONTROL

IC manufacturing processes must be stable, repeatable, and of high quality to yield products with acceptable performance. This implies that all individuals involved in manufacturing an IC (including operators, engineers, and management) must continuously seek to improve manufacturing process output and reduce variability. Variability reduction is accomplished in large part by strict process control. This section focuses on *statistical process control* (SPC) techniques as a means to achieve high-quality products.

SPC refers to a powerful collection of problem-solving tools used to achieve process stability and reduce variability. Perhaps the primary and most technically sophisticated of these tools is the control chart. The control chart was developed by Dr. Walter Shewhart of Bell Telephone Laboratories in the 1920s.⁶ For this reason, control charts are also often referred to as *Shewhart control charts*.

A control chart is an online SPC technique that is used to detect the occurrence of shifts in process performance so that investigation and corrective action may be undertaken to bring an incorrectly behaving manufacturing process back under control. A typical control chart is shown in Figure 10.19. This chart is a graphical display of a quality characteristic that has been measured from a sample versus the sample number or time. The chart consists of (a) a *center line*, which represents the mean value of the characteristic corresponding to an in-control state; (b) an *upper control limit* (UCL); and (c) a *lower control limit* (LCL). The control limits are selected such that if the process is under statistical control, nearly all the sample points will plot between them. If the variance of the quality characteristic is σ^2 and the standard deviation of the characteristic is σ , then the control limits are typically set at $\pm 3\sigma$ from the center line. Points that plot outside of the control limits are interpreted as evidence that the process is out of control.

10.3.1 Control Charts for Attributes

Some quality characteristics cannot be easily represented numerically. For example, we may be concerned with whether or not a wire bond is defective. In this case, the bond is classified as either defective or nondefective (or equivalently, conforming or nonconforming), and there is no numerical value associated with the quality of the bond. Quality characteristics of this type are referred to as *attributes*.

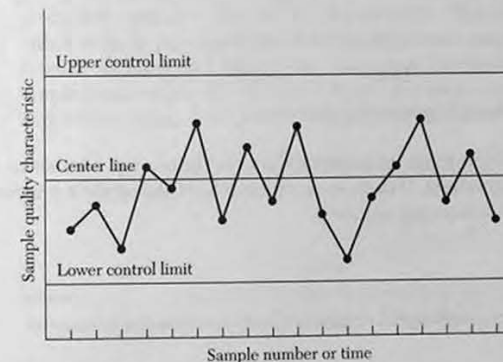


Figure 10.19 Typical control chart.⁶

Two commonly used control charts for attributes are the *defect chart* (or *c-chart*) and the *defect density chart* (or *u-chart*). When a specification is not satisfied in a product, a defect or nonconformity may result. In such cases, it is possible to develop control charts for either the total number of defects or the defect density. These charts assume that the presence of defects in samples of constant size is appropriately modeled by the *Poisson distribution*,⁶ in which the probability of a defect occurring is given by

$$P(x) = \frac{e^{-c} c^x}{x!} \quad (1)$$

where x is the number of defects and c is a constant greater than zero. For the Poisson distribution, c is both the mean and variance. Therefore, the *c-chart* with $\pm 3\sigma$ control limits is given by

$$\begin{aligned} \text{UCL} &= c + 3\sqrt{c} \\ \text{Centerline} &= c \\ \text{LCL} &= c - 3\sqrt{c} \end{aligned} \quad (2)$$

assuming that c is known. (Note: if these calculations yield a negative value for the LCL, the standard practice is to set the LCL equal to 0). If c is not known, it may be estimated from an observed average number of defects in a sample ($\bar{c}$). In this case, the control chart becomes

$$\begin{aligned} \text{UCL} &= \bar{c} + 3\sqrt{\bar{c}} \\ \text{Center line} &= \bar{c} \\ \text{LCL} &= \bar{c} - 3\sqrt{\bar{c}} \end{aligned} \quad (3)$$

EXAMPLE 1

Suppose the inspection of 25 silicon wafers yields 37 defects. Set up a *c-chart* for this situation.

SOLUTION We estimate c using

$$\bar{c} = \frac{37}{25} = 1.48$$

This is the center line for the *c-chart*. The upper and lower control limits can be found from Eq. 3 as follows:

$$\begin{aligned} \text{UCL} &= \bar{c} + 3\sqrt{\bar{c}} = 5.13 \\ \text{LCL} &= \bar{c} - 3\sqrt{\bar{c}} = -2.17 \end{aligned}$$

Since $-2.17 < 0$, we set the LCL equal to 0 in this case.

Suppose we would like to set up a control chart for the *average* number of defects over a sample size of n products. If there were c total defects among the n samples, then the average number of defects per sample is

$$u = \frac{c}{n} \quad (4)$$

The parameters of a 3σ defect density chart (*u-chart*) are then given by

$$\begin{aligned} \text{UCL} &= \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} \\ \text{Center line} &= \bar{u} \\ \text{LCL} &= \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} \end{aligned} \quad (5)$$

where u is the average number of defects over m groups of sample size n .

EXAMPLE 2

Suppose an IC manufacturer wants to establish a defect density chart. Twenty different samples of size $n = 5$ wafers are inspected, and a total of 183 defects are found. Set up the *u-chart* for this situation.

SOLUTION We estimate u using

$$\bar{u} = \frac{u}{m} = \frac{c}{mn} = \frac{183}{(20)(5)} = 1.83$$

This is the center line for the *u-chart*. The upper and lower control limits can be found from Eq. 5 as follows:

$$\begin{aligned} \text{UCL} &= \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 3.64 \\ \text{LCL} &= \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 0.02 \end{aligned}$$

10.3.2 Control Charts for Variables

In many cases, quality characteristics are expressed as specific numeric measurements, rather than assessing the probability of presence of defects. For example, the thickness of a film is an important characteristic to be measured and controlled. Control charts for continuous variables such as this can provide more information regarding manufacturing process performance than attribute control charts like the *c-* and *u-*charts.

When attempting to control continuous variables, it is important to control both the mean and variance of the quality characteristic. This is true because shifts or drifts in either of these parameters can result in significant misprocessing. Control of the mean is achieved using an $\bar{x}$ -chart, and variance can be monitored using the standard deviation as in an *s*-chart. The names of these two charts originate from the *sample mean* ($\bar{x}$) and *sample variance* (s^2), which, respectively, are given by

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7)$$

where $x_1, x_2, \dots, x_n$ are observations in a sample of size n . The square root of the sample variance is known as the *sample standard deviation* (s).

Suppose m samples of size n are collected. If $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ are the sample means, the best estimator for the true mean (μ) is the *grand average* ($\bar{\bar{x}}$), which is given by

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (8)$$

Since $\bar{\bar{x}}$ estimates μ , it is used as the center line of the $\bar{x}$ -chart. It can also be shown⁶ that if a quality characteristic is normally distributed with a known mean μ and standard deviation σ , then $\bar{\bar{x}}$ is also normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. Thus, the center line and control limits for the $\bar{x}$ -chart are

$$\begin{aligned} \text{UCL} &= \bar{\bar{x}} + 3\sqrt{\frac{\sigma}{n}} \\ \text{Center line} &= \bar{\bar{x}} \\ \text{LCL} &= \bar{\bar{x}} - 3\sqrt{\frac{\sigma}{n}} \end{aligned} \quad (9)$$

Since σ is unknown, it must also be estimated by analyzing past data. Caution must be applied in doing so since s itself cannot be used directly as the estimate because s is not an *unbiased* estimator of σ . (The term *unbiased* refers to the situation in which the expected value of the estimator is equal to the parameter being estimated.) Instead s actually estimates c_s , where c_s is a statistical parameter that is dependent on the sample size (see Table 10.1). For m samples of size n , the average sample standard deviation is

$$\bar{s} = \frac{1}{m} \sum_{j=1}^m s_j \quad (10)$$

It turns out that the statistic $\bar{s}/c_s$ is, in fact, an unbiased estimator of σ .

In addition, the standard deviation of s is $\sigma\sqrt{1-c_s^2}$. Using this information, the control limits for the s -chart can be set up as follows:

$$\begin{aligned} \text{UCL} &= \bar{s} + 3\frac{\bar{s}}{c_s}\sqrt{1-c_s^2} \\ \text{Center line} &= \bar{s} \\ \text{LCL} &= \bar{s} - 3\frac{\bar{s}}{c_s}\sqrt{1-c_s^2} \end{aligned} \quad (11)$$

When $\bar{s}/c_s$ is used to estimate σ , the limits on the corresponding $\bar{x}$ -chart may be defined as

$$\begin{aligned} \text{UCL} &= \bar{\bar{x}} + \frac{3\bar{s}}{c_s\sqrt{n}} \\ \text{Center line} &= \bar{\bar{x}} \\ \text{LCL} &= \bar{\bar{x}} - \frac{3\bar{s}}{c_s\sqrt{n}} \end{aligned} \quad (12)$$

EXAMPLE 3

Suppose $\bar{x}$ - and s -charts are to be established to control linewidth for a lithography process. Twenty-five different samples of size $n = 5$ linewidths are measured. Suppose the grand average for the 125 lines is 4.01 μm . If $\bar{s} = 0.09$ mm, what are the control limits for the s -chart?

TABLE 10.1 c_4 Parameter for s -Chart

Sample Size (n)	c_4
2	0.7979
3	0.8862
4	0.9213
5	0.9400
6	0.9515
7	0.9594
8	0.9650
9	0.9693
10	0.9727
11	0.9754
12	0.9776
13	0.9794
14	0.9810
15	0.9823
16	0.9835
17	0.9845
18	0.9854
19	0.9862
20	0.9869
21	0.9876
22	0.9882
23	0.9887
24	0.9892
25	0.9896
$n > 25$	$c_4 \approx \frac{4(n-1)}{4n-3}$

SOLUTION The value for c_4 for $n = 5$ (found in Table 10.1) is 0.94. The upper and lower control limits for $\bar{x}$ can be found from Eq. 12 as follows:

$$\text{UCL} = \bar{\bar{x}} + \frac{3\bar{s}}{c_s\sqrt{n}} = 4.14 \mu\text{m}$$

$$\text{LCL} = \bar{\bar{x}} - \frac{3\bar{s}}{c_s\sqrt{n}} = 3.88 \mu\text{m}$$

The upper and lower control limits for s can be found from Eq. 11 as follows:

$$\text{UCL} = \bar{s} + 3\frac{\bar{s}}{c_s}\sqrt{1-c_s^2} = 0.19 \mu\text{m}$$

$$\text{LCL} = \bar{s} - 3\frac{\bar{s}}{c_s}\sqrt{1-c_s^2} = 0$$

Note: Since the LCL is (slightly) negative, we automatically set it to zero.

▶ 10.4 STATISTICAL EXPERIMENTAL DESIGN

Experiments allow investigators to determine the effects of several variables on a given process or product. A *designed experiment* is a test or series of tests that involve purposeful changes to these variables in order to observe the effect of the changes on that process or product. *Statistical experimental design* is an efficient approach for systematically varying these controllable process variables and ultimately determining their impact on process or product quality, or both. This approach is useful for comparing methods, deducing dependencies, and creating models to predict effects.

Statistical process control and experimental design are closely interrelated. Both techniques can be used to reduce variability. However, SPC is a passive approach in which a process is monitored and data collected, whereas experimental design requires active intervention in performing tests on the process under different conditions. Experimental design can also be beneficial in implementing SPC, since designed experiments may help to identify the most influential process variables, as well as their optimum settings.

Overall, experimental design is a powerful engineering tool for improving a manufacturing process. Application of experimental design techniques can lead to improved yield, reduced variability, reduced development time, and reduced cost. Ultimately, the result is enhanced manufacturability, performance, and product reliability. The following sections illustrate the use of experimental design methods in IC fabrication.

10.4.1 Comparing Distributions

Consider the yield data in Table 10.2 obtained from an IC manufacturing process in which two batches of 10 wafers were fabricated using a standard method (Method A) and a modified method (Method B). The question to be answered from the experiment is what evidence (if any) does the data collected provide that Method B is really better than Method A?

To answer this question, we examine the average yields for each process. The modified method (Method B) gave an average yield that was 1.30% higher than the standard method. However, due to the considerable variability in the individual test results, it might not be correct to immediately conclude that Method B is superior to Method A. In fact,

TABLE 10.2 Yield Data from a Hypothetical IC Manufacturing Process

Wafer	Method A Yield (%)	Method B Yield (%)
1	89.7	84.7
2	81.4	86.1
3	84.5	83.2
4	84.8	91.9
5	87.3	86.3
6	79.7	79.3
7	85.1	86.2
8	81.7	89.1
9	83.7	83.7
10	84.5	88.5
Average	84.24	85.54

it is conceivable that the difference observed could be due to experimental error, operator error, or even pure chance.

The proper approach to determine whether the difference between the two manufacturing processes is significant is a *hypothesis test*. A statistical hypothesis is a statement about the values of the parameters of a probability distribution. A hypothesis test is an evaluation of the validity of the hypothesis according to some criterion. Hypotheses are expressed in the following manner:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_1: \mu &\neq \mu_0 \end{aligned} \quad (13)$$

where the statement $H_0: \mu = \mu_0$ is called the *null hypothesis*, and $H_1: \mu \neq \mu_0$ is called the *alternative hypothesis*. To perform a hypothesis test, we select a random sample from a population, compute an appropriate test statistic, and then either accept or reject the null hypothesis. For the yield experiment, the hypothesis test can be represented as

$$\begin{aligned} H_0: \mu_A &= \mu_B \\ H_1: \mu_A &\neq \mu_B \end{aligned} \quad (14)$$

where μ_A and μ_B represent the mean yields for the two methods.

To evaluate this hypothesis, a *test statistic* is required. The appropriate test statistic in this case is⁶

$$t_0 = \frac{(\bar{y}_A - \bar{y}_B)}{s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \quad (15)$$

where $\bar{y}_A$ and $\bar{y}_B$ are the sample means of the yields for each method, n_A and n_B are the number of trials in each sample (10 each in this case), and

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} \quad (16)$$

which is referred to as the *pooled estimate of the common variance* of the two processes. The denominator of Eq. 16 is called the number of *degrees of freedom* for the hypothesis test. The values of the sample variances are calculated using Eq. 7: $s_A = 2.90$ and $s_B = 3.65$. Using Eqs. 16 and 15 then gives values of $s_p = 3.30$ and $t_0 = 0.88$, respectively.

We can use Appendix K to determine the probability of computing a given *t* statistic with a certain number of degrees of freedom. This probability is represented by the shaded region in the figure in the appendix. Interpolating from Appendix K, we find that the likelihood of computing a *t* statistic with $n_A + n_B - 2 = 18$ degrees of freedom equal to 0.88 is 0.195. The value 0.195 is the *statistical significance* of the hypothesis test. This means that there is only a 19.5% chance that the observed difference between the mean yields is due to pure chance. In other words, we can be 80.5% confident that Method B is really superior to Method A.

10.4.2 Analysis of Variance

The previous example shows how we might use hypothesis testing to compare two distributions. However, it is often important in IC manufacturing applications to be able to compare several distributions. Moreover, we might also be interested in determining which process conditions in particular have a significant impact on process quality. *Analysis of variance* (ANOVA) is an excellent technique for accomplishing these objectives. ANOVA

builds on the idea of hypothesis testing and allows us to compare different sets of process conditions (i.e., treatments), as well as to determine whether a given treatment results in a statistically significant variation in quality.

The ANOVA procedure is best illustrated by example. In the following discussion, consider the data in Table 10.3, which represents hypothetical defect densities measured on wafers fabricated using four different sets of process recipes (labeled 1 through 4). Through the use of ANOVA, we will determine whether the discrepancies *between* recipes (i.e., treatments) is truly greater than the variation of the via diameters *within* the individual groups of wafers processed with the same recipe.

Let k be the number of treatments ($k = 4$ in this case). Note that the sample size (n) for each treatment varies ($n_1 = 4$, $n_2 = n_3 = 6$, and $n_4 = 8$). The treatment means (in cm^{-2}) are as follows: $\bar{y}_1 = 61$, $\bar{y}_2 = 66$, $\bar{y}_3 = 68$, and $\bar{y}_4 = 61$. The total number of samples (N) is 24, and the grand average of all 24 samples is $\bar{y} = 64 \text{ cm}^{-2}$.

Sums of Squares

To perform ANOVA, several key parameters must be computed. These parameters, called *sums of squares*, serve to quantify deviations within and between different treatments. Let y_{it} represent the i th observation for the t th treatment. The sum of squares within the t th treatment is given by

$$S_t = \sum_{i=1}^{n_t} (y_{it} - \bar{y}_t)^2 \quad (17)$$

where n_t is the sample size for the treatment in question and y_{it} is the treatment mean. The *within-treatment sum of squares* for all treatments is

$$S_R = S_1 + S_2 + \dots + S_k = \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y}_t)^2 \quad (18)$$

To quantify the deviations of the treatment averages from the grand average, we use the *between-treatment sum of squares*, which is given by

$$S_T = \sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2 \quad (19)$$

Finally, the total sum of squares for all the data about the grand average is

$$S_D = \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y})^2 \quad (20)$$

TABLE 10.3 Hypothetical Defect Densities (in cm^{-2}) for Four Different Process Recipes

Recipe 1	Recipe 2	Recipe 3	Recipe 4
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59

Each sum of squares has an associated number of degrees of freedom required for its computation. The degrees of freedom for the within-treatment, between-treatment, and total sums of squares, respectively, are

$$\begin{aligned} v_R &= N - k \\ v_T &= k - 1 \\ v_D &= N - 1 \end{aligned} \quad (21)$$

The final quantity needed to carry out analysis of variance is the pooled estimate of the variance quantified by each sum of squares. This quantity, known as the *mean square*, is the ratio of the sum of squares to its associated number of degrees of freedom. The within-treatment, between-treatment, and total mean squares are therefore

$$\begin{aligned} s_R^2 &= \frac{S_R}{v_R} = \frac{\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y}_t)^2}{N - k} \\ s_T^2 &= \frac{S_T}{v_T} = \frac{\sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2}{k - 1} \\ s_D^2 &= \frac{S_D}{v_D} = \frac{\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y})^2}{N - 1} \end{aligned} \quad (22)$$

ANOVA Table

Once the parameters just described have been computed, it is customary to arrange them in a tabular format called an *ANOVA table*. The general form of the ANOVA table is depicted in Table 10.4. The ANOVA table that corresponds to the defect density data in Table 10.3 is shown in Table 10.5. Note that in both the Sum of Squares and Degrees of Freedom columns, the values for between and within treatments add up to give the corresponding total value. This additive property of the sum of squares arises from the algebraic identity

$$\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y})^2 = \sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2 + \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{it} - \bar{y}_t)^2 \quad (23)$$

or equivalently,

$$S_D = S_T + S_R$$

The complete ANOVA table provides a mechanism for testing the hypothesis that all of the treatment means are equal. The null hypothesis in this case is thus

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

TABLE 10.4 General Format of the ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Between treatments	S_T	$v_T = k - 1$	s_T^2	s_T^2 / s_R^2
Within treatments	S_R	$v_R = N - k$	s_R^2	
Total	S_D	$v_D = N - 1$	s_D^2	

TABLE 10.5 ANOVA Table for Via Diameter Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Between treatments	$S_T = 228$	$\nu_T = 3$	$s_T^2 = 76.0$	$s_T^2/s_R^2 = 13.6$
Within treatments	$S_R = 112$	$\nu_R = 20$	$s_R^2 = 5.6$	
Total	$S_D = 340$	$\nu_D = 23$	$s_D^2 = 14.8$	

If the null hypothesis were true, the ratio s_T^2/s_R^2 would follow the F distribution with ν_T and ν_R degrees of freedom. Interpolating from Appendix L, the *significance level* (i.e., the first subscript of the F values in the tables) for the observed F ratio of 13.6 with 3 and 20 degrees of freedom is 0.00046. This means that there is only a 0.0046% chance that the means are in fact equal, and the null hypothesis is discredited. In other words, we can be 99.9954% sure that real differences exist among the four different processes in our example.

10.4.3 Factorial Designs

Experimental design is an organized method of conducting experiments to extract the maximum amount of information from a limited number of experiments. Experimental design techniques are employed in manufacturing to systematically and efficiently explore the effects of a set of input variables, or *factors* (such as processing temperature), on responses (such as yield). The unifying feature in statistically designed experiments is that all factors are varied simultaneously, as opposed to the more traditional "one-variable-at-a-time" technique. A properly designed experiment can minimize the number of experimental runs that would otherwise be required if this approach or random sampling were used.

Factorial experimental designs are of great practical importance for IC manufacturing applications. To perform a factorial experiment, an investigator selects a fixed number of levels for each of a number of variables (factors) and runs experiments at all possible combinations of the levels. Two of the most important issues in factorial experimental designs are choosing the set of factors to be varied in the experiment and specifying the ranges over which variation will take place. The choice of the number of factors directly affects the number of experimental runs (and therefore the cost of the experiment).

Two-Level Factorials

The ranges of the process variables investigated in factorial experiments can be discretized into minimum, maximum, and "center" levels. In a *two-level factorial design*, the minimum and maximum levels of each factor (normalized to take on values -1 and $+1$, respectively) are used together in every possible combination. Thus, a full two-level factorial experiment with n factors requires 2^n experimental runs. The various factor level combinations of a three-factor experiment can be represented pictorially as the vertices of a cube, as shown in Figure 10.20.

Table 10.6 shows a 2^3 factorial experiment for a CVD process. The three factors are temperature (T), pressure (P), and gas flow rate (F). The response being measured is the deposition rate (D) in angstroms per minute. The highest and lowest levels of each factor are represented by the $+$ and $-$ signs, respectively. The display of levels depicted in the first three columns of this table is called a *design matrix*.

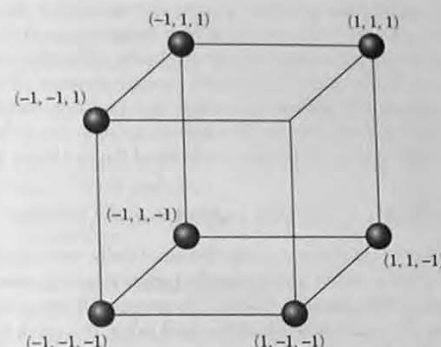


Figure 10.20 Factor level combinations of a three-factor experiment represented as the vertices of a cube.

TABLE 10.6 Two-Level Factorial Experiment

Run	P	T	F	D ($\text{\AA}/\text{min}$)
1	-	-	-	$d_1 = 94.8$
2	+	-	-	$d_2 = 110.96$
3	-	+	-	$d_3 = 214.12$
4	+	+	-	$d_4 = 255.82$
5	-	-	+	$d_5 = 94.14$
6	+	-	+	$d_6 = 145.92$
7	-	+	+	$d_7 = 286.71$
8	+	+	+	$d_8 = 340.52$

What can we determine from this factorial design? Furthermore, what does the data collected tell us about the effect of pressure on deposition rate? The effect of any single variable on the response is called a *main effect*. The method used to compute such a main effect is to find the difference between the average deposition rate when the pressure is high (i.e., runs 2, 4, 6, and 8) and the average deposition rate when the pressure is low (runs 1, 3, 5, and 7). Mathematically, this is expressed as

$$P = d_p - d_{p-} = 1/4[(d_2 + d_4 + d_6 + d_8) - (d_1 + d_3 + d_5 + d_7)] = 40.86 \quad (24)$$

where P is the main effect for pressure, d_p is the average deposition rate when the pressure is high, and d_{p-} is the average deposition rate when the pressure is low. We interpret this result as showing that the average effect of increasing pressure from its lowest to its highest level is to increase the deposition rate by 40.86 $\text{\AA}/\text{min}$. The other main effects for temperature and flow rate are computed in a similar manner. In general, the main effect for each variable in a two-level factorial experiment is the difference between the two averages of the response (y), or

$$\text{Main effect} = y_+ - y_- \quad (25)$$

We are also interested in quantifying how two or more factors *interact*. For example, suppose that the pressure effect is much greater at high temperatures than it is at low temperatures. A measure of this interaction is provided by the difference between the average pressure effect with temperature high and the average pressure effect with temperature low. By convention, *half* of this difference is called the *pressure-by-temperature interaction*, or symbolically, the $P \times T$ interaction. This interaction may also be thought of as one-half the difference in the average temperature effects at the two levels of pressure. Mathematically, this is

$$P \times T = d_{PT+} - d_{PT-} = 1/4[(d_1 + d_4 + d_5 + d_8) - (d_2 + d_3 + d_6 + d_7)] = 6.89 \quad (26)$$

The $P \times F$ and $T \times F$ interactions are obtained in a similar fashion. Finally, we might also be interested in the interaction of all three factors, denoted as the *pressure-by-temperature-by-flow rate* or the $P \times T \times F$ interaction. This interaction defines the average difference between any two-factor interaction at the high and low levels of the third factor. It is given by

$$P \times T \times F = d_{PTF+} - d_{PTF-} = -5.88 \quad (27)$$

It is important to note that the main effect of any factor can be individually interpreted only if there is no evidence that the factor interacts with other factors.

The Yates Algorithm

It can be tedious to calculate the effects and interactions for two-level factorial experiments using the method just described, particularly if there are more than three factors involved. Fortunately, the *Yates algorithm* provides a quicker method of computation that is also relatively easily programmed via computer. To implement this algorithm, the experimental design matrix is first arranged in what is called *standard order*. A 2^n factorial design is in standard order when the first column of the design matrix consists of alternating minus and plus signs, the second column consists of successive pairs of minus and plus signs, the third column consists of four minus signs followed by four plus signs, and so on. In general, the k th column consists of 2^{k-1} minus signs followed by 2^{k-1} plus signs.

The Yates calculations for the deposition rate data are shown in Table 10.7. Column y contains the deposition rates for each run. These are considered in successive pairs. The first four entries in column (1) are obtained by adding the pairs together, and the next four are obtained by subtracting the top number from the bottom number of each pair. Column (2) is obtained from column (1) in the same way, and column (3) is obtained from column (2). To obtain the experimental effects, one need only divide the column

(3) entries by the Divisor column entries. In general, the first divisor will be 2^n , and the remaining divisors will be 2^{n-1} . The first element in the Identification (ID) column is the grand average of all of the observations, and the remaining identifications are derived by locating the plus signs in the design matrix.

Although the Yates algorithm provides a relatively straightforward methodology for computing experimental effects, it should be pointed out that modern analysis of statistical experiments is accomplished almost exclusively by commercially available statistical software packages. A few of the more common packages include RS/1, SAS, and Minitab. These packages completely alleviate the necessity of performing any tedious hand calculations.

Fractional Factorial Designs

A disadvantage of the two-level factorial design is that the number of experimental runs increases exponentially with the number of factors. To alleviate this concern, *fractional factorial* designs are constructed by systematically eliminating some of the runs in a full factorial design. For example, a half fractional design with n factors requires only 2^{n-1} runs. Full or fractional two-level factorial designs can be used to estimate the main effects of individual factors as well as the interaction effects between factors. However, they cannot be used to estimate quadratic or higher-order effects. This is not a serious shortcoming, since higher-order effects and interactions tend to be smaller than low-order effects (i.e., main effects tend to be larger than two-factor interactions, which tend to be larger than three-factor interactions, etc.). Ignoring high-order effects is conceptually similar to ignoring higher-order terms in a Taylor series expansion.

To illustrate the use of fractional factorial designs, let $n = 5$ and consider a 2^5 factorial design. The full factorial implementation of this design would require 32 experimental runs. However, a 2^{5-1} fractional factorial design only requires 16 runs. This 2^{5-1} design is generated by first writing the design matrix for a 2^4 full factorial design in standard order. Then plus and minus signs in the four columns of the 2^4 design matrix are each "multiplied" together to form a fifth column.

For example, let's take another look at our CVD experiment. Suppose we only have the time or resources available to perform four deposition experiments, rather than the eight required for a 2^3 full factorial design. This calls for a 2^{3-1} fractional factorial alternative. This new design could be generated by writing the full 2^3 design for the pressure and temperature variables, and then multiplying those columns to obtain a third column for flow rate. This procedure is illustrated in Table 10.8. The only drawback in using this procedure is that since we have used the PT relation to define column F , we can no longer distinguish between the effects of the $P \times T$ interaction and the F main effect. When this occurs, the two effects are said to be *confounded*.

TABLE 10.7 Illustration of the Yates Algorithm

P	T	F	y	(1)	(2)	(3)	Divisor	Effect	ID
-	-	-	94.8	205.76	675.70	1543.0	8	192.87	Avg
+	-	-	110.96	469.94	867.29	163.45	4	40.86	P
-	+	-	214.12	240.06	57.86	651.35	4	162.84	T
+	+	-	255.82	627.23	105.59	27.57	4	6.89	PT
-	-	+	94.14	16.16	264.18	191.59	4	47.90	F
+	-	+	145.92	41.70	387.17	47.73	4	11.93	PF
-	+	+	286.71	51.78	25.54	122.99	4	30.75	TF
+	+	+	340.52	53.81	2.03	-23.51	4	-5.88	PTF

TABLE 10.8 Illustration of 2^{3-1} Fractional Factorial Design for CVD Example

Run	P	T	F
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

► 10.5 YIELD

Variability in IC manufacturing processes can lead to deformations or nonconformities in finished products. Such process disturbances often result in *faults*, or unintentional changes in the performance or conformance of electronic products. The presence of such faults is quantified by the manufacturing yield. *Yield* is defined as the percentage of devices or circuits that meet a nominal performance specification.

Yield can be categorized as either *functional* or *parametric*. Functional yield is determined by the proportion of fully functional products. Often referred to as hard yield, the functional yield of ICs is usually characterized by open circuits or short circuits caused by physical defects (such as particles). In some cases, however, a fully functional product still fails to meet performance specifications for one or more parameters (such as speed, noise level, or power consumption). These situations are described by parametric yield (or soft yield).

10.5.1 Functional Yield

The development of models to estimate the functional yield of ICs is fundamental to manufacturing. A model that provides accurate estimates of manufacturing yield can help predict product cost, determine optimum equipment utilization, or be used as a metric against which actual measured manufacturing yields can be evaluated. Yield models are also critical to support decisions involving new technologies and the identification of problematic products or processes.

As previously mentioned, functional yield is significantly affected by the presence of defects. Defects can result from many random sources, including contamination from equipment, processes or handling, mask imperfections, and airborne particles. Physically, these defects include shorts, opens, misalignment, photoresist splatters and flakes, pinholes, scratches, and crystallographic flaws. This is illustrated by Figure 10.21.

Yield models are usually presented as a function of the average number of defects per unit area (D_0) and the *critical area* (A_c) of the electronic system. In other words,

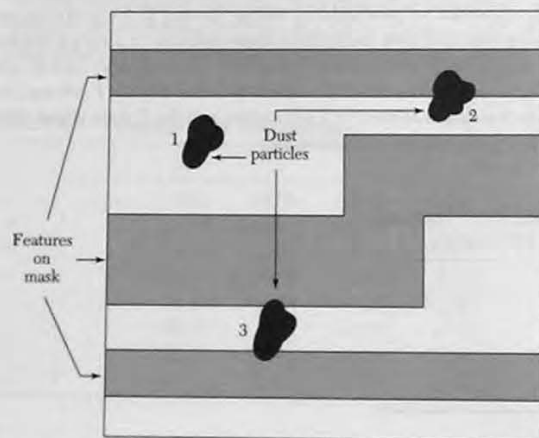


Figure 10.21 Various ways in which dust particles can interfere with interconnect mask patterns.

$$Y = f(A_c, D_0) \quad (28)$$

where Y is the functional yield. The critical area is the area in which a defect occurring has a high probability of resulting in a fault. For example, if particle 3 in Figure 10.21 is large enough and conductive, it has fallen into an area in which it causes a short between the two metal lines it bridges. The relationship between the yield, defect density, and critical area is complex. It depends on the circuit geometry, the density of photolithographic patterns, the number of photolithography steps used in the manufacturing process, and other factors. A few of the more common models that attempt to quantify this relationship are described next.

Poisson Model

The Poisson yield model assumes that defects are uniformly distributed across a substrate, and that each defect results in a fault. Pineda de Gyvez provides an excellent derivation of this model.⁷ Let C be the number of circuits on a substrate (i.e., the number of ICs), and let M be the number of possible defect types. Under these conditions, there are C^M unique ways in which the M defects can be distributed on the C circuits. For example, if there are three circuits ($C1$, $C2$, and $C3$) and three defect types (such as $M1$ = metal open, $M2$ = metal short, and $M3$ = metal 1 to metal 2 short, for example), then there are

$$C^M = 3^3 = 27 \quad (29)$$

possible ways in which these three defects can be distributed over three chips. These combinations are illustrated in Table 10.9.

If one circuit is removed (i.e., is found to contain no defects), the number of ways to distribute the M defects among the remaining circuits is

$$(C - 1)^M \quad (30)$$

Thus, the probability that a circuit will contain zero defects of any type is

$$\frac{(C - 1)^M}{C^M} = \left(1 - \frac{1}{C}\right)^M \quad (31)$$

TABLE 10.9 Truth Table for Unique Fault Combinations

Combination	C1	C2	C3	Combination	C1	C2	C3
1	M1M2M3			15	M3		M2M1
2		M1M2M3		16		M1M2	M3
3			M1M2M3	17		M1M3	M2
4	M1M2	M3		18		M2M3	M1
5	M1M3	M2		19		M1	M2M3
6	M2M3	M1		20		M2	M1M3
7	M1M2		M3	21		M3	M2M1
8	M1M3		M2	22	M1	M2	M3
9	M2M3		M1	23	M1	M3	M2
10	M1	M2M3		24	M2	M1	M3
11	M2	M1M3		25	M2	M3	M1
12	M3	M2M1		26	M3	M1	M2
13	M1		M2M3	27	M3	M2	M1
14	M2		M1M3				

Substituting $M = CA_c D_0$, the yield is the number of circuits with zero defects, or

$$Y = \lim_{C \rightarrow \infty} \left(1 - \frac{1}{C}\right)^{CA_c D_0} = \exp(-A_c D_0) \quad (32)$$

For N circuits to have zero defects, this becomes

$$Y = \exp(-A_c D_0)^N = \exp(-NA_c D_0) \quad (33)$$

The Poisson model is simple and relatively easy to derive. It provides a reasonably good estimate of yield when the critical area is small. However, if D_0 is calculated based on small-area circuits, using the same D_0 for large-area yield computations results in a yield estimate that is overly pessimistic compared with actual measured data.

Murphy's Yield Integral

B. T. Murphy first proposed that the value of the defect density (D) should not be constant.⁸ Instead, he reasoned that D must be summed over all circuits and substrates using a normalized probability density function, $f(D)$. The yield can then be calculated using the integral

$$Y = \int_0^{\infty} e^{-A_c D} f(D) dD \quad (34)$$

Various forms of $f(D)$ form the basis for the differences between many analytical yield models. The Poisson model assumes that $f(D)$ is a delta function, that is,

$$f(D) = \delta(D - D_0) \quad (35)$$

where D_0 is the average defect density as before (see Fig. 10.22a). Using this density function, the yield is determined from Eq. 34 to be

$$Y_{\text{Poisson}} = \int_0^{\infty} e^{-A_c D} f(D) dD = \exp(-A_c D_0) \quad (36)$$

as shown before.

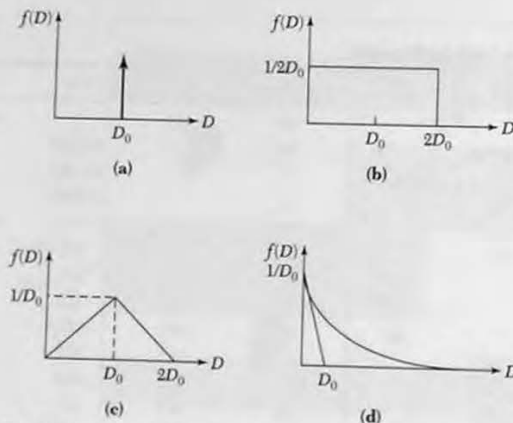


Figure 10.22 Probability density functions for (a) the Poisson model, (b) the uniform Murphy model, (c) the triangular Murphy model, and (d) the exponential Seeds model.⁷

Murphy initially investigated a uniform density function, as shown in Figure 10.22b. The evaluation of the yield integral for the uniform density function gives

$$Y_{\text{uniform}} = \frac{1 - e^{-2D_0 A_c}}{2D_0 A_c} \quad (37)$$

Murphy later believed that a Gaussian distribution would be a better reflection of the true defect density distribution than the delta function. However, since he was unable to integrate the yield integral with a Gaussian function substituted for $f(D)$, he approximated it using the triangular function shown in Figure 10.22c. This function results in the yield expression

$$Y_{\text{triangular}} = \left(\frac{1 - e^{-D_0 A_c}}{2D_0 A_c} \right)^2 \quad (38)$$

The triangular Murphy yield model is widely used today in industry to determine the effect of manufacturing process defect density.

R. B. Seeds was the first to verify Murphy's predictions.⁹ However, Seeds theorized that high yields were caused by a large population of low defect densities (which are not high enough to cause faults) and a small proportion of high defect densities (i.e., high enough to cause faults). He therefore proposed the exponential density function given by

$$f(D) = \frac{1}{D_0} \exp\left(\frac{-D}{D_0}\right) \quad (39)$$

and shown in Figure 10.22d. This function implies that the probability of observing a low defect density is significantly higher than that of observing a high defect density. Substituting this exponential function in the Murphy integral and integrating yields

$$Y_{\text{exponential}} = \frac{1}{1 + D_0 A_c} \quad (40)$$

Although the Seeds model is simple, its yield predictions for large-area substrates are too optimistic. Therefore, this model has not been widely used.

Okabe, Nagata, and Shimada recognized the physical nature of defect distributions and proposed the gamma probability density function.¹⁰ Stapper has likewise developed and applied yield models using the gamma density function.¹¹ The gamma distribution is given by

$$f(D) = [\Gamma(\alpha)\beta^\alpha]^{-1} D^{\alpha-1} e^{-D/\beta} \quad (41)$$

where α and β are two parameters of the distribution, and $\Gamma(\alpha)$ is the gamma function. The shape of $\Gamma(\alpha)$ is shown for several values of α in Figure 10.23. In this distribution, the average defect density is $D_0 = \alpha\beta$.

The yield model derived by substituting Eq. 41 into Murphy's integral is

$$Y_{\text{gamma}} = \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha} \quad (42)$$

This model is commonly referred to as the *negative binomial* model. The parameter α must be empirically determined. It is generally called the cluster parameter because it increases with decreasing variance in the distribution of defects. If α is high, then the variability of defects is low (little clustering). Under these conditions, the gamma density function approaches a delta function, and the negative binomial model reduces to the Poisson model. Mathematically, this means

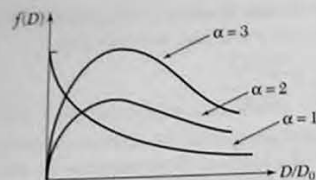


Figure 10.23 Probability density function for the gamma distribution.

$$Y = \lim_{\alpha \rightarrow \infty} \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha} = \exp(-A_c D_0) \quad (43)$$

If α is low, on the other hand, the variability of defects across the wafer is significant (much clustering), and the gamma model reduces to the Seeds exponential model, or

$$Y = \lim_{\alpha \rightarrow 0} \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha} = \frac{1}{1 + A_c D_0} \quad (44)$$

If the critical area and defect density are known (or can be accurately measured), the negative binomial model is an excellent general-purpose yield predictor that can be used for a variety of IC manufacturing processes.

10.5.2 Parametric Yield

Even in a defect-free manufacturing environment, random processing variations can lead to varying levels of system performance. These variations result from the fluctuation of numerous physical and environmental parameters (linewidths, film thicknesses, ambient humidity, etc.), which in turn manifest themselves as variations in final system performance (such as speed or noise level). These performance variations lead to "soft" faults and are characterized by the *parametric yield* of the manufacturing process. Parametric yield is a measure of the quality of functioning systems, whereas functional yield measures the proportion of functioning units produced by the manufacturing process.

A common method used to evaluate parametric yield is *Monte Carlo simulation*. In the Monte Carlo approach, a large number of pseudo-random sets of values for circuit or system parameters are generated according to an assumed probability distribution (usually the normal distribution) based on sample means and standard deviations extracted from measured data. For each set of parameters, a simulation is performed to obtain information about the predicted behavior of a circuit or system. The overall performance distribution is then extracted from the set of simulation results.

To illustrate the Monte Carlo technique, consider as a performance metric the drive current of an n -channel MOSFET in saturation (I_{Dsat}). It can be shown that¹²

$$I_{Dsat} \equiv \left(\frac{Z\mu_n C_o}{2L}\right)(V_G - V_T)^2 \quad (45)$$

where Z is the width of the device, L is its length, μ_n is the electron mobility in the channel, C_o is the oxide capacitance per unit area, V_G is the applied gate voltage, and V_T is

the threshold voltage of the transistor. In this equation, C_o is a function of the thickness of the oxide (d), and V_T is a function of the oxide thickness as well as the doping in the channel, or $I_{Dsat} = f(C_o, V_T)$. Both of these dimensions are subject to manufacturing process variations. They can thus be characterized as varying according to normal distributions with means μ_c and μ_v and standard deviations σ_c and σ_v , respectively (see Fig. 10.24).

Using the Monte Carlo approach, we can estimate the parametric yield of MOSFETs produced by a given manufacturing process within a certain range of saturation drain currents by computing the value of I_{Dsat} for every possible combination of C_o and V_T . The result of these computations is a final performance distribution like the one shown in Figure 10.24b. This probability density function can then be used to compute the proportion of transistors having a given range of drive currents. For example, if we wanted to compute the percentage of MOSFETs manufactured that would have a value of I_{Dsat} between two limits a and b , we would evaluate the integral

$$Y(\text{MOSFETs with } a < I_{Dsat} < b) = \int_a^b f(x) dx \quad (46)$$

Thus, once the overall distribution of a given output metric is known, it is possible to estimate the fraction of manufactured parts with any range of performance. Estimation of parametric yield is useful for circuit designers because it helps identify the limits of the manufacturing process to facilitate and encourage design for manufacturability.

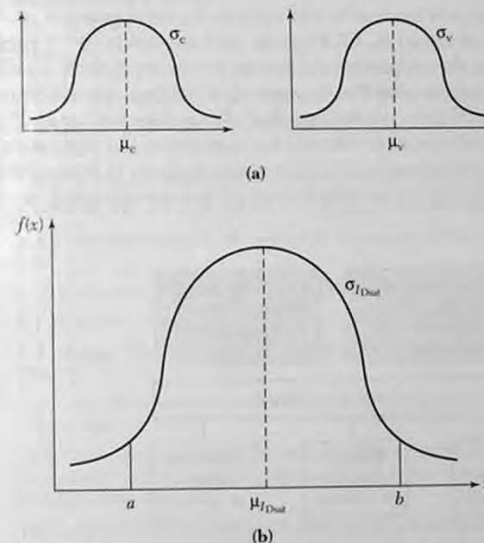


Figure 10.24 (a) Normal probability density functions for C_o and V_T . (b) Overall probability density function for I_{Dsat} .

► 10.6 COMPUTER-INTEGRATED MANUFACTURING

The vast majority of quantitative evaluation of IC manufacturing processes is accomplished via computers rather than by hand calculations. Not only is this more efficient, but in today's economy, it is essential. The fabrication of integrated circuits can be quite expensive. In fact, the last decade has seen electronics manufacturing become so capital intensive that small companies often find it too expensive to support their own manufacturing operations. A typical state-of-the-art high-volume manufacturing facility today costs several orders of magnitude more than a comparable facility 20 years ago. This has led to the rise of a large contract manufacturing industry.

As a result of rising costs, the challenge before manufacturers today is to offset such large capital investment with a greater amount of technological innovation in the fabrication process. In other words, the objective now is to make use of the latest developments in computer hardware and software technology to enhance manufacturing methods that have become prohibitively expensive. In effect, this effort in *computer-integrated manufacturing* of integrated circuits (IC-CIM) is aimed at optimizing the cost-effectiveness of electronics manufacturing in the same manner in which *computer-aided design* (CAD) has dramatically affected the economics of circuit design.

Under the overall heading of reducing manufacturing cost, several subtasks have been identified. These include increasing fabrication yield, reducing product cycle time, maintaining consistent levels of product quality and performance, and improving the reliability of processing equipment. Since fabrication processes often consist of hundreds of sequential steps, yield loss may potentially occur at every step. Consequently, maintaining product quality in an electronics manufacturing facility requires the strict control of literally hundreds or even thousands of process variables. The interdependent issues of high yield, high quality, and low cycle time can be addressed by the development of several critical capabilities in a state-of-the-art IC-CIM system: work-in-process (WIP) monitoring, equipment communication, data acquisition and storage, process/equipment modeling, and real-time process control, to name a few. The emphasis of each of these activities is to increase throughput and reduce yield loss by preventing potential misprocessing, but each presents significant engineering challenges in its effective implementation and deployment.

A block diagram of a typical modern IC-CIM system is shown in Figure 10.25. This diagram outlines many of the key features required for efficient manufacturing operations.¹³

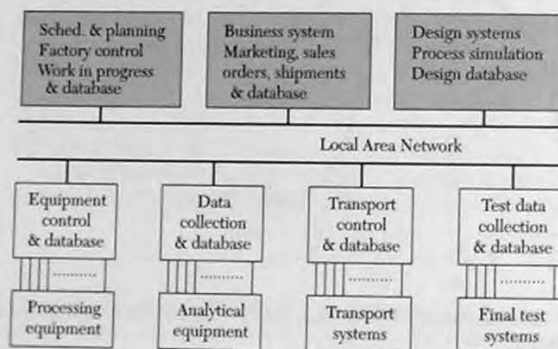


Figure 10.25 Two-level CIM architecture.¹³

The lower level of this two-level architecture includes embedded controllers that provide real-time control and analysis of fabrication equipment. These controllers often consist of personal computers and the associated control software dedicated to each individual piece of equipment. The second level of this IC-CIM architecture is composed of a distributed local area network of computer workstations and file servers linked by a common distributed database. Equipment communication with host computers is facilitated by an electronics manufacturing standard called the *generic equipment model* (GEM). The GEM standard is used in both semiconductor manufacturing and printed circuit board assembly. This standard is based on the *semiconductor equipment communications standard* (SECS) protocol.

This type of IC-CIM architecture has great flexibility, allowing extension and adaptation to meet constantly changing requirements. Over the past several years, powerful, flexible, and cost-effective information systems based on models such as this have become an integral part of the IC manufacturing enterprise.

► 10.7 SUMMARY

This chapter provided an overview of the relevant issues in IC manufacturing. This included a description of electrical testing and basic packaging processes, as well as a presentation of statistical process control, statistical experimental design, and yield modeling. The chapter concluded with a brief introduction to IC-CIM systems. In IC manufacturing, process and equipment reliability directly influence throughput, yield, and ultimately cost. Over the next several years, significant enhancement of manufacturing operations will be required to reach projected targets for future generations of microelectronic devices, packages, and systems.

► REFERENCES

1. Pineda de Gyvez and D. Pradhan, *Integrated Circuit Manufacturability*, IEEE Press, Piscataway, NJ, 1999.
2. A. Landzberg, *Microelectronics Manufacturing Diagnostics Handbook*, Van Nostrand Reinhold, New York, 1993.
3. R. Tummala, Ed., *Fundamentals of Microsystems Packaging*, McGraw-Hill, New York, 2001.
4. W. Brown, Ed., *Advanced Electronic Packaging*, IEEE Press, New York, 1999.
5. R. Jaeger, *Introduction to Microelectronic Fabrication*, 2nd Ed., Prentice-Hall, Upper Saddle River, NJ, 2002.
6. D. Montgomery, *Introduction to Statistical Quality Control*, Wiley, New York, 1985.
7. J. Pineda de Gyvez and D. Pradhan, *Integrated Circuit Manufacturability*, IEEE Press, New York, 1999.
8. B. Murphy, "Cost-Size Optima of Monolithic Integrated Circuits," *Proc. IEEE*, **52**, (12), 1537-1545 (1964).
9. R. Seeds, "Yield and Cost Analysis of Bipolar LSI," *IEEE Int. Electron Devices Meet.*, Washington, DC, October 1967.
10. T. Okabe, M. Nagata, and S. Shimada, "Analysis of Yield of Integrated Circuits and a New Expression for the Yield," in C. Strapper, Ed., *Defect and Fault Tolerance in VLSI Systems*, Vol. 2, Plenum Press, New York, pp. 47-61, 1990.
11. C. Stapper, "Fact and Fiction in Yield Modeling," *Microelectronics J.*, **210**, 129-151 (1989).
12. S. Sze, *Semiconductor Devices: Physics and Technology*, 2nd Ed., Wiley, New York, 2002.
13. D. Hodges, L. Rowe, and C. Spanos, "Computer Integrated Manufacturing of VLSI," Proceedings of the 11th IEEE/CHMT International Electronics Manufacturing Technology Symposium, September 1989, pp. 1-3.

► PROBLEMS

Asterisks denote difficult problems.

SECTION 10.3: STATISTICAL PROCESS CONTROL

- Control charts for $\bar{x}$ and s are to be maintained for the threshold voltage of short-channel MOSFETs using sample sizes of $n = 10$. It is known that the process is normally distributed with $\mu = 0.75$ V and $\sigma = 0.10$ V. Find the center line and control limits for each of these control charts.
- Repeat Problem 1 assuming that μ and σ are unknown and that we have collected 50 observations of sample size 10. These samples yielded a grand average of 0.734 V and an average s_i of 0.125 V.

SECTION 10.4: STATISTICAL EXPERIMENTAL DESIGN

- The following 2^3 factorial experiment was used to analyze a photolithography process. Analyze the experimental results using the Yates algorithm.

Run	Exposure Dose	Develop Time	Bake Temperature	Yield (%)
1	—	—	—	60
2	+	—	—	77
3	—	+	—	59
4	+	+	—	68
5	—	—	+	57
6	+	—	+	83
7	—	+	+	45
8	+	+	+	85

- Consider the throughputs (i.e., wafers processed per hour) of five different manufacturing processes (labelled A to E in the following table). For each process, data was collected on three different dates. Perform an analysis of variance to determine whether the processes and processing dates are significantly different.

DAY	A	B	C	D	E
1	509	512	532	506	509
2	505	507	542	520	519
3	465	472	498	483	475

SECTION 10.5: YIELD

- Assuming a Poisson model, calculate the maximum defect density allowable on 100,000 NMOS transistors in order to achieve a functional yield of 95%. Assume the gate of each device is 10 μm wide and 1 μm long.
- Use Murphy's yield integral to derive Eqs. 37, 38, and 40.
- Suppose the probability density function of the defect density for a given interconnect manufacturing process is given by

$$f(D) = -100D + 10 \quad 0 \leq D \leq 0.1$$

If the critical area for this interconnect is 100 cm^2 , calculate the functional yield we can expect for the process over the range of defect densities from 0.05 to 0.1 cm^{-2} .

Future Trends and Challenges

Since the beginning of the integrated circuit era in 1959, the minimum device dimension, also called the *minimum feature length*, has been reduced at an annual rate of about 13% (i.e., a reduction of 30% every 3 years). According to the prediction by the *International Technology Roadmap for Semiconductors*,¹ the minimum feature length will shrink from 130 nm (0.13 μm) in the year 2002 to 35 nm (0.035 μm) around 2014, as shown in Table 11.1. Also shown in Table 11.1 is DRAM size. The DRAM has increased its memory cell capacity four times every 3 years, and 64-Gbit DRAM is expected to be available in 2011 using 50-nm design rules. The table also shows that wafer size will increase to 450 mm (18 in. diameter) in 2014. In addition to the feature size reduction, challenges come from the device level, material level, and system level, as discussed in the following sections.

► 11.1 CHALLENGES FOR INTEGRATION

Figure 11.1 shows the trends of power supply voltage (V_{DD}), threshold voltage (V_T), and gate oxide thickness (d) versus channel length for CMOS logic technology.² From this figure, one can see that the gate oxide thickness will soon approach the tunneling-current limit of 2 nm. V_{DD} scaling will slow down because of the nonscalable V_T (i.e., to a minimum V_T of about 0.3 V due to subthreshold leakage and circuit noise immunity). Some challenges of the 180-nm technology and beyond are shown in Figure 11.2.³ The most stringent requirements are detailed in the following subsections.

TABLE 11.1 The Technology Generation from 1997 to 2014

Year of the first product shipment	1997	1999	2002	2005	2008	2011	2014
Feature size (nm)	250	180	130	100	70	50	35
DRAM size (bit)	256M	1 G	—	8 G	—	64 G	—
Wafer size (mm)	200	300	300	300	300	300	450
Gate oxide (nm)	3–4	1.9–2.5	1.3–1.7	0.9–1.1	<1.0	—	—
Junction depth (nm)	50–100	42–70	25–43	20–33	15–30	—	—

DRAM, dynamic random access memory.

Source: *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, San Jose, CA, 1999.

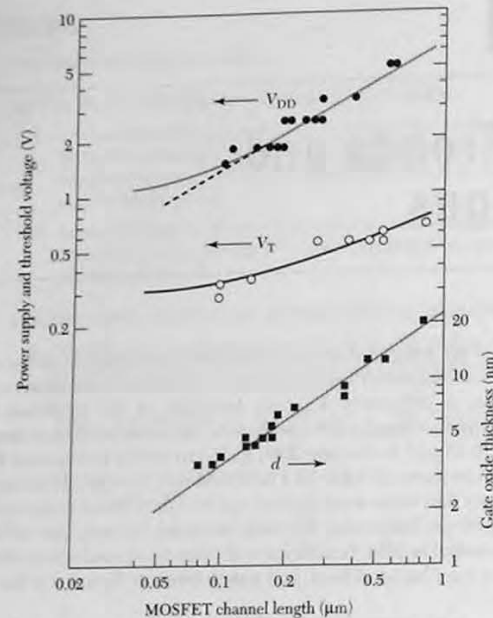


Figure 11.1 Trends of power supply voltage V_{DD} , threshold voltage V_T , and gate oxide thickness d versus channel length for CMOS logic technologies. Points are collected from data published over recent years.²

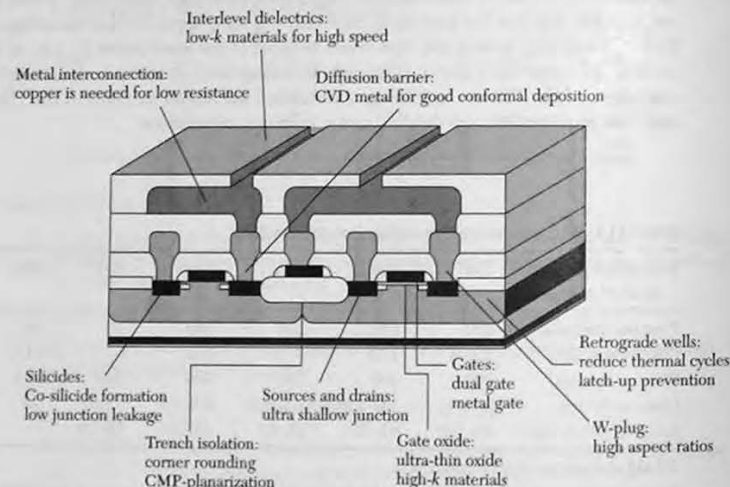


Figure 11.2 Challenges for 180-nm and smaller MOSFETs.³

11.1.1 Ultrashallow Junction Formation

So-called short-channel effects happen as the channel length is reduced. This problem becomes critical as the device dimension is scaled down to 100 nm. To achieve an ultrashallow junction with low sheet resistance, low-energy (i.e., less than 1 keV) implantation technology with high dosage must be employed. Table 11.1 shows the required junction depth versus the technology generation. The requirements of the junction for 100 nm are depths around 20 to 33 nm with a doping concentration of $1 \times 10^{20}/\text{cm}^3$.

11.1.2 Ultrathin Oxide

As the gate length shrinks below 130 nm, the oxide equivalent thickness of the gate dielectric must be reduced around 2 nm to maintain performance. However, if only SiO_2 (with a dielectric constant of 3.9) is used, the leakage through the gate becomes very high because of direct tunneling. For this reason, thicker high- k dielectric materials that have lower leakage current are needed. Candidates for the short term are silicon nitride (with a dielectric constant of 7), Ta_2O_5 (25), and TiO_2 (60–100).

11.1.3 Silicide Formation

Silicide-related technology has become an integral part of submicron devices for reducing the parasitic resistance to improve device and circuit performance. The conventional Ti-silicide process has been widely used in 350- to 250-nm technology. However, the sheet resistance of a TiSi_2 line increases with decreasing linewidth, which limits the use of TiSi_2 in 100-nm CMOS applications and beyond. CoSi_2 or NiSi processes will replace TiSi_2 for technology beyond 100 nm.

11.1.4 New Materials for Interconnection

To achieve high-speed operation, the RC time delay of the interconnection must be reduced. Figure 8.14 showed the delay as a function of feature size.⁴ It is obvious that gate delay decreases as the channel length decreases. Meanwhile, the delay resulting from interconnect increases significantly as the size decreases. This causes the total delay time to increase as the dimension of the device size scales down to 250 nm. Consequently, both high-conductivity metals, such as Cu, and low-dielectric-constant (low- k) insulators, such as organic (polyimide) or inorganic (F-doped oxide) materials, offer major performance gains. Cu exhibits superior performance because of its high conductivity ($1.7 \mu\Omega\text{-cm}$ compared with $2.7 \mu\Omega\text{-cm}$ for Al) and is 10 to 100 times more resistant to electromigration. The delay using Cu and the low- k material shows a significant decrease compared with that of conventional Al and oxide. Hence, Cu with the low- k material is essential in multilevel interconnection for future deep-submicron technology.

11.1.5 Power Limitations

The power required merely to charge and discharge circuit nodes in an IC is proportional to the number of gates and the frequency at which they are switched (clock frequency). The power can be expressed as $P \approx 1/2 CV^2 nf$, where C is the capacitance per device, V is the applied voltage, n is the number of devices per chip, and f is the clock frequency. The temperature rise caused by this power dissipation in an IC package is limited by the thermal conductivity of the package material, unless auxiliary liquid or gas cooling is used. The maximum allowable temperature rise is limited by the bandgap of the semiconductor

($\sim 100^\circ\text{C}$ for Si with a bandgap of 1.1 eV). For such a temperature rise, the maximum power dissipation of a typical high-performance package is about 10 W. As a result, we must limit either the maximum clock rate or the number of gates on a chip. As an example, in an IC containing 100-nm MOS devices with $C = 5 \times 10^{-2}$ fF, running at a 20 GHz clock rate, the maximum number of gates we can have is about 10^7 if we assume a 10% duty cycle. This is a design constraint fixed by basic material parameters.

11.1.6 SOI Integration

Section 9.2.2 mentioned the isolation of SOI wafers. Recently, SOI technology has received a great deal of attention. The advantages of SOI integration become significant as the minimum feature length approaches 100 nm. From a process point of view, SOI does not need the complex well structure and isolation processes. In addition, shallow junctions are directly obtained through the SOI film thickness. There is no risk of nonuniform interdiffusion of silicon and Al in the contact regions because of oxide isolation at the bottom of the junction. Hence, the contact barrier is not necessary. From a device point of view, the modern bulk silicon device needs high doping at the drain and substrate to eliminate short-channel effects and punchthrough. This high doping results in high capacitance when the junction is reverse biased. On the other hand, in SOI, the maximum capacitance between the junction and substrate is the capacitance of the buried insulator, whose dielectric constant is three times smaller than that of silicon (3.9 versus 11.9). Based on ring oscillator performance, 130-nm SOI CMOS technology can achieve 25% faster speeds or require 50% less power compared with a similar bulk technology.⁵ SRAM, DRAM, CPU, and rf CMOS have all been successfully fabricated using SOI technology. Therefore, SOI is a key candidate for the future system-on-a-chip technology, considered in the following section.

EXAMPLE 1

For an equivalent oxide thickness of 1.5 nm, what will be the physical thickness when the high- k materials nitride ($\epsilon_r/\epsilon_0 = 7$), Ta_2O_5 (25), or TiO_2 (80) are used?

SOLUTION For nitride,

$$\left(\frac{\epsilon_r}{1.5}\right) = \left(\frac{\epsilon_{\text{nitride}}}{d_{\text{nitride}}}\right)$$

$$d_{\text{nitride}} = 1.5 \left(\frac{7}{3.9}\right) = 2.69 \text{ nm}$$

Using the same calculation, we obtain 9.62 nm for Ta_2O_5 and 10.77 nm for TiO_2 .

► 11.2 SYSTEM-ON-A-CHIP

Increased component density and improved fabrication technology have helped the realization of the *system-on-a-chip* (SOC), that is, an IC chip that contains a complete electronic system. Designers can build all the circuitry needed for a complete electronic system, such as a camera, radio, television, or personal computer (PC), on a single chip. Figure 11.3 shows an SOC application for a PC's motherboard. Components (11 chips in this case) once found on boards become virtual components on the chip at the right.⁶

There are two obstacles in the realization of the SOC. The first is the huge complexity of the design. Since the component board is presently designed by different companies

A/MS = analog/mixed signal
ASIC = application-specific IC
CPU = central processing unit
PLD = programmable logic device

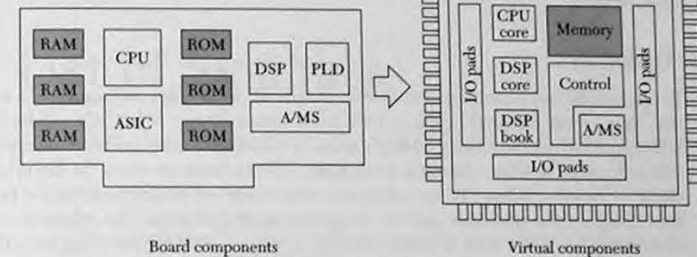


Figure 11.3 System-on-a-chip of a conventional personal computer motherboard.⁶

using different design tools, it is difficult to integrate the components into one chip. Another difficulty lies in fabrication. In general, the fabrication processes of a DRAM are significantly different from those of logic ICs (e.g., CPUs). Speed is the first priority for the logic, whereas leakage of the stored charge is the priority for memory. Therefore, multilevel interconnection schemes using five to six levels of metal are essential for logic ICs to improve the speed. However, DRAM circuits need only two to three levels. In addition, to increase the speed, a silicide process must be used to reduce the series resistance, and ultrathin gate oxide is needed to increase the drive current in logic circuits. These requirements are not critical for memory.

To achieve the SOC goal, an embedded DRAM technology has been introduced to merge logic and DRAM into a single chip with compatible processes. Figure 11.4 shows a schematic cross section of an embedded DRAM, including the DRAM cells and the logic CMOS devices.⁷ Some processing steps are modified as a compromise. The trench-type

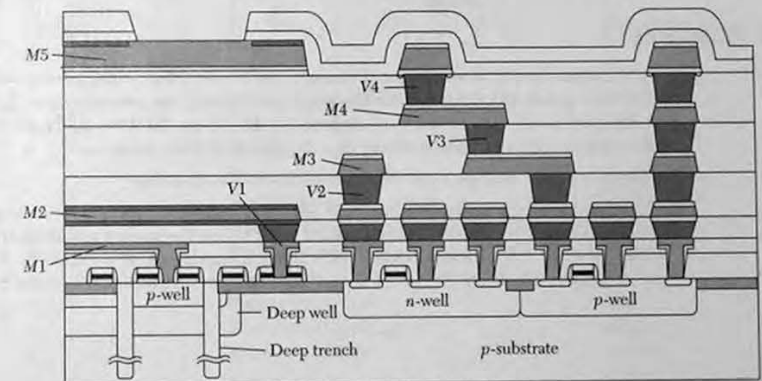


Figure 11.4 Schematic cross section of the embedded DRAM including DRAM cells and logic MOSFETs. There is no height difference in the trench capacitor cell because of the DRAM cell structure. M1 to M5 are metal interconnections, and V1 to V4 are via holes.⁷

capacitor, instead of the stacked type, is used so that there is no height difference in the DRAM cell structure. In addition, multiple gate oxide thicknesses exist on the same wafer to accommodate multiple supply voltages or to combine memory and logic circuits on one chip, or both.

► 11.3 SUMMARY

Because of the rapid reduction in feature length, IC technology will soon reach its practical limit as the channel length is reduced to about 20 nm. What ICs will be beyond CMOS is a key question being asked by research scientists. Major candidates include innovative devices based on quantum mechanical effects, because when the lateral dimension is reduced to below 100 nm, electronic structures will exhibit nonclassical behavior depending on the materials and the temperature of operation. The operation of such devices will be on the scale of single-electron transport. This approach has been demonstrated by the single-electron memory cell. The realization of such systems with trillions of components will be a major challenge beyond CMOS.⁸

► REFERENCES

1. *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, San Jose, 1999.
2. Y. Taur and E. J. Nowak, "CMOS Devices below 0.1 μm : How High Will Performance Go?" *IEEE Tech. Dig. Int. Electron Devices Meet.*, p. 215 (1997).
3. L. Peters, "Is the 0.18 μm Node Just a Roadside Attraction?" *Semicond. Int.*, 22, 46 (1999).
4. M. T. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *IEEE Tech. Dig. Int. Electron Devices Meet.*, p. 241 (1995).
5. E. Leobandung, et al., "Scalability of SOI Technology into 0.13 μm 1.2 V CMOS Generation," *IEEE Int. Electron Devices Meet.*, p. 403 (1998).
6. B. Martin, "Electronic Design Automation," *IEEE Spectr.*, 36, 61 (1999).
7. H. Ishiuchi, et al., "Embedded DRAM Technologies," *IEEE Tech. Dig. Int. Electron Devices Meet.*, p. 33 (1997).
8. S. Luryi, J. Xu, and A. Zaslavsky, Eds., *Future Trends in Microelectronics*, Wiley, New York, 1999.

► PROBLEMS

1. (a) Calculate the RC time constant of an aluminum runner 0.5 μm thick formed on a thermally grown SiO_2 0.5 μm thick. The length and width of the runner are 1 cm and 1 μm , respectively. The resistivity of the runner is $10^{-5} \Omega\text{-cm}$. (b) What will be the RC time constant for a polysilicon runner ($R_s = 30 \Omega/\square$) of identical dimension?
2. Why do we need multiple oxide thicknesses for a system-on-a-chip?
3. Normally we need a buffered layer placed between a high- k Ta_2O_5 and the silicon substrate. Calculate the effective oxide thickness (EOT) when the stacked gate dielectric is Ta_2O_5 ($k = 25$) with a thickness of 75 $\text{\AA}$ on a buffered nitride layer ($k = 7$ and a thickness of 10 $\text{\AA}$). Also calculate EOT for a buffered oxide layer ($k = 3.9$ and a thickness of 5 $\text{\AA}$).

Appendix A

List of Symbols

Symbol	Description	Unit
a	Lattice constant	$\text{\AA}$
c	Speed of light in vacuum	cm/s
C	Capacitance	F
D	Diffusion coefficient	cm^2/s
E	Energy	eV
$\mathcal{E}$	Electric field	V/cm
f	Frequency	Hz(cps)
h	Planck constant	J $\cdot$ s
I	Current	A
J	Current density	A/cm^2
k	Boltzmann constant	J/K
L	Length	cm or μm
m_0	Electron rest mass	kg
$\bar{n}$	Refractive index	
n	Density of free electrons	cm^{-3}
n_i	Intrinsic carrier concentration	cm^{-3}
p	Density of free holes	cm^{-3}
P	Pressure	Pa
q	Magnitude of electronic charge	C
Q_{it}	Interface trapped charge	charges/ cm^2
R	Resistance	Ω
t	Time	s
T	Absolute temperature	K
v	Carrier velocity	cm/s
V	Voltage	V
ϵ_0	Permittivity in vacuum	F/cm
ϵ_s	Semiconductor permittivity	F/cm
ϵ_{ox}	Insulator permittivity	F/cm
ϵ_s/ϵ_0 or ϵ_{ox}/ϵ_0	Dielectric constant	
λ	Wavelength	μm or nm
ν	Frequency of light	Hz
μ_0	Permeability in vacuum	H/cm
μ_n	Electron mobility	$\text{cm}^2/\text{V}\cdot\text{s}$
μ_p	Hole mobility	$\text{cm}^2/\text{V}\cdot\text{s}$
ρ	Resistivity	$\Omega\text{-cm}$
Ω	Ohm	Ω

Appendix B

International System of Units (SI Units)

Quantity	Unit	Symbol	Dimensions
Length ^a	Meter	m	
Mass	Kilogram	kg	
Time	Second	s	
Temperature	Kelvin	K	
Current	Ampere	A	
Light intensity	Candela	Cd	
Angle	Radian	rad	
Frequency	Hertz	Hz	1/s
Force	Newton	N	kg·m/s ²
Pressure	Pascal	Pa	N/m ²
Energy ^a	Joule	J	N·m
Power	Watt	W	J/s
Electric charge	Coulomb	C	A·s
Potential	Volt	V	J/C
Conductance	Siemens	S	A/V
Resistance	Ohm	Ω	V/A
Capacitance	Farad	F	C/V
Magnetic flux	Weber	Wb	V·s
Magnetic induction	Tesla	T	Wb/m ²
Inductance	Henry	H	Wb/A
Light flux	Lumen	Lm	Cd·rad

^a It is more common in the semiconductor field to use cm for length and eV for energy (1 cm = 10⁻² m, 1 eV = 1.6 × 10⁻¹⁹ J).

Appendix C

Unit Prefixes*

Multiple	Prefix	Symbol
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^9	giga	G
10^6	mega	M
10^3	kilo	k
10^2	hecto	h
10	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

* Adopted by International Committee on Weights and Measures. (Compound prefixes should not be used, e.g., not $\mu\mu$ but p.)

Appendix D

Greek Alphabet

Letter	Lowercase	Uppercase
Alpha	α	A
Beta	β	B
Gamma	γ	Γ
Delta	δ	Δ
Epsilon	ϵ	E
Zeta	ζ	Z
Eta	η	H
Theta	θ	Θ
Iota	ι	I
Kappa	κ	K
Lambda	λ	Λ
Mu	μ	M
Nu	ν	N
Xi	ξ	Ξ
Omicron	$\omicron$	O
Pi	π	Π
Rho	ρ	P
Sigma	σ	Σ
Tau	τ	T
Upsilon	υ	Υ
Phi	ϕ	Φ
Chi	χ	X
Psi	ψ	Ψ
Omega	ω	Ω

Appendix E

Physical Constants

Quantity	Symbol	Value
Angstrom unit	$\text{\AA}$	$10 \text{ \AA} = 1 \text{ nm} = 10^{-3} \text{ }\mu\text{m} = 10^{-7} \text{ cm} = 10^{-9} \text{ m}$
Avogadro constant	N_{av}	6.02214×10^{23}
Bohr radius	a_B	$0.52917 \text{ \AA}$
Boltzmann constant	k	$1.38066 \times 10^{-23} \text{ J/K } (R/N_{\text{av}})$
Elementary charge	q	$1.60218 \times 10^{-19} \text{ C}$
Electron rest mass	m_e	$0.91094 \times 10^{-30} \text{ kg}$
Electron volt	eV	$1 \text{ eV} = 1.60218 \times 10^{-19} \text{ J}$ $= 23.053 \text{ kcal/mol}$
Gas constant	R	$1.98719 \text{ cal/mol}\cdot\text{K}$
Permeability in vacuum	μ_0	$1.25664 \times 10^{-6} \text{ H/cm } (4\pi \times 10^{-9})$
Permittivity in vacuum	ϵ_0	$8.85418 \times 10^{-14} \text{ F/cm } (1/\mu_0 c^2)$
Planck constant	h	$6.62607 \times 10^{-34} \text{ J}\cdot\text{s}$
Reduced Planck constant	$\hbar$	$1.05457 \times 10^{-34} \text{ J}\cdot\text{s } (h/2\pi)$
Proton rest mass	M_p	$1.67262 \times 10^{-27} \text{ kg}$
Speed of light in vacuum	c	$2.99792 \times 10^{10} \text{ cm/s}$
Standard atmosphere		$1.01325 \times 10^5 \text{ Pa}$
Thermal voltage at 300 K	kT/q	0.025852 V
Wavelength of 1 eV quantum	λ	$1.23984 \text{ }\mu\text{m}$

Appendix F

Properties of Si and GaAs at 300 K

Properties	Si	GaAs
Atoms/cm ³	5.02×10^{23}	4.42×10^{22}
Atomic weight	28.09	144.63
Breakdown field (V/cm)	$\sim 3 \times 10^5$	$\sim 4 \times 10^5$
Crystal structure	Diamond	Zincblende
Density (g/cm ³)	2.329	5.317
Dielectric constant	11.9	12.4
Effective density of states in conduction band, N_C (cm ⁻³)	2.86×10^{19}	4.7×10^{17}
Effective density of states in valence band, N_V (cm ⁻³)	2.66×10^{19}	7.0×10^{18}
Effective mass (conductivity)		
Electrons (m_e/m_0)	0.26	0.063
Holes (m_h/m_0)	0.69	0.57
Electron affinity, χ (V)	4.05	4.07
Energy gap (eV)	1.12	1.42
Index of refraction	3.42	3.3
Intrinsic carrier concentration (cm ⁻³)	9.65×10^9	2.25×10^6
Intrinsic resistivity (Ω -cm)	3.3×10^5	2.9×10^8
Lattice constant (Å)	3.5656	3.5959
Linear coefficient of thermal expansion, $\Delta L/L \times T$ (°C ⁻¹)	2.59×10^{-6}	5.75×10^{-6}
Melting point (°C)	1412	1240
Minority-carrier lifetime (s)	3×10^{-2}	$\sim 10^{-8}$
Mobility (cm ² /V-s)		
μ_n (electrons)	1450	9200
μ_p (holes)	505	320
Specific heat (J/g · °C)	0.7	0.35
Thermal conductivity (W/cm-K)	1.31	0.46
Vapor pressure (Pa)	1 at 1650°C 10^{-6} at 900°C	100 at 1050°C 1 at 900°C

Appendix G

Some Properties of the Error Function

w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$
0.00	0.000 000	0.32	0.349 126	0.64	0.634 586	0.96	0.825 424
0.01	0.011 283	0.33	0.359 279	0.65	0.642 029	0.97	0.829 870
0.02	0.022 565	0.34	0.369 365	0.66	0.649 377	0.98	0.834 232
0.03	0.033 841	0.35	0.379 382	0.67	0.656 628	0.99	0.838 508
0.04	0.045 111	0.36	0.389 330	0.68	0.663 782	1.00	0.842 701
0.05	0.056 372	0.37	0.399 206	0.69	0.670 840	1.01	0.846 810
0.06	0.067 622	0.38	0.409 009	0.70	0.677 801	1.02	0.850 838
0.07	0.078 858	0.39	0.418 739	0.71	0.684 666	1.03	0.854 784
0.08	0.090 078	0.40	0.428 392	0.72	0.691 433	1.04	0.858 650
0.09	0.101 281	0.41	0.437 969	0.73	0.698 104	1.05	0.862 436
0.10	0.112 463	0.42	0.447 468	0.74	0.704 678	1.06	0.866 144
0.11	0.123 623	0.43	0.456 887	0.75	0.711 156	1.07	0.869 773
0.12	0.134 758	0.44	0.466 225	0.76	0.717 537	1.08	0.873 326
0.13	0.145 867	0.45	0.475 482	0.77	0.723 822	1.09	0.876 803
0.14	0.156 947	0.46	0.484 655	0.78	0.730 010	1.10	0.880 205
0.15	0.167 996	0.47	0.493 745	0.79	0.736 103	1.11	0.883 533
0.16	0.179 012	0.48	0.502 750	0.80	0.742 101	1.12	0.886 788
0.17	0.189 992	0.49	0.511 668	0.81	0.748 003	1.13	0.889 971
0.18	0.200 936	0.50	0.520 500	0.82	0.753 811	1.14	0.893 082
0.19	0.211 840	0.51	0.529 244	0.83	0.759 524	1.15	0.896 124
0.20	0.222 703	0.52	0.537 899	0.84	0.765 143	1.16	0.899 096
0.21	0.233 522	0.53	0.546 464	0.85	0.770 668	1.17	0.902 000
0.22	0.244 296	0.54	0.554 939	0.86	0.776 110	1.18	0.904 837
0.23	0.255 023	0.55	0.563 323	0.87	0.781 440	1.19	0.907 608
0.24	0.265 700	0.56	0.571 616	0.88	0.786 687	1.20	0.910 314
0.25	0.276 326	0.57	0.579 816	0.89	0.791 843	1.21	0.912 956
0.26	0.286 900	0.58	0.587 923	0.90	0.796 908	1.22	0.915 534
0.27	0.297 418	0.59	0.595 936	0.91	0.801 883	1.23	0.918 050
0.28	0.307 880	0.60	0.603 856	0.92	0.806 768	1.24	0.920 505
0.29	0.318 283	0.61	0.611 681	0.93	0.811 564	1.25	0.922 900
0.30	0.328 627	0.62	0.619 411	0.94	0.816 271	1.26	0.925 236
0.31	0.338 908	0.63	0.627 046	0.95	0.820 891	1.27	0.927 514

(continued)

(continued)		w		$\text{erf}(w)$		w		$\text{erf}(w)$	
w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$
1.28	0.929 734	1.74	0.986 135	2.21	0.998 224	2.67	0.999 841		
1.29	0.931 899	1.75	0.986 672	2.22	0.998 308	2.68	0.999 849		
1.30	0.934 008	1.76	0.987 190	2.23	0.998 388	2.69	0.999 858		
1.31	0.936 063	1.77	0.987 691	2.24	0.998 464	2.70	0.999 866		
1.32	0.938 065	1.79	0.988 641	2.25	0.998 537	2.71	0.999 873		
1.33	0.940 015	1.80	0.989 091	2.26	0.998 607	2.72	0.999 880		
1.34	0.941 914	1.81	0.989 525	2.27	0.998 674	2.73	0.999 887		
1.35	0.943 762	1.82	0.989 943	2.28	0.998 738	2.74	0.999 893		
1.36	0.945 561	1.83	0.990 347	2.29	0.998 799	2.75	0.999 899		
1.37	0.947 312	1.84	0.990 736	2.30	0.998 857	2.76	0.999 905		
1.38	0.949 016	1.85	0.991 111	2.31	0.998 912	2.77	0.999 910		
1.39	0.950 673	1.86	0.991 472	2.32	0.998 966	2.78	0.999 916		
1.40	0.952 285	1.87	0.991 821	2.33	0.999 016	2.79	0.999 920		
1.41	0.953 852	1.88	0.992 156	2.34	0.999 065	2.80	0.999 925		
1.42	0.955 376	1.89	0.992 479	2.35	0.999 111	2.81	0.999 929		
1.43	0.956 857	1.90	0.992 790	2.36	0.999 155	2.82	0.999 933		
1.44	0.958 297	1.91	0.993 090	2.37	0.999 197	2.83	0.999 937		
1.45	0.959 695	1.92	0.993 378	2.38	0.999 237	2.85	0.999 944		
1.46	0.961 054	1.93	0.993 656	2.39	0.999 275	2.86	0.999 948		
1.47	0.962 373	1.94	0.993 923	2.40	0.999 311	2.87	0.999 951		
1.48	0.963 654	1.95	0.994 179	2.41	0.999 346	2.88	0.999 954		
1.49	0.964 898	1.96	0.994 426	2.42	0.999 379	2.89	0.999 956		
1.50	0.966 105	1.97	0.994 664	2.43	0.999 411	2.90	0.999 959		
1.51	0.967 277	1.98	0.994 892	2.44	0.999 441	2.91	0.999 961		
1.52	0.968 413	1.99	0.995 111	2.45	0.999 469	2.92	0.999 964		
1.53	0.969 516	2.00	0.995 322	2.46	0.999 497	2.93	0.999 966		
1.54	0.970 586	2.01	0.995 525	2.47	0.999 523	2.94	0.999 968		
1.55	0.971 623	2.02	0.995 719	2.48	0.999 547	2.95	0.999 970		
1.56	0.972 628	2.03	0.995 906	2.49	0.999 571	2.96	0.999 972		
1.57	0.973 603	2.04	0.996 086	2.50	0.999 593	2.97	0.999 973		
1.58	0.974 547	2.05	0.996 258	2.51	0.999 614	2.98	0.999 975		
1.59	0.975 462	2.06	0.996 423	2.52	0.999 634	2.99	0.999 976		
1.60	0.976 348	2.07	0.996 582	2.53	0.999 654	3.00	0.999 977 91		
1.61	0.977 207	2.08	0.996 734	2.54	0.999 672	3.01	0.999 979 26		
1.62	0.978 038	2.09	0.996 880	2.55	0.999 689	3.02	0.999 980 53		
1.63	0.978 843	2.10	0.997 021	2.56	0.999 706	3.03	0.999 981 73		
1.64	0.979 622	2.11	0.997 155	2.57	0.999 722	3.04	0.999 982 86		
1.65	0.980 376	2.12	0.997 284	2.58	0.999 736	3.05	0.999 983 92		
1.66	0.981 105	2.13	0.997 407	2.59	0.999 751	3.06	0.999 984 92		
1.67	0.981 810	2.14	0.997 525	2.60	0.999 764	3.07	0.999 985 86		
1.68	0.982 493	2.15	0.997 639	2.61	0.999 777	3.08	0.999 986 74		
1.69	0.983 153	2.16	0.997 747	2.62	0.999 789	3.09	0.999 987 57		
1.70	0.983 790	2.17	0.997 851	2.63	0.999 800	3.10	0.999 988 35		
1.71	0.984 407	2.18	0.997 951	2.64	0.999 811	3.11	0.999 989 08		
1.72	0.985 003	2.19	0.998 046	2.65	0.999 822	3.12	0.999 989 77		
1.73	0.985 578	2.20	0.998 137	2.66	0.999 831	3.13	0.999 990 42		

(continued)

(continued)		w		$\text{erf}(w)$		w		$\text{erf}(w)$	
w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$	w	$\text{erf}(w)$
3.14	0.999 991 03	3.36	0.999 997 983	3.58	0.999 999 587	3.80	0.999 999 923		
3.15	0.999 991 60	3.37	0.999 998 120	3.59	0.999 999 617	3.81	0.999 999 929		
3.16	0.999 992 14	3.38	0.999 998 247	3.60	0.999 999 644	3.82	0.999 999 934		
3.17	0.999 992 64	3.39	0.999 998 367	3.61	0.999 999 670	3.83	0.999 999 939		
3.18	0.999 993 11	3.40	0.999 998 478	3.62	0.999 999 694	3.84	0.999 999 944		
3.19	0.999 993 56	3.41	0.999 998 582	3.63	0.999 999 716	3.85	0.999 999 948		
3.20	0.999 993 97	3.42	0.999 998 679	3.64	0.999 999 736	3.86	0.999 999 952		
3.21	0.999 994 36	3.43	0.999 998 770	3.65	0.999 999 756	3.87	0.999 999 956		
3.22	0.999 994 73	3.44	0.999 998 855	3.66	0.999 999 773	3.88	0.999 999 959		
3.23	0.999 995 07	3.45	0.999 998 934	3.67	0.999 999 790	3.89	0.999 999 962		
3.24	0.999 995 40	3.46	0.999 999 008	3.68	0.999 999 805	3.90	0.999 999 965		
3.25	0.999 995 70	3.47	0.999 999 077	3.69	0.999 999 820	3.91	0.999 999 968		
3.26	0.999 995 98	3.48	0.999 999 141	3.70	0.999 999 833	3.92	0.999 999 970		
3.27	0.999 996 24	3.49	0.999 999 201	3.71	0.999 999 845	3.93	0.999 999 973		
3.28	0.999 996 49	3.50	0.999 999 257	3.72	0.999 999 857	3.94	0.999 999 975		
3.29	0.999 996 72	3.51	0.999 999 309	3.73	0.999 999 867	3.95	0.999 999 977		
3.30	0.999 996 94	3.52	0.999 999 358	3.74	0.999 999 877	3.96	0.999 999 979		
3.31	0.999 997 15	3.53	0.999 999 403	3.75	0.999 999 886	3.97	0.999 999 980		
3.32	0.999 997 34	3.54	0.999 999 445	3.76	0.999 999 895	3.98	0.999 999 982		
3.33	0.999 997 51	3.55	0.999 999 485	3.77	0.999 999 903	3.99	0.999 999 983		
3.34	0.999 997 68	3.56	0.999 999 521	3.78	0.999 999 910				
3.35	0.999 997 838	3.57	0.999 999 555	3.79	0.999 999 917				

Appendix H

Basic Kinetic Theory of Gases

The ideal gas law states that

$$PV = RT = N_{av} kT \quad (1)$$

where P is the pressure, V is the volume of 1 mole of gas, R is the gas constant (1.98 cal/mol-K, or 82 atm-cm³/mol-K), T is the absolute temperature in K, N_{av} is the Avogadro constant (6.02×10^{23} molecules/mole), and k is the Boltzmann constant (1.38×10^{-23} J/K, or 1.37×10^{-22} atm-cm²/K). Since real gases behave more and more like the ideal gas as the pressure is lowered, Eq. 1 is valid for most vacuum processes. We can use Eq. 1 to calculate the molecular concentration n (the number of molecules per unit volume):

$$n = \frac{N_{av}}{V} = \frac{P}{kT} \quad (2)$$

$$= 7.25 \times 10^{16} \frac{P}{T} \text{ molecules/cm}^3 \quad (2a)$$

where P is in Pa. The density ρ_d of a gas is given by the product of its molecular weight and its concentration:

$$\rho_d = \text{Molecular weight} \times \left(\frac{P}{kT} \right) \quad (3)$$

The gas molecules are in constant motion and their velocities are temperature dependent. The distribution of velocities is described by the Maxwell-Boltzmann distribution law, which states that for a given speed v ,

$$\frac{1}{n} \frac{dn}{dv} = f_v = \frac{4}{\sqrt{\pi}} \left(\frac{m}{2kT} \right)^{3/2} v^2 \exp \left(-\frac{mv^2}{2kT} \right) \quad (4)$$

where m is the mass of a molecule. This equation states that if there are n molecules in the volume, there will be dn molecules having a speed between v and $v + dv$. The average speed is obtainable from Eq. 4:

$$v_{av} = \frac{\int_0^\infty v f_v dv}{\int_0^\infty f_v dv} = \frac{2}{\sqrt{\pi}} \sqrt{\frac{2kT}{m}} \quad (5)$$

An important parameter for vacuum technology is the molecular *impingement rate*, that is, how many molecules impinge on a unit area per unit time. To obtain this parameter, first consider the distribution function f_{vx} for the velocities of molecules in the x direction. This function can be expressed by an equation similar to Eq. 4:

$$\frac{1}{n} \frac{dn_x}{dv_x} \equiv f_{vx} = \left(\frac{m}{2\pi kT} \right)^{1/2} v_x^2 \exp \left(\frac{-mv_x^2}{2kT} \right) \quad (6)$$

The molecular impingement rate ϕ is given by

$$\phi = \int_0^\infty v_x dn_x \quad (7)$$

Substituting dn_x from Eq. 6 and integrating gives

$$\phi = n \sqrt{\frac{kT}{2\pi m}} \quad (8)$$

The relationship between the impingement rate and the gas pressure is obtained by using Eq. 2:

$$\phi = P(2\pi mkT)^{-1/2} \quad (9)$$

$$= 2.64 \times 10^{20} \left(\frac{P}{\sqrt{MT}} \right) \quad (9a)$$

where P is the pressure in Pa and M is the molecular weight.

Appendix I

SUPREM Commands

The Stanford University Process Engineering Modeling (SUPREM) program is a simulation package that allows a user to model various process steps used in the fabrication of integrated circuits. SUPREM can predict the results of oxidation, deposition, etching, diffusion, epitaxial growth, and ion implantation processes. SUPREM III models the changes to the semiconductor structure that result from these processes in one dimension. The primary results are the thicknesses of various layers and the distribution of impurities within those layers. The program can also determine certain material properties, such as the sheet resistance of diffused regions in silicon layers.

To run SUPREM, an input deck must be provided. This file contains a series of statements and comments. The deck begins with a **TITLE** statement, which is merely a comment repeated on each page of the program output. The next command, **INITIALIZE**, is a control statement that sets the substrate type, orientation, and doping. This command can also be used to specify the thickness of the region to be simulated and establish a grid. After the substrate and materials are established, a series of statements is used to specify the sequence of process steps as they occur. Finally, the output of the simulation can be printed or plotted using **PRINT** or **PLOT** statements, respectively. Simulation ends with a **STOP** statement. Several **COMMENT** statements will typically appear throughout the deck.

A description of a few commonly used SUPREM statements is provided in Table I.1. *This table is by no means complete.* To obtain the complete SUPREM software package and its associated documentation, contact

Silvaco Data Systems, Inc.
4701 Patrick Henry Drive
Building 2
Santa Clara, CA 95054
Phone: 408-654-4372
Fax: 408-727-5297
www.silvaco.com

SUPREM is a trademark of the Board of Trustees of Stanford University.

TABLE I.1 Common SUPREM Commands

Name	Description	Basic Syntax	Typical Flags and Parameters
COMMENT	Outputs character string to label an input sequence	COMMENT <text>	None
DEPOSITION	Deposits specified material on top of current structure	DEPOSITION <Material> Thickness=<n> Temperature=<n>	Aluminum Nitride Oxide Polysilicon Silicon C.Phosphor (cm ⁻³) C.Arsenic (cm ⁻³) C.Boron (cm ⁻³) Thickness (μm) Temperature (°C)
DIFFUSION	Models high-temperature diffusion in oxidizing and nonoxidizing ambients	DIFFUSION Time=<n> Temperature=<n> <Dopant> <Ambient>	Arsenic Boron DryO2 Nitrogen Phosphorus WetO2 Solidso1 HClx (%) T.Rate (°C/min) Temperature (°C) Time (minutes)
ETCH	Etches specified material from the top of current structure	ETCH <Material> Thickness=<n>	All Aluminum Nitride Oxide Polysilicon Silicon Thickness (μm)
IMPLANT	Simulates ion implantation of impurities	IMPLANT <Dopant> Dose=<n> Energy=<n>	Arsenic Boron Phosphorus Dose (cm ⁻²) Energy (keV)
INITIALIZE	Sets up initial coefficients and structure to be used in the simulation	INITIALIZE <Structure> <Substrate> <Dopant> Concentration=<n> INITIALIZE Structure=<filename>	<100> <110> <111> Silicon Arsenic Boron Phosphorus Concentration (cm ⁻³)

(continued)

(continued)

Name	Description	Basic Syntax	Typical Flags and Parameters
PLOT	Specifies that impurity concentrations or results of electrical calculations versus depth into the substrate are to be plotted	PLOT <Parameters> Cmin=<n> Cmax=<n>	Active Arsenic Boron Chemical Net Phosphorus Cmin (cm ⁻³) Cmax (cm ⁻³)
PRINT	Outputs information about the structure being simulated and coefficients used	PRINT <Parameters>	Arsenic Boron Chemical Concentration Layers Net Phosphorus
SAVEFILE	Saves the current structure being processed, the coefficients being used, or both	SAVEFILE <Feature> Filename=<Text>	Structure
STOP	Terminates simulation	STOP <Text>	None
TITLE	Inputs a character string to label the following input	TITLE <Text>	None

Appendix J

Running PROLITH

PROLITH is a Windows-based photolithography simulation program marketed by FINLE Technologies in Austin, Texas. PROLITH simulates the complete one- and two-dimensional optical lithography process from aerial image formation through resist exposure and development. The output of the program is an accurate prediction of the final resist profile, which is presented in a wide variety of images, plots, graphs, and calculations.

PROLITH accepts lithography information in the form of data files and input parameters and uses this information to simulate standard and advanced lithography processes. After the software has been installed, a user can run PROLITH by simply clicking on the PROLITH icon from the Windows Start menu. After a successful license search, the Imaging Tool parameters window appears (see Fig. 4.20). Before running a simulation, the user must choose the simulation options and enter a set of input parameters. This is accomplished by selecting Options from the File menu to open the Options dialog. The settings in this dialog box are used to establish basic simulation options, such as the Image Calculation Mode, Physical Model, and Speed Factor.

Once the options have been established, the user may enter simulation input parameters from the appropriate parameters window. These windows are opened by selecting various options from the View menu or by clicking the corresponding toolbar button for the window. In each window, parameters may be entered by marking or clearing checkboxes or option buttons, entering values in text boxes, selecting files or other values from lists, and so forth. Many parameter windows, such as the Resist parameters window, for example, provide instant graphical views of the information entered.

After input parameters have been entered, PROLITH displays simulation results from the Graphs menu. PROLITH can produce graphs simulating formation of a mask feature by an optical projection system, exposure of photoresist using this image, or development of the exposed photoresist. The following options from the Graphs menu (or corresponding toolbar buttons) are available to display such simulations:

- *Aerial Image*: The relative intensity of the image as a function of position
- *Image in Resist*: The image projected into the photoresist at the start of exposure
- *Exposed Latent Image*: The latent image before post-exposure bake (PEB)
- *PEB Latent Image*: The latent image after post-exposure bake
- *Develop Time Contours*: Contours of constant develop time as a function of position in the resist
- *Resist Profile*: Two-dimensional photoresist profile after development

This overview of PROLITH capabilities is by no means complete. To obtain the complete PROLITH software package and its associated documentation, contact

FINLE Technologies, Inc.

P.O. Box 162712

Austin, TX 78716

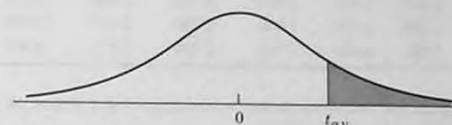
Phone: 512-327-3781

Fax: 512-327-1510

www.finle.com

Appendix K

Percentage Points of the t Distribution



α v	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792

(continued)

(continued)

α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

 v , degrees of freedom.Source: Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

Appendix L

Percentage Points of the F Distribution

		$F_{0.25, v_1, v_2}$																			
v_2	v_1	Degrees of Freedom for the Numerator (v_1)																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
Degrees of Freedom for the Denominator (v_2)	1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85	
	2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48	
	3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47	
	4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	
	5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87	
	6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74	
	7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.68	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65	
	8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58	
	9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53	
	10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48	
	11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45	
	12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42	
	13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40	
	14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38	
	15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	
	16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	
	17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	
	18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32	
	19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30	
	20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29	
	21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	
	22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.27	
	23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.26	
	24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26	
	25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25	
	26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25	
	27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24	
	28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24	
	29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23	
	30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23	
	40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19	
	60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15	
	120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10	
	∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00	

$$F_{0.75, v_1, v_2} = 1/F_{0.25, v_2, v_1}$$

(continued)

Source: Adapted with permission from Biometrika Tables for Statisticians, Vol. 1, 3rd ed., by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

		$F_{0.10, v_1, v_2}$																		
v_1		Degrees of Freedom for the Numerator (v_1)																		
v_2		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
Degrees of Freedom for the Denominator (v_2)	1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
	2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
	4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
	6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
	7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
	8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
	9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
	11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
	13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
	14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
	15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
	16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.78	1.75	1.72
	17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
	18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
	19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
	21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
	22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
	23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
	24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.19	
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00	

		$F_{0.05, v_1, v_2}$																			
		Degrees of Freedom for the Numerator (v_1)																			
v_2	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
Degrees of Freedom for the Denominator (v_2)	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81		
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78		
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76		
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73		
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71		
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69		
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67		
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65		
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64		
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62		
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51		
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39		
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25		
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00		

$$F_{0.95, v_1, v_2} = 1/F_{0.05, v_2, v_1}$$

(continued)

		$F_{0.25, v_1, v_2}$																			
v_1		Degrees of Freedom for the Numerator (v_1)																			
v_2		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
Degrees of Freedom for the Denominator (v_2)	1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001.0	1006.0	1010.0	1014.0	1018.0	
	2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50	
	3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	
	4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	
	5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	
	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	
	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	
	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	
	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	
	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	
	11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	
	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	
	13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	
	14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	
	15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40	
	16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	
	17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	
	18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	
	19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	
	20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04		
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00		
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97		
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94		
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91		
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88		
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85		
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83		
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81		
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79		
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64		
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48		
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31		
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00		

		$F_{0.25, v_1, v_2}$																		
		Degrees of Freedom for the Numerator (v_1)																		
v_2	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
Degrees of Freedom for the Denominator (v_2)	1	4052.0	4999.5	5403.0	5625.0	5764.0	5859.0	5928.0	5982.0	6022.0	6056.0	6106.0	6157.0	6209.0	6235.0	6261.0	6287.0	6313.0	6339.0	6366.0
	2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
	13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
	15	8.68	6.36	5.42	4.89	4.54	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
	16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
	17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
	18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
	19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.50
	20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
	21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
	22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
	23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
	24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
	25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
	26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
	27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
	28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
	29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
	30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
	40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
	60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
	120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
	∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

 $F_{0.90, v_1, v_2} = 1/F_{0.01, v_2, v_1}$

Index

- Advanced isolation technology, 207–208
- Aluminum, 144
- etching, 88
- metallization, 169–173
- Ammonium fluoride (NH_4F), 87
- Amorphous silica, 43
- Analysis of variance (ANOVA)
- table, 245–246
- technique, 243–244
- Annealing
- boron and phosphorus, 134
- rapid thermal, 135–136
- ANOVA, *see* Analysis of variance (ANOVA)
- Anisotropic etch profiles, 92
- APCVD, Atmospheric pressure CVD (APCVD)
- Area defect, 34
- Aspect ratio–dependent etching, 98
- ATE, *see* Automated test equipment (ATE)
- Atmospheric pressure CVD (APCVD), 145
- Attributes, 237
- Automated test equipment (ATE), 228
- Automobile industry, 1
- Ball grid array (BGA), 231
- Barium strontium titanate (BST), 164
- BGA, *see* Ball grid array (BGA)
- BiCMOS technology, 210–211
- Bipolar technology, 188–195
- Bipolar transistor, 5, 189
- Blanket etches, 85
- Boron, conventional annealing of, 134–135
- Bow, 32
- Bridgman technique, 6, 30–31
- BST, *see* Barium strontium titanate (BST)
- Buffered HF solution (BHF), 88
- Buffered oxide etch (BOE), *see* Buffered HF solution (BHF)
- Bulk micromachining, 215
- CAD, *see* Computer-aided design (CAD)
- Capacitance, 162, 186, 201
- Capacitance–voltage technique, 114
- CAR, *see* Chemical-amplified resist (CAR)
- Carbon concentration, 36
- (CBE), *see* Chemical beam epitaxy (CBE)
- CCD, *see* Charge-coupled device (CCD)
- C-chart, *see* Defect chart (C-chart)
- CD, *see* Critical dimension (CD)
- Cell projection, 75
- Central processing unit (CPU), 8, 203
- CF, *see* Correction factor (CF)
- Chanstop, *see* p+ channel stop
- Charge-coupled device (CCD), 5
- Chemical-amplified resist (CAR), 70
- Chemical beam epitaxy (CBE), 152
- Chemical etchants, 85
- Chemical etching, 92
- Chemical mechanical polishing (CMP), 8, 100, 173–175
- Chemical vapor deposition (CVD), 14, 50, 88, 145, 168
- for gallium arsenide, 147
- metallorganic, 147–148
- for silicon, 145–147
- silicon oxides and dioxide, 156
- TiN, 169
- tungsten, 169
- Chip scale package (CSP), 232
- Clean room, importance for lithography, 60–62
- Close proximity printing, 62
- Clustered plasma processing, 97
- CMOS, *see* Complementary MOSFET (CMOS)
- CMP, *see* Chemical mechanical polishing (CMP)
- Cobalt silicide, 176
- Comb-meander-comb structure, 228
- Complementary MOSFET (CMOS), 8, 196
- technology, 203–205
- well of, 206
- Computer-aided design (CAD), 65, 256
- Computer-integrated manufacturing of integrated circuits (IC-CIM), 256–257
- Concentration-dependent diffusivity, 115–117, 119
- Conformal step coverage, 159–160

- Constant surface concentration, 109–111, 118–119
- Constant total dopant, 111–112
- Contact printing, 62–63
- Continuous-wave (CW) power, 5
- Control charts, 237
- for attributes, 237–239
- for variables, 239–241
- COP. *see* Poly-glycidyl methacrylate-co-ethyl-acrylate (COP)
- Copper (Cu), 109, 261
- interconnect, 8
- metallization, 173
- Correction factor (CF), 113
- CPU. *see* Central processing unit (CPU)
- Critical dimension (CD), 63
- Critical layer thickness, 152
- Crystal defects, 33–35
- Crystal growth
- epitaxial growth and, 144
- techniques, 30–31
- Crystal puller, 18
- CSP. *see* Chip scale package (CSP)
- CVD. *see* Chemical vapor deposition (CVD)
- Czochralski technique, 18, 30–31
- Damascene technology, 99–100, 173
- Deep-trench isolation, 207
- Defect chart (C-chart), 238
- Defect density chart (U-chart), 238
- Defect-free zone. *see* Denuded zone
- Degrees of freedom, 243
- Denuded zone, 37–38
- Deposition
- rate, 166
- simulation, 177
- Depth of focus (DOF), 65
- Designed experiment, 242
- Design matrix, 246
- DIBL. *see* Drain-induced barrier lowering (DIBL)
- Dichlorosilane ($\text{Si H}_2\text{Cl}_2$), 145
- Dielectric deposition, 155–165
- Dielectric etching, 99
- Dielectric isolation, 192–193
- Dielectric layers, 41, 144, 162–163
- Die separation, 230
- Diffraction, 69
- Diffused layers, evaluation of, 113–114
- Diffused well, 206
- Diffusion
- coefficient, 107–108, 116
- constant, 52
- equation, 107–109
- extrinsic, 114–117
- introductory note, 105–106
- ion implantation and, 14
- lateral, 118–120
- process, 106–114
- profiles, 109–112, 114, 117
- in silicon, 117
- simulation, 120–121
- theory, 6
- Diffusivity. *see* Diffusion, coefficient
- DIP. *see* Dual in-line package (DIP)
- Dislocation. *see* Line defect
- Dislocation loops, 154
- DOF. *see* Depth of focus (DOF)
- Donor vacancy, 115
- Dopant
- distribution of, 19–21
- impurities redistribution, 50–51
- profile of the diffused layer, 113–114
- segregation coefficient, 30–31
- Double charged acceptor vacancy, 115
- Double-comb structure, 228
- Double-polysilicon structure, self-aligned, 193–195
- Drain, 197, 200, 206
- Drain-induced barrier lowering (DIBL), 208
- DRAM. *see* Dynamic random access memory (DRAM)
- Drive-in diffusion, 112
- Dry etching, 71, 89–90
- Dry oxidation, 42
- Dual in-line package (DIP), 230–231
- Dynamic random access memory (DRAM), 8, 11, 164, 199–200, 263–264
- E-beam lithography. *see* Electron beam lithography
- ECR. *see* Electron cyclotron resonance (ECR)
- Edge dislocation, 33, 154
- EGS. *see* Polycrystalline silicon (EGS)
- Electrical testing of integrated circuits, 227–228
- Electroluminescence phenomenon, 2
- Electromigration, 172–173
- Electron beam lithography, 73–76
- Electron cyclotron resonance (ECR)
- plasma etching, 95
- Electronics industry, 1
- Electron resist, 75
- Ellipsometry, 54
- Emitter, 192, 194
- End-point control, 93
- Epitaxial growth techniques, 144–152
- Epitaxial layers, 144, 145
- n-type, 190–191
- structures and defects in, 152–154
- Epitaxial process, 7, 144–145
- Epitaxy, 7, 144
- Equilibrium segregation coefficient, 19–21, 30
- Erfc distribution, 118
- Etch
- mechanism, 91–92
- rates, 85
- simulation, 101–102
- Etchants
- chemical, 85
- commonly used, 87
- Etches, blanket, 85
- Etching
- aluminum, 88
- aspect ratio-dependent, 98
- chemical, 92
- dielectric, 99
- dry, 71, 89–90
- equipment, 93–94
- gallium arsenide, 88–89
- interconnect metal, 99–101
- orientation-dependent, 86, 87
- photolithography and, 13–14
- physical, 92
- plasma, 91–92
- polycide gate, 98–99
- polysilicon, 88, 98–99
- silicon, 86–87
- silicon dioxide, 87–89
- silicon nitride, 88
- silicon trench, 97–98
- single-crystal silicon, 88
- wet, 88
- wet chemical, 85–89
- Etching, polysilicon. *see* Polysilicon, etching
- Etching single-crystal silicon. *see* Single-crystal silicon etching
- Etch rate uniformity, 85
- Etch-resistant polymers, 6
- Eutectic temperature, 170
- EUV lithography. *see* Extreme ultraviolet (EUV) lithography
- Experimental design, statistical, 242
- Exposure tools, 62–65
- Extreme ultraviolet (EUV) lithography, 76–78
- Extrinsic p^+ base regions, 194
- Extrinsic diffusion, 114–117
- Extrinsic stacking fault, 34–35
- Fabrication steps, integrated circuit, 196–198
- Factorial experimental designs, 246
- Fermi level, 116
- Fick's diffusion equation, 107
- Field oxide, 41, 197
- Film deposition
- dielectric, 155–165
- epitaxial growth techniques, 144–152
- metallization, 167–177
- polysilicon, 165–167
- simulation, 177
- structures, 152–154
- Films, 144
- Flats, 31
- Flip-chip bonding, 235
- Floating-gate nonvolatile memory, 201–202
- Float-zone process, 24–26
- Four-point probe technique, 35–36, 113
- Fractional factorial designs, 249
- Frenkel defect, 34
- Functional testing, 228
- Functional yield, 250–251
- Gallium arsenide (GaAs), 2, 109
- chemical vapor deposition (CVD) for, 147
- crystal growth techniques, 26–31
- etching, 88–89
- vs. silicon, 32, 212
- zinc diffusion in, 117
- Gate
- oxide, 60
- oxide layer, 41
- Gate engineering technology, 208–210
- Gaussian distribution, 114
- Germanium (Ge), 2
- Gettering, 37

G-line, 65
 Grain boundary, 34
 Gunn diode. *see* Transferred-electron diode (TED)
 HDP. *see* High-density plasma (HDP)
 Heteroepitaxy, 144, 152
 Heterojunction bipolar transistor, 2
 Heterostructure laser, 4
 Hexamethyl-disilazane (HMDS), 70
 HF. *see* Hydrofluoric acid (HF)
 High-density plasma (HDP) etchers, 90, 95–97, 99
 High-dielectric-constant materials, 164–165
 High-energy and high-current implantation, 139–140
 H-line, 65
 HMDS. *see* Hexamethyl-disilazane (HMDS)
 Homoepitaxy, 144, 152
 Hydrofluoric acid (HF), 86
 Hypothesis test, 243
 IC-CIM. *see* Computer-integrated manufacturing of integrated circuits (IC-CIM)
 IC manufacturing technology, 80–81
 ICP source. *see* Inductively coupled plasma (ICP) source
 ILD. *see* Interlayer dielectric (ILD)
 I-line, 65
 Image system, simple, 64–65
 IMPATT diode, 5, 182
 Implant damage, 131–134
 Impurity distribution, 113
 Impurity doping, 105
 Impurity redistribution during oxidation, 50–51
 Inductively coupled plasma (ICP) source, 95
 In situ doping, 167
 Integrated circuit (IC) manufacturing computer-integrated, 256–257
 electrical testing, 227–228
 packaging, 228–236
 statistical experimentation design, 242–249
 statistical process control, 237–241
 yield, 250–255
 Integrated circuits (ICs), 1, 7, 182

capacitor, 185–187
 inductor, 187–188
 monolithic, 7
 pattern transfer, 70–71
 resistor, 184–185
 Interconnections, 14
 Interconnect metal etching, 99–101
 Interface defects, 154
 Interlayer dielectric (ILD), 173
 International Technology Roadmap for Semiconductors, CA, 1999, 80, 259
 Interstitial diffusion, 107, 108
 Intrinsic base region, 194
 Intrinsic carrier concentration, 114–115
 Intrinsic diffusivity, 115
 Intrinsic stacking fault, 34–35
 Ion
 channeling, 130–131
 distribution, 125–126
 stopping, 127–130
 Ion beam
 lithography, 79–80
 sputtering, 167
 Ion implantation, 70
 damage and annealing, 131–136
 and diffusion, 14
 range of, 125–131
 related processes, 136–140
 simulation, 140–141
 Isolation technology, advanced, 207–208
 Junction depth, 113, 117
 Junction spiking, 169–172
 Kinetics of growth, 42–49
 Large-scale integration (LSI), 184
 Laser diodes, 4
 Laser interferometry, 93
 Latchup, 205
 Lateral diffusion, 118–120
 Lateral oxide isolation, 191
 Lateral straggle (σ_L), 125
 Lattice disorder, 131
 Lattice-matched epitaxy, 152–153
 LCL. *see* Lower control limit (LCL)
 Lead zirconium titanate (PZT), 164
 Level 1 packaging, 232
 Liftoff technique, 70–71
 LIGA. *see* Lithographic, galvanofarming, abfarming (LIGA)

Lightly doped drain (LDD) MOSFET, 138
 Linear rate constant, 46–47
 Line defect, 34
 Lithographic, galvanofarming, abfarming (LIGA), 215–218
 Lithographic methods
 comparison of, 80–81
 next-generation, 73–80
 Lithography process, 5, 6
 Liftoff technique, 70–71
 Local oxidation of silicon (LOCOS), 191, 197
 LOCOS, Local oxidation of silicon (LOCOS)
 Long-throw sputtering, 168
 Low-angle grain boundaries, 154
 Low-dielectric-constant (low- k) insulators, 261
 Low-dielectric-constant materials, 162–164
 Lower control limit (LCL), 237
 Low pressure chemical vapor deposition (LPCVD), 101, 145, 155, 162
 LPCVD. *see* Low pressure chemical vapor deposition (LPCVD)
 LSI. *see* Large-scale integration (LSI)
 Magnetically enhanced reactive ion etching (MERIE), 90
 Magnetron sputtering, 168
 Main effect, 247
 Mask-beam system, 80
 Mask damage, 63
 Masking
 multiple implantation and, 136–138
 oxide, 7
 properties of silicon dioxide, 51–52
 Masks, 65–67, 79
 MBE. *see* Molecular beam epitaxy (MBE)
 Meander structure, 227–228
 Mean free path, 150
 Mean square, 245
 Mean time to failure (MTF), 172
 Medium-energy ion implantor, 125
 Medium scale integration (MSI), 184
 Memory devices, 199–203
 MEMS. *see* Microelectromechanical systems (MEMS)
 MERIE. *see* Magnetically enhanced reactive ion etching (MERIE)
 MESFET. *see* Metal-semiconductor field-effect transistor (MESFET)

Metal, films, 14, 41, 144
 Metallization, 14, 167–177, 198
 Metal-nitride-oxide-semiconductor (MOS), 201
 Metalorganic chemical vapor deposition (MOCVD), 8, 147–148
 Metalorganic molecular beam epitaxy (MOMBE), 152
 Metal-oxide-semiconductor field-effect transistor (MOSFET), 4, 182
 technology of, 196
 Metal-oxide-semiconductor (MOS) capacitors, 185–186
 devices, 53, 98
 Metal-semiconductor field-effect transistor (MESFET), 5, 37, 182
 Metal silicides, 175–176
 Microelectromechanical systems (MEMS), 212–214
 technology, 212–214
 Micromachining, surface and bulk, 215
 Minimum device dimension, 259
 Minimum feature length. *see* Minimum device dimension
 MMICs. *see* Monolithic microwave integrated circuits (MMICs)
 MNOS. *see* Metal-nitride-oxide-semiconductor (MNOS)
 Mobile ionic charges, 54
 MOCVD. *see* Metalorganic chemical vapor deposition (MOCVD)
 MOCVD reactor, 148
 MODFET. *see* Modulation-doped field-effect transistor (MODFET)
 Modulation-doped field-effect transistor (MODFET), 5
 Molecular beam epitaxy (MBE), 8, 145, 148–152
 MOMBE. *see* Metalorganic molecular beam epitaxy (MOMBE)
 Monolithic IC, 7
 Monolithic microwave integrated circuits (MMICs), 5, 187, 212
 Monte Carlo simulation, 254
 MOSFET. *see* Metal-oxide-semiconductor field-effect transistor (MOSFET)
 MSI. *see* Medium scale integration (MSI)
 MTF. *see* Mean time to failure (MTF)
 Multiple implantation and masking, 136–138
 Murphy's yield integral, 252–254

- n*-channel MOSFET (NMOS), 8, 41, 196, 197, 203–205
- Negative photoresist, 68–69, 75
- Neutral vacancy, 115
- Neutron irradiation, 25
- Next-generation lithographic methods, 73–81
- Nitric acid (HNO_3), 87
- Nonconformal step coverage, 159–160
- Nonvolatile memory device, 201–202
- Nonvolatile semiconductor memory (NVSM), 5, 203
- n-p-n* bipolar transistor, 189
- n+* polysilicon, 194
- n*-type dopants, 106
- NVSM. *see* Nonvolatile semiconductor memory (NVSM)
- OES. *see* Optical emission spectroscopy (OES)
- Ohmic contact, 14, 144
- OPC. *see* Optical proximity correction (OPC)
- Optical emission spectroscopy (OES), 93
- Optical lithographic pattern transfer process, 70–71
- Optical lithography, 60–72, 79
- Optical proximity correction (OPC), 72
- Orientation-dependent etching, 86, 87
- Oscillation, 93
- Oxidation
- impurity redistribution during, 50–51
 - in planar process, 11–12
 - simulation, 54, 56–57
 - vertical, 50
- Oxide
- growth (thin), 49–50
 - layer, 7
 - masking, 7
 - quality, 53–54
 - thickness characterization, 54
 - trapped charges, 54
- Packaging, integrated circuit
- levels of, 228–230
 - types of, 230–232
- Parabolic rate constant, 46, 47
- Parametric yield, 250, 254–255
- Pattern transfer, integrated circuit, 70–71
- PBS. *see* Poly-butene-1 sulfone (PBS)
- p*-channel MOSFET (PMOS), 196, 197, 203–205, 208
- p+* channel stop, 191–192
- PCMs. *see* Process control monitors (PCMs)
- PECVD. *see* Plasma-enhanced chemical vapor deposition (PECVD)
- PGA. *see* Pin grid array (PGA)
- P-glass flow process, 160, 196
- Phase, 26–27
- Phase diagram, 27
- Phase-shifting mask (PSM), 72
- Phosphorus
- conventional annealing of, 134–135
 - diffusion profiles, 117, 118
- Photolithography
- defined, 60
 - etching and, 13–14
 - simulation, 81–82
- Photomask, 60, 69, 73
- Photoresist, 13, 60, 67–70, 192. *see also*
- Etch-resistant polymers
 - negative, 68–69, 75
 - positive, 67–68, 75
- Physical etching, 92
- Physical vapor deposition (PVD), 14, 167–168
- Pin grid array (PGA), 231
- Planar process, 7, 11–14
- Plasma
- diagnostics, 93
 - fundamentals of, 90–91
- Plasma-assisted etching. *see* Dry etching
- Plasma-enhanced chemical vapor deposition (PECVD), 42, 155, 156
- Plasma etching
- as process, 91–92
 - reactors, 93–94
- PMMA. *see* Poly-methyl methacrylate (PMMA)
- PMOS. *see* *p*-channel MOSFET (PMOS)
- p-n* junction, 7, 14
- Point defect, 34
- Poisson distribution, 238
- Poisson yield model, 251–252
- Poly 1. *see* *p+* polysilicon
- Poly 2. *see* *n+* polysilicon
- Poly-butene-1 sulfone (PBS), 75
- Polyicide gate etching, 98–99
- Polyicide process, 175

- Polycrystalline silicon (EGS), 18, 41, 144.
- see also* Polysilicon
- Poly-glycidyl methacrylate-co-ethylacrylate (COP), 75
- Polymers, 68, 75
- Poly-methyl methacrylate (PMMA), 75, 79
- Polysilicon, 144
- capacitor, 201
 - deposition, 165–167
 - electrode, 201
 - etching, 98–99
 - silicon nitride and, 88
- Positive photoresist, 67–68, 75
- Power limitations, 261–262
- Power supply voltage, 259–260
- p+* polysilicon, 194
- gate, 208
- Precipitates, 154
- Predeposition diffused layer, 112
- Pressure-by-temperature-by-flow rate, 248
- Pressure-by-temperature interaction, 248
- Primary flat, 31
- Process control monitors (PCMs), 227
- Process simulation, 218–222
- Profilometry, 54
- Projected range (R_p), 125
- Projected straggle (σ_p), 125
- PROLITH, 81
- Proximity effect, 75–76
- PSM. *see* Phase-shifting mask (PSM)
- p* tub, 204–205
- p*-type impurity, 106
- p*-type silicon substrate, 211
- p* well. *see* *p* tub
- PZT. *see* Lead zirconium titanate (PZT)

Quartzite, 18

- Radio frequency (RF) heater, 24
- RAM. *see* Random access memory (RAM)
- Random access memory (RAM), 199
- Range (R), 125
- Rapid thermal annealing (RTA), 134, 135–136
- Raster scan system, 74–75
- Reactive ion beam etching, 90
- Reactive ion etching (RIE), 90, 94–95, 173
- Reactive plasma etching
- applications, 97–101
 - techniques, 93–94

- Reactive reactor technology, 93
- Reactor, 93
- Rectifying metal-semiconductor barriers, 144
- Redistribution diffusion. *see* Drive-in diffusion
- Refilled trench, 205
- Registration, 62
- Resistance-heated oxidation furnace, 41
- Resistivity, 113, 114
- Resistors, 185
- Resolution
- defined, 62
 - enhancement techniques, 72
- Resonant tunneling diode (RTD), 5
- Retrograde well. *see* Twin well
- RIE. *see* Reactive ion etching (RIE)
- R_p . *see* Projected range (R_p)
- RTA. *see* Rapid thermal annealing (RTA)
- RTD. *see* Resonant tunneling diode (RTD)
- Salicide, 176
- Scanning focused-beam system, 80
- Schottky barrier diode, 114
- Screw dislocation, 33
- Secondary flat, 31
- Secondary mass spectroscopic (SIMS) technique, 36, 114
- Segregation coefficient
- effective, 22–23
 - equilibrium, 30, 50
- Self-aligned structure, 193–195
- SEMC. *see* Single-electron memory cell (SEMC)
- SEMI. *see* Semiconductor Equipment and Materials Institute (SEMI)
- Semiconductor
- devices, 2–5
 - industries, 1
 - materials, 2
 - process technologies, 5–8
- Semiconductor Equipment and Materials Institute (SEMI), 32
- Separation by implanted oxygen (SIMOX), 140, 193
- Shadow printing, 62–63, 78
- Shallow trench isolation, 207
- Sheet resistance, 113–114
- Shewhart control charts. *see* Control charts
- Shmoo plot. *see* Two-dimensional plot

- Silica. *see* Silicon dioxide (SiO_2)
- Silicide, 144
formation, 261
metal, 175–177
- Silicon crystal growth, 18–23
- Silicon dioxide (SiO_2), 2, 11–12, 43–44, 144, 156
deposition methods, 157
diffusion constant for dopants in, 52
etching, 87–88
masking properties of, 51–52
properties of, 157–159
- Silicone float-zone process. *see* Float-zone process
- Silicon-on-insulator (SOI), 140
integration, 262
- Silicon (Si), 2, 109
characteristics, 35–36
chemical vapor deposition (CVD) for, 145–147
compared with Gallium arsenide, 212
crystal, 18
diffusion in, 117
etching, 86–87
nitride, 85, 144, 160–162
oxidation, 41–59
tetrachloride (SiCl_4), 145, 146
trench etching, 97–98
wafer, 33
- SIMOX. *see* Separation by implanted oxygen (SIMOX)
- SIMS technique. *see* Secondary mass spectroscopic (SIMS) technique
- Simulation
deposition, 177
diffusion, 120–121
etch, 101–102
ion implantation, 140–141
oxidation, 54, 56–57
photolithography, 81–82
process, 218–222
- Single-crystal silicon etching, 88
- Single-electron memory cell (SEMC), 5
- SLS. *see* Strained-layer superlattice (SLS)
- Small-scale integration (SSI), 184
- SOC. *see* System-on-a-chip (SOC)
- SOI. *see* Silicon-on-insulator (SOI)
- Source forming, 197
- SPC. *see* Statistical process control (SPC)
- Sputter-deposition rate, 167
- Sputtering, 168
- SRAM. *see* Static random access memory (SRAM)
- SSI. *see* Small-scale integration (SSI)
- Stacking fault, 34
- Stanford University Process Engineering Modeling (SUPREM) program, 56, 81, 101, 120, 177, 211, 218–222
- Static random access memory (SRAM), 199, 201
- Statistical experimental design, 242
- Statistical process control (SPC), 237, 242
- Step-and-repeat projection, 64, 65
- Step coverage, 159–160
- Strained-layer epitaxy, 152
- Strained-layer superlattice (SLS), 153
- Substrates defects, 153
- Sums of squares, 244–245
- Superlattice, 152
- SUPREM. *see* Stanford University Process Engineering Modeling (SUPREM) program
- Surface concentration, 113, 114, 116
- Surface micromachining, 215
- Surface-mount packages, 231
- Surface orientation, 33
- Surface wave-coupled plasma (SWP) source, 95
- SWP source. *see* Surface wave-coupled plasma (SWP) source
- System-on-a-chip (SOC), 262–264
- TAB. *see* Tape-automated bonding (TAB)
- Tape-automated bonding (TAB), 235
- Taper, 32
- Target, 127
- TCP source. *see* Transformer-coupled plasma (TCP) source
- TED. *see* Transferred-electron diode (TED)
- Test statistic, 243
- Thermal oxidation process, 42–50
- Thermal oxide, 144, 194
- Thermocompression technique, 233
- Thermosonic bonding, 233
- Thickness, 32
- Thin oxide growth, 49–50
- Throughput, 62
- Thyristor, 2
- Tilt-angle ion implantation, 138–139

- Titanium nitride (TiN), chemical vapor deposition (CVD), 169
- Transferred-electron diode (TED), 5
- Transformer-coupled plasma (TCP) source, 95, 96
- Transistors, bipolar, 189
- Trench isolation, 8, 207
- Trench type capacitor, 201
- Trichlorosilane (SiHCl_3), 18, 145
- Tungsten, chemical vapor deposition (CVD), 169
- Twining, 34
- Twins, 154
- Twins defect, 34
- Twin tub, 205
- Twin well process, 206
- Two-dimensional plot, 228
- Two-level factorial design, 246–248
- U-chart. *see* Defect density chart (U-chart)
- UCL. *see* Upper control limit (UCL)
- ULSI. *see* Ultralarge-scale integration (ULSI)
- Ultralarge-scale integration (ULSI), 35–37, 95, 184
- Ultrasonic bonding, 233
- Ultrashallow junction formation, 261
- Ultrathin oxide, 261
- Upper control limit (UCL), 237
- Vacancy diffusion, 107, 108
- Vacancy model of Fair, 120
- Vapor-phase epitaxy (VPE), 145
- Vector scan system, 74–75
- Vertical oxidation, 50
- Volume defect, 35
- VPE. *see* Vapor-phase epitaxy (VPE)
- Wafer, 13, 183
shaping, 31–33
slicing and parameters, 32
- Well formation technology, 206
- Wet chemical etching, 85–86
- Wet etching, 88
- Wet oxidation, 42
- Wire bonding, 232–233
- X-ray lithography (XRL), 78–79
- XRL. *see* X-ray lithography (XRL)
- Yates algorithm, 248–249
- Yield, 67, 226, 250
- Zinc diffusion, in Gallium arsenide, 117
- Zone-refining technique, 25