

33

Processing of Integrated Circuits

Chapter Contents

33.1 Overview of IC Processing

- 33.1.1 Processing Sequence
- 33.1.2 Clean Rooms

33.2 Silicon Processing

- 33.2.1 Production of Electronic Grade Silicon
- 33.2.2 Crystal Growing
- 33.2.3 Shaping of Silicon into Wafers

33.3 Lithography

- 33.3.1 Optical Lithography
- 33.3.2 Other Lithography Techniques

33.4 Layer Processes Used in IC Fabrication

- 33.4.1 Thermal Oxidation
- 33.4.2 Chemical Vapor Deposition
- 33.4.3 Introduction of Impurities into Silicon
- 33.4.4 Metallization
- 33.4.5 Etching

33.5 Integrating the Fabrication Steps

33.6 IC Packaging

- 33.6.1 IC Package Design
- 33.6.2 Processing Steps in IC Packaging

33.7 Yields in IC Processing

An **integrated circuit** (IC) is a collection of electronic devices such as transistors, diodes, and resistors that have been fabricated and electrically interconnected onto a small flat chip of semiconductor material. Silicon (Si) is the most widely used semiconductor material for ICs, because of its combination of properties and low cost. Less common semiconductor chips are made of gallium arsenide (GaAs) and germanium (Ge). Because the circuits are fabricated into one solid piece of material, the term **solid-state** electronics is used to denote these devices. Semiconductor devices such as integrated circuits are the basis for virtually all modern electronics products, which constitute the world's largest industry, overtaking the automobile industry in gross sales in 1998 [15].

The IC was invented in 1959 and has been the subject of continual development ever since (Historical Note 33.1). The most fascinating aspect of microelectronics technology is the huge number of devices that can be packed onto a single small chip. Various terms have been developed to define the level of integration and density of packing, such as large-scale integration (LSI) and very-large-scale integration (VLSI). Table 33.1 lists these terms, their definitions (although there is not complete agreement over the dividing lines between levels), and the period during which the technology was introduced. In 1975, Gordon Moore[†] formulated what has come to be known as **Moore's law**, which states that the number of transistors (the building blocks of logic and memory devices) on an integrated circuit doubles every 2 years [23]. The predictive ability of this law has remained accurate to the present time. Today's gigascale technology is capable of fabricating millions of transistors per square millimeter of processable area on the chip surface.

[†]Gordon Moore was a cofounder of Intel Corporation, and at time of writing, is the company's Chairman Emeritus.

TABLE • 33.1 Levels of integration in microelectronics.

Integration Level	Number of Devices on a Chip	Approx. Year Introduced
Small-scale integration (SSI)	10–50	1959
Medium-scale integration (MSI)	50–10 ³	1960s
Large-scale integration (LSI)	10 ³ –10 ⁴	1970s
Very-large-scale integration (VLSI)	10 ⁴ –10 ⁶	1980s
Ultra-large-scale integration (ULSI)	10 ⁶ –10 ⁸	1990s
Giga-scale integration (GSI)	10 ⁸ –10 ¹⁰	2000s

Recent advances in semiconductor technology include system-on-chip and three-dimensional integrated circuits [21]. **System-on-chip** refers to the fabrication of an integrated circuit that contains all of the components required in a computer. Conventional computers include multiple integrated circuits and other components that are interconnected on a printed circuit board (Chapter 34). The system-on-chip concept minimizes assembly costs and power requirements for the computer. A **three-dimensional integrated circuit** is an IC consisting of components that have both vertical and horizontal features, enabling faster operation because the average conduction distance between components is reduced compared with a two-dimensional layer with the same number of components. Intel Corporation’s Tri-Gate technology uses 3-D transistors with vertical fins that project upward from the silicon chip surface, enabling switching speeds to be increased and power requirements to be reduced.

Historical Note 33.1

Integrated circuit technology

The history of integrated circuits includes inventions of electronic devices and the processes for making these devices. The development of radar immediately before World War II (1939 to 1945) identified germanium and silicon as important semiconductor elements for the diodes used in radar circuitry. Owing to the importance of radar in the war, commercial sources of germanium and silicon were developed.

In 1947, the transistor was developed at the Bell Telephone Laboratories by J. Bardeen and W. Brattain. An improved version was subsequently invented by W. Shockley of Bell Labs in 1952. These three inventors shared the 1956 Nobel Prize in Physics for their research on semiconductors and the discovery of the transistor. The interest of the Bell Labs was to develop electronic switching systems that were more reliable than the electromechanical relays and vacuum tubes used at that time.

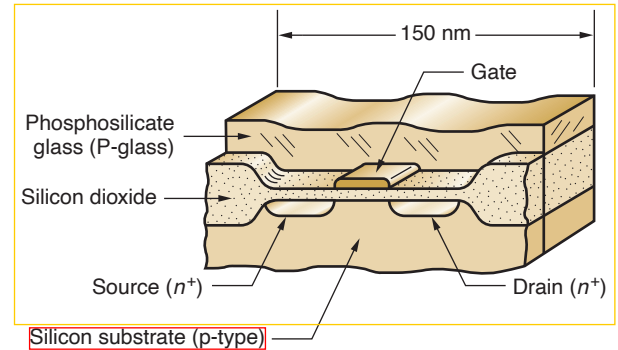
In February 1959, J. Kilby of Texas Instruments Inc. filed a patent application for the fabrication of multiple electronic devices and their interconnection to form a circuit on a single piece of semiconductor material. Kilby was describing an integrated circuit (IC). In May 1959, J. Hoerni of Fairchild Semiconductor Corp. applied for a patent describing the planar process for fabricating transistors. In July of the same year, R. Noyce also of

Fairchild filed a patent application similar to the Kilby invention but specifying the use of planar technology and adherent leads.

Although filed later than Kilby’s, Noyce’s patent was issued first, in 1961 (the Kilby patent was awarded in 1964). This discrepancy in dates and similarity in inventions have resulted in considerable controversy over who was really the inventor of the IC. The issue was argued in legal suits stretching all the way to the U.S. Supreme Court. The high court refused to hear the case, leaving stand a lower court ruling that favored several of Noyce’s claims. The result (at the risk of oversimplifying) is that Kilby is generally credited with the concept of the monolithic integrated circuit, whereas Noyce is credited with the method for fabricating it.

The first commercial ICs were introduced by Texas Instruments in March 1960. Early integrated circuits contained about 10 devices on a small silicon chip—about 3 mm (0.12 in) square. By 1966, silicon had overtaken germanium as the preferred semiconductor material. Since that year, Si has been the predominant material in IC fabrication. Since the 1960s, a continual trend toward miniaturization and increased integration of multiple devices in a single chip has occurred in the electronics industry (the progress can be seen in Table 33.1), leading to the components described in this chapter.

FIGURE 33.1 Cross section of a transistor (specifically, a MOSFET) in an integrated circuit. Approximate size of the device is shown. Processing technologies at time of publication are approaching feature sizes in ICs as small as 22 nm.

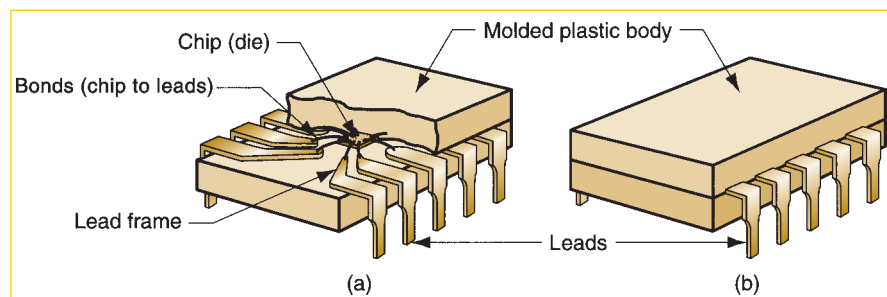


33.1 Overview of IC Processing

Structurally, an integrated circuit consists of hundreds, thousands, millions, or billions of microscopic electronic devices that have been fabricated and electrically interconnected within the surface of a silicon chip. A **chip**, also called a **die**, is a square or rectangular flat plate that is about 0.5 mm (0.020 in) thick and typically 5 to 25 mm (0.200–1.0 in) on a side. Each electronic device (i.e., transistor, diode, etc.) on the chip surface consists of separate layers and regions with different electrical properties combined to perform the particular electronic function of the device. A typical cross section of such a MOSFET² device is illustrated in Figure 33.1. The approximate size of the device is shown, but the feature sizes within it are smaller. Current technology allows feature sizes as small as 32 nm, and 22 nm will be achieved within a few years. The devices are electrically connected to one another by very fine lines of conducting material, so that the interconnected devices (that is, the integrated circuit) function in the specified way. Conducting lines and pads are also provided to electrically connect the IC to leads, which in turn permit the IC to be connected to external circuits. MOSFET is the most important device technology for ultra large scale integration [15].

To connect the IC to the outside world, and to protect it from damage, the chip is attached to a lead frame and encapsulated inside a suitable package, as in Figure 33.2. The package is an enclosure, usually made of plastic or ceramic, which provides

FIGURE 33.2 Packaging of an integrated circuit chip: (a) cutaway view showing the chip attached to a lead frame and encapsulated in a plastic enclosure, and (b) the package as it would appear to a user. This type of package is called a dual in-line package (DIP).



²MOSFET stands for Metal-Oxide-Semiconductor Field-Effect Transistor. A transistor is a semiconductor device capable of performing various functions such as amplifying, controlling, or generating electrical signals. A field-effect transistor is one in which current flows between source and drain regions through a channel, the flow depending on the application of voltage to the channel gate. A metal-oxide-semiconductor FET uses silicon dioxide to separate the channel and gate metallization.

mechanical and environmental protection for the chip and includes leads by which the IC can be electrically connected to external circuits. The leads are attached to conducting pads on the chip that access the IC.

33.1.1 PROCESSING SEQUENCE

The sequence to fabricate a silicon-based IC chip begins with the processing of silicon (Section 7.5.2). Briefly, silicon of very high purity is reduced in several steps from sand (silicon dioxide, SiO_2). The silicon is grown from a melt into a large solid single-crystal log, with typical length of 1 to 3 m (3–10 ft) and diameter up to 450 mm (18 in). The log, called a *boule*, is then sliced into thin wafers, which are disks of thickness equal to about 0.5 mm (0.020 in).

After suitable finishing and cleaning, the wafers are ready for the sequence of processes by which microscopic features of various chemistries will be created in their surface to form the electronic devices and their interconnections. The sequence consists of several types of processes, most of them repeated many times. A total of 200 or more processing steps may be required to produce a modern IC. Basically, the objective in the sequence is to add, alter, or remove a layer of material in selected regions of the wafer surface. The layering steps in IC fabrication are sometimes referred to as the *planar process*, because the processing relies on the geometric form of the silicon wafer being a plane. The processes by which the layers are added include thin film deposition techniques such as physical vapor deposition and chemical vapor deposition (Section 27.5), and existing layers are altered by diffusion and ion implantation (Section 27.2). Additional layer-forming techniques, such as thermal oxidation, are also employed. Layers are removed in selected regions by etching, using chemical etchants (usually acid solutions) and other more advanced technologies such as plasma etching.

The addition, alteration, and removal of layers must be done selectively; that is, only in certain extremely small regions of the wafer surface to create the device details such as in Figure 33.1. To distinguish which regions will be affected in each processing step, a procedure involving *lithography* is used. In this technique, masks are formed on the surface to protect certain areas and allow other areas to be exposed to the particular process (e.g., film deposition, etching). By repeating the steps many times, exposing different areas in each step, the starting silicon wafer is gradually transformed into many integrated circuits.

Processing of the wafer is organized in such a way that many individual chip surfaces are formed on a single wafer. Because the wafer is round with diameters ranging from 150 to 450 mm (6 to 18 in), whereas the final chip may only be 12 mm (0.5 in) square, it is possible to produce hundreds of chips on a single wafer. At the conclusion of planar processing, each IC on the wafer is visually and functionally tested, the wafer is cut into individual chips, and each chip that passes the quality test is packaged as in Figure 33.2.

Summarizing the preceding discussion, the production of silicon-based integrated circuits consists of the following stages, portrayed in Figure 33.3: (1) *Silicon processing*, in which sand is reduced to very pure silicon and then shaped into wafers; (2) *IC fabrication*, consisting of multiple processing steps that add, alter, and remove thin layers in selected regions to form the electronic devices; lithography is used to define the regions to be processed on the surface of the wafer; and (3) *IC packaging*, in which the wafer is tested, cut into individual dies (IC chips), and the dies

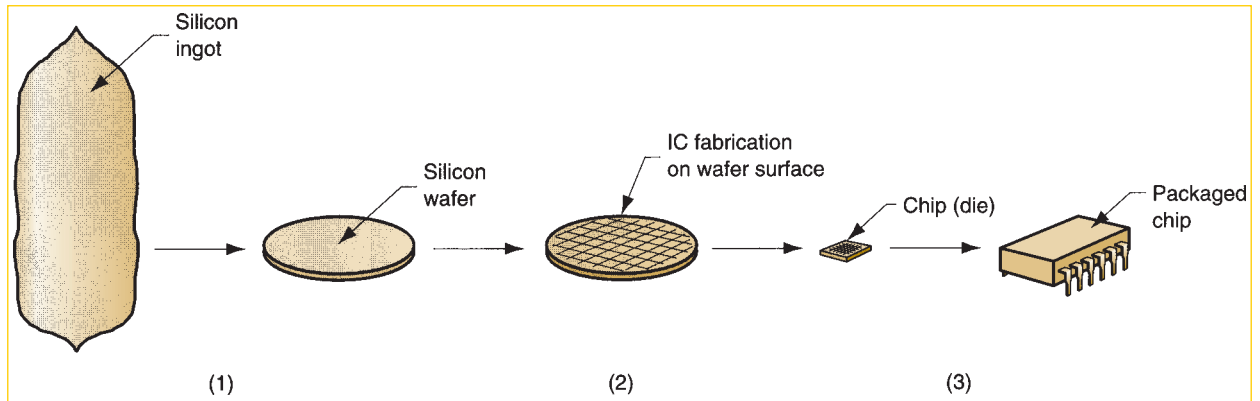


FIGURE 33.3 Sequence of processing steps in the production of integrated circuits: (1) pure silicon is formed from the molten state into an ingot and then sliced into wafers; (2) fabrication of integrated circuits on the wafer surface; and (3) wafer is cut into chips and packaged.

are encapsulated in an appropriate package. Subsequent sections of the chapter are concerned with the details of these processing stages. Before beginning the coverage of the processing details, it is important to note that the microscopic dimensions of the devices in integrated circuits impose special requirements on the environment in which IC fabrication is accomplished.

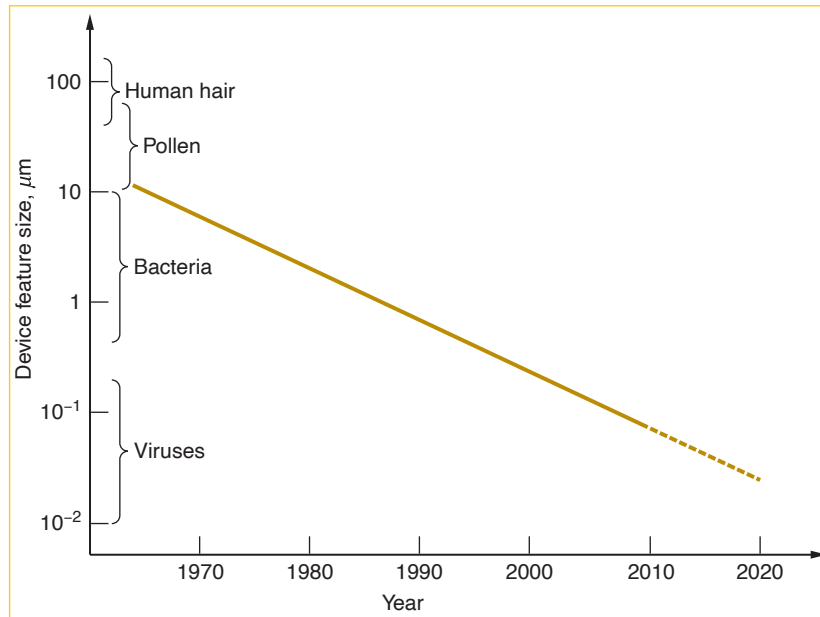
33.1.2 CLEAN ROOMS

Much of the processing sequence for integrated circuits must be carried out in a clean room, the ambience of which is more like a hospital operating room than a production factory. Cleanliness is dictated by the microscopic feature sizes in an IC, the scale of which continues to decrease with each passing year. Figure 33.4 shows the trend in IC device feature size; also displayed are common airborne particles that are potential contaminants in IC processing. These particles can cause defects in the integrated circuits, reducing yields and increasing costs.

A clean room provides protection from these contaminants. The air is purified to remove most of the particles from the processing environment; temperature and humidity are also controlled. The clean room is air conditioned to a temperature of 21°C (70°F) and 45% relative humidity. The air is passed through a high-efficiency particulate air (HEPA) filter to capture particle contaminants. Several classification systems are used to specify the cleanliness of a clean room. Two are outlined here: ISO and US [18]. In both systems, a number is used to indicate the number of particles of size $0.5\ \mu\text{m}$ or greater in a specified volume of air. In the ISO system, the volume of air is 1 cubic meter, whereas 1 cubic foot is used in the US system.³ A ISO class 5 clean room is required to maintain a count of particles of size $0.5\ \mu\text{m}$ or greater at less than $3520/\text{m}^3$. That corresponds to a US class 100 clean room which must maintain a count of particles of size $0.5\ \mu\text{m}$ or greater at less than $100/\text{ft}^3$. ($1\ \text{ft} = 0.3048\ \text{m}$, so $1\ \text{m} = (0.3048)^{-1} = 3.28\ \text{ft}$, and $1\ \text{m}^3 = (3.28)^3 \approx 35.2\ \text{ft}^3$. Thus, 100 particles per cubic foot is equivalent to 3520 particles per cubic meter.)

³Only in the United States would metric units ($0.5\ \mu\text{m}$) be mixed with U.S. customary units (ft^3).

FIGURE 33.4 Trend in device feature size in IC fabrication; also shown is the size of common airborne particles that can contaminate the processing environment. Minimum feature sizes for logic type ICs are expected to be about 13 nm in the year 2016 [10].



Modern VLSI processing requires ISO class 4 or US class 10 clean rooms, which means that the number of particles of size equal to or greater than $0.5 \mu\text{m}$ is less than $352/\text{m}^3$ or $10/\text{ft}^3$. By comparison, outside air in a typical urban atmosphere contains 35,000,000 particles/ m^3 or 1,000,000 particles/ ft^3 of size equal to or greater than $0.5 \mu\text{m}$ [18].

Humans are the biggest source of contaminants in IC processing; emanating from humans are bacteria, tobacco smoke, viruses, hair, and other particles. Human workers in IC processing areas are required to wear special clothing, generally consisting of white cloaks, gloves, and hair nets. Where extreme cleanliness is required, workers are completely encased in bunny suits. Processing equipment is a second major source of contaminants; machinery produces wear particles, oil, dirt, and similar contaminants. IC processing is usually accomplished in laminar-flow hooded work areas, which can be purified to greater levels of cleanliness than the general environment of the clean room.

In addition to the very pure atmosphere provided by the clean room, the chemicals and water used in IC processing must be very clean and free of particles. Modern practice requires that chemicals and water be filtered before using them.

33.2 Silicon Processing

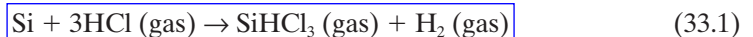
Microelectronic chips are fabricated on a substrate of semiconductor material. Silicon (Si) is the leading semiconductor material today, constituting more than 95% of all semiconductor devices produced in the world. The discussion in this introductory treatment is limited to Si. The preparation of the silicon substrate can be divided into three steps: (1) production of electronic grade silicon, (2) crystal growing, and (3) shaping of Si into wafers.

33.2.1 PRODUCTION OF ELECTRONIC GRADE SILICON

Silicon is one of the most abundant materials in the Earth's crust (Table 7.1), occurring naturally as silica (e.g., sand) and silicates (e.g., clay). Electronic grade silicon (EGS) is polycrystalline silicon of ultra high purity—so pure that the impurities are in the range of parts per billion (ppb). They cannot be measured by conventional chemical laboratory techniques but must be inferred from measurements of resistivity on test ingots. The reduction of the naturally occurring Si compound to EGS involves the following processing steps.

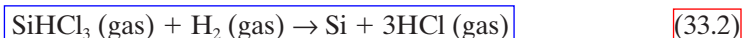
The first step is carried out in a submerged-electrode arc furnace. The principal raw material for silicon is *quartzite*, which is very pure SiO_2 . The charge also includes coal, coke, and wood chips as sources of carbon for the various chemical reactions that occur in the furnace. The net product consists of metallurgical grade silicon (MGS), and the gases SiO and CO . MGS is only about 98% Si, which is adequate for metallurgical alloying but not for electronics components. The major impurities (the remaining 2% of MGS) include aluminum, calcium, carbon, iron, and titanium.

The second step involves grinding the brittle MGS and reacting the Si powders with anhydrous HCl to form trichlorsilane:



The reaction is performed in a fluidized-bed reactor at temperatures around 300°C ($\sim 570^\circ\text{F}$). Trichlorsilane (SiHCl_3), although shown as a gas in Equation (33.1), is a liquid at room temperature. Its low boiling point of 32°C (90°F) permits it to be separated from the leftover impurities of MGS by fractional distillation.

The final step in the process is reduction of the purified trichlorsilane by means of hydrogen gas. The process is carried out at temperatures up to 1000°C ($\sim 1800^\circ\text{F}$), and a simplified equation of the reaction can be written as follows:



The product of this reaction is electronic grade silicon—nearly 100% pure Si.

33.2.2 CRYSTAL GROWING

The silicon substrate for microelectronic chips must be made of a single crystal whose unit cell is oriented in a certain direction. The properties of the substrate and the way it is processed are both influenced by these requirements. Accordingly, silicon used as the raw material in semiconductor device fabrication must not only be of ultra high purity, as in electronic grade silicon; it must also be prepared in the form of a single crystal and then cut in a direction that achieves the desired planar orientation. The crystal-growing process is covered here, and the next section details the cutting operation.

The most widely used crystal-growing method in the semiconductor industry is the *Czochralski process*, illustrated in Figure 33.5, in which a single crystal ingot, called a *boule*, is pulled upward from a pool of molten silicon.⁴ The setup includes a furnace, a mechanical apparatus for pulling the boule, a vacuum system, and

⁴The process is named after J. Czochralski, a Polish scientist who discovered it in 1916 while doing research on metal crystallization [15].

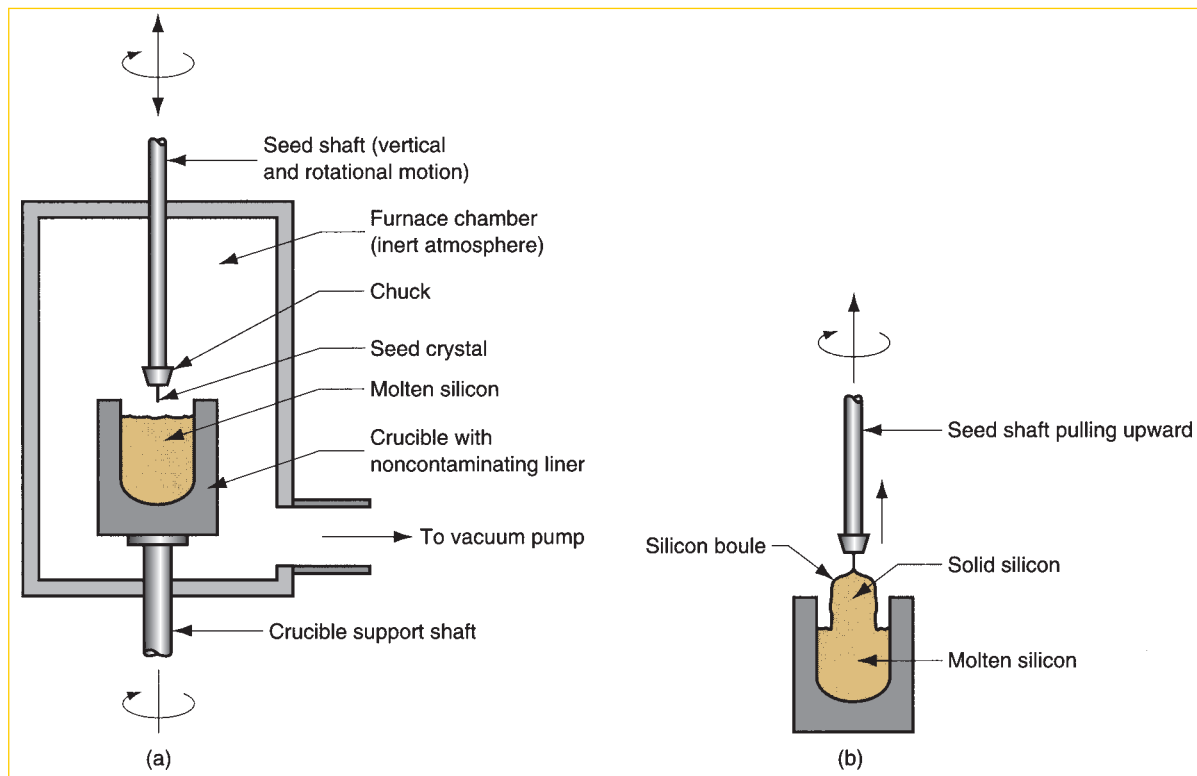


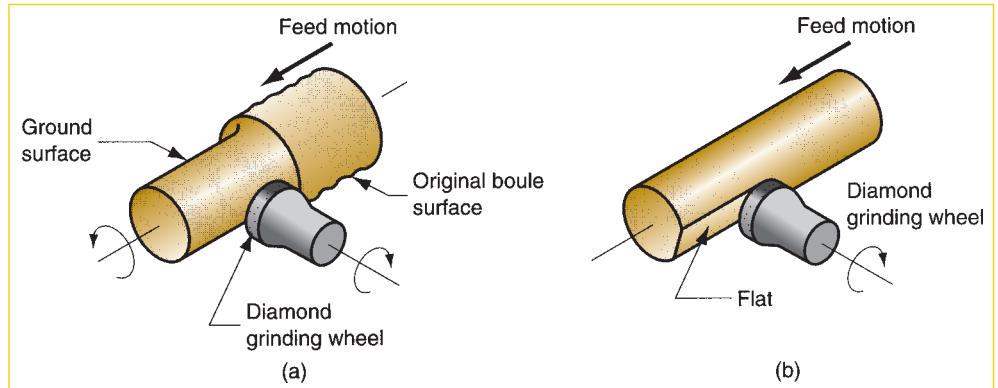
FIGURE 33.5 The Czochralski process for growing single-crystal ingots of silicon: (a) initial setup prior to start of crystal pulling, and (b) during crystal pulling to form the boule.

supporting controls. The furnace consists of a crucible and heating system contained in a vacuum chamber. The crucible is supported by a mechanism that permits rotation during the crystal-pulling procedure. Chunks of EGS are placed in the crucible and heated to a temperature slightly above the melting point of silicon: 1410°C (2570°F). Heating is by induction or resistance, the latter being used for large melt sizes. The molten silicon is doped[§] before pulling begins to make the crystal either p-type or n-type.

To initiate crystal growing, a seed crystal of silicon is dipped into the molten pool and then withdrawn upward under carefully controlled conditions. At first the pulling rate (vertical velocity of the pulling apparatus) is relatively rapid, which causes a single crystal of silicon to solidify against the seed, forming a thin neck. The velocity is then reduced, causing the neck to grow into the desired larger diameter of the boule while maintaining its single crystal structure. In addition to pulling rate, rotation of the crucible and other process parameters are used to control boule size. Single-crystal ingots of diameter $= 450\text{ mm}$ (18 in) and up to 3 m (10 ft) long can be produced for subsequent fabrication of microelectronic chips.

[§]The term **dope** (doped, doping) refers to the introduction of impurities into the semiconductor material to alter its electrical properties, making the semiconductor either an n-type (excess electrons in its structure) or a p-type (missing electrons in its structure).

FIGURE 33.6 Grinding operations used in shaping the silicon ingot: (a) a form of cylindrical grinding provides diameter and roundness control, and (b) a flat is ground on the cylinder.



It is important to avoid contamination of the silicon during crystal growing, because contaminants, even in small amounts, can dramatically alter the electrical properties of Si. To minimize unwanted reactions with silicon and the introduction of contaminants at the elevated temperatures of crystal growing, the procedure is carried out either in an inert gas (argon or helium) or a vacuum. Choice of crucible material is also important; fused silica (SiO_2), although not perfect for the application, represents the best available material and is used almost exclusively. Gradual dissolution of the crucible introduces oxygen as an unwanted impurity in the silicon boule. Unfortunately, the level of oxygen in the melt increases during the process, leading to a variation in concentration of the impurity throughout the length and diameter of the ingot.

33.2.3 SHAPING OF SILICON INTO WAFERS

A series of processing steps are used to reduce the boule into thin, disc-shaped wafers. The steps can be grouped as follows: (1) ingot preparation, (2) wafer slicing, and (3) wafer preparation. In ingot preparation, the seed and tang ends of the ingot are first cut off, as well as portions of the ingot that do not meet the strict resistivity and crystallographic requirements for subsequent IC processing. Next, a form of cylindrical grinding, as shown in Figure 33.6(a), is used to shape the ingot into a more perfect cylinder, because the crystal-growing process cannot achieve sufficient control over diameter and roundness. One or more flats are then ground along the length of the ingot, as in Figure 33.6(b). After the wafers have been cut from the ingot, these flats serve several functions: (1) identification, (2) orientation of the ICs relative to crystal structure, and (3) mechanical location during processing.

The ingot is now ready to be sliced into wafers, using the abrasive cutoff process illustrated in Figure 33.7. A very thin saw blade with diamond grit bonded to the

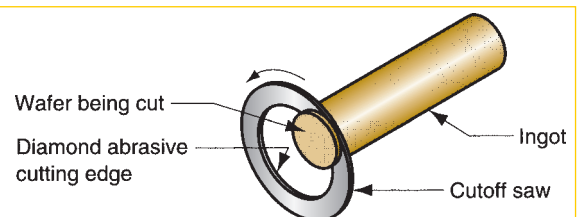


FIGURE 33.7 Wafer slicing using a diamond abrasive cutoff saw.

internal diameter serves as the cutting edge. Use of the ID for slicing rather than the OD of the saw blade provides better control over flatness, thickness, parallelism, and surface characteristics of the wafer. The wafers are cut to a thickness of around 0.4 to 0.7 mm (0.016–0.028 in), depending on diameter (greater thicknesses for larger wafer diameters). For every wafer cut, a certain amount of silicon is wasted because of the kerf width of the saw blade. To minimize kerf loss, the blades are made as thin as possible—around 0.33 mm (0.013 in).

Next the wafer must be prepared for the subsequent processes and handling in IC fabrication. After slicing, the rims of the wafers are rounded using a contour-grinding operation, such as in Figure 33.8 (a). This reduces chipping of the wafer edges during handling and minimizes accumulation of photoresist solutions at the wafer rims. The wafers are then chemically etched to remove surface damage that occurred during slicing. This is followed by a flat polishing operation to provide very smooth surfaces that will accept the subsequent optical lithography processes. The polishing step, seen in Figure 33.8(b), uses a slurry of very fine silica (SiO_2) particles in an aqueous solution of sodium hydroxide (NaOH). The NaOH oxidizes the Si wafer surface, and the abrasive particles remove the oxidized surface layers—about 0.025 mm (0.001 in) is removed from each side during polishing. Finally, the wafer is chemically cleaned to remove residues and organic films.

It is of interest to know how many IC chips can be fabricated on a wafer of a given size. The number depends on the chip size relative to the wafer size. Assuming that

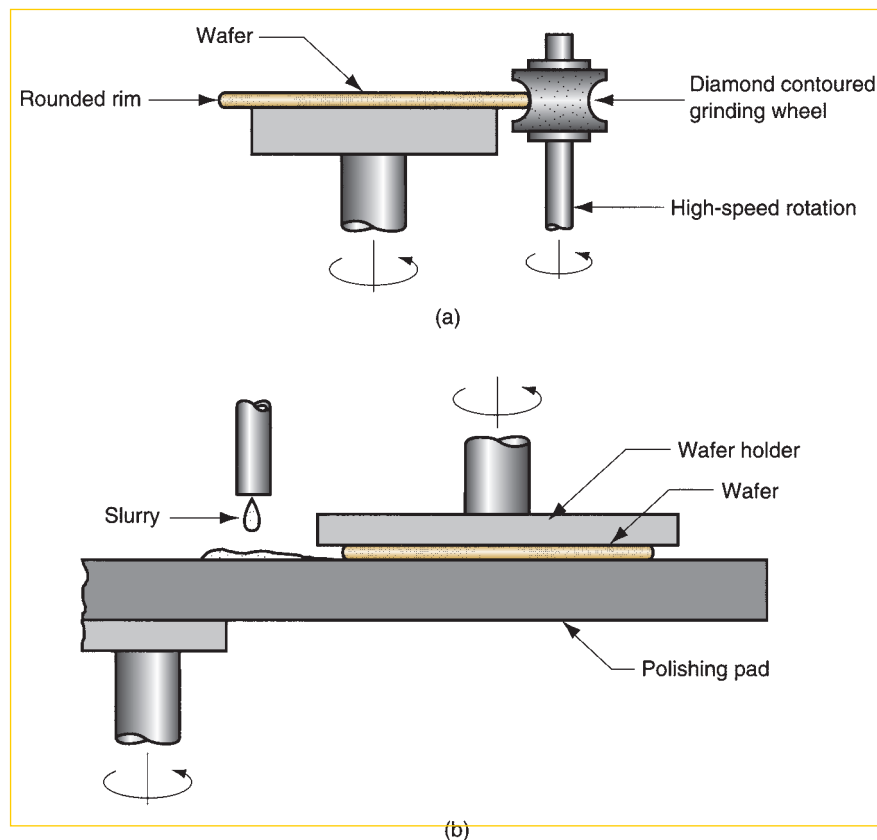


FIGURE 33.8 Two of the steps in wafer preparation: (a) contour grinding to round the wafer rim, and (b) surface polishing.

the chips are square, the following equation can be used to estimate the number of chips on the wafer:

$$n_c = 0.34 \left(\frac{D_w}{L_c} \right)^{2.25} \quad (33.3)$$

where n_c = estimated number of chips on the wafer; D_w = diameter of the processable area of the wafer, assumed circular, mm (in); and L_c = side dimension of the chip, assumed square, mm (in). The diameter of the processable area of the wafer will be slightly less than the outside diameter of the wafer. The actual number of chips on the wafer may be different from the value given by Equation (33.3), depending on the way the chips are laid out on the wafer.

Example 33.1 Number of chips on wafer

A 200-mm diameter silicon wafer has a processable area whose diameter is 190 mm. The IC chips to be fabricated on the wafer surface are square with 18 mm on a side. How many IC chips can be placed onto the wafer?

Solution: $n_c = 0.34 \left(\frac{190}{18} \right)^{2.25} = 0.34(10.56)^{2.25} = 68.3$ round to **68 chips**.

33.3 Lithography

An IC consists of many microscopic regions on the wafer surface that make up the transistors, other devices, and interconnections in the circuit design. In the planar process, the regions are fabricated by a sequence of steps that add, alter, or remove layers in selected areas of the surface. The form of each layer is determined by a geometric pattern representing circuit design information that is transferred to the wafer surface by a procedure known as lithography—basically the same procedure used by artists and printers for centuries.

Several lithographic technologies are used in semiconductor processing: (1) optical lithography, also known as **photolithography**; (2) electron-beam lithography, (3) X-ray lithography, and (4) ion beam lithography. As their names indicate, the differences are in the type of radiation used to transfer the mask pattern to the surface by exposing the photoresist. The traditional technique is optical lithography, and most of the discussion will be directed at this topic.

33.3.1 OPTICAL LITHOGRAPHY

Optical lithography uses light radiation to expose a coating of photoresist on the surface of the silicon wafer; a mask containing the required geometric pattern for each layer separates the light source from the wafer, so that only the portions of the photoresist not blocked by the mask are exposed. The **mask** consists of a flat plate of transparent glass onto which a thin film of an opaque substance has been deposited in certain areas to form the desired pattern. Thickness of the glass plate is around 2 mm (0.080 in), whereas the deposited film is only a few μm thick—for some film materials, less than one μm . The mask itself is fabricated by lithography, the pattern being based on circuit design data, usually in the form of digital output from the CAD system used by the circuit designer.

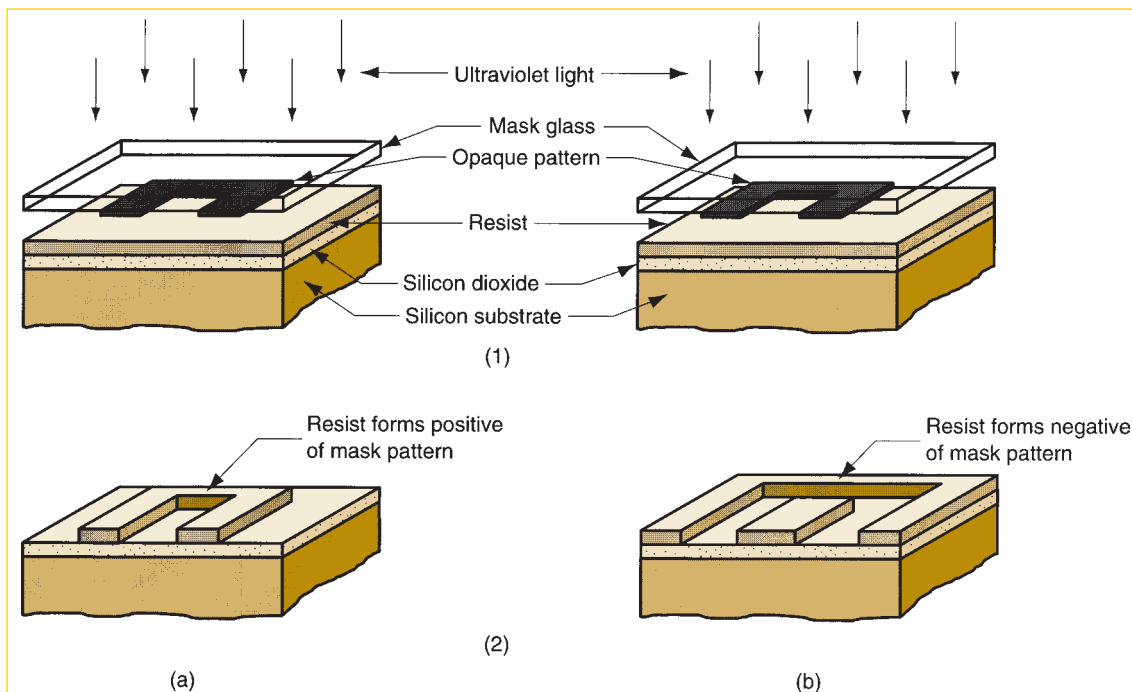


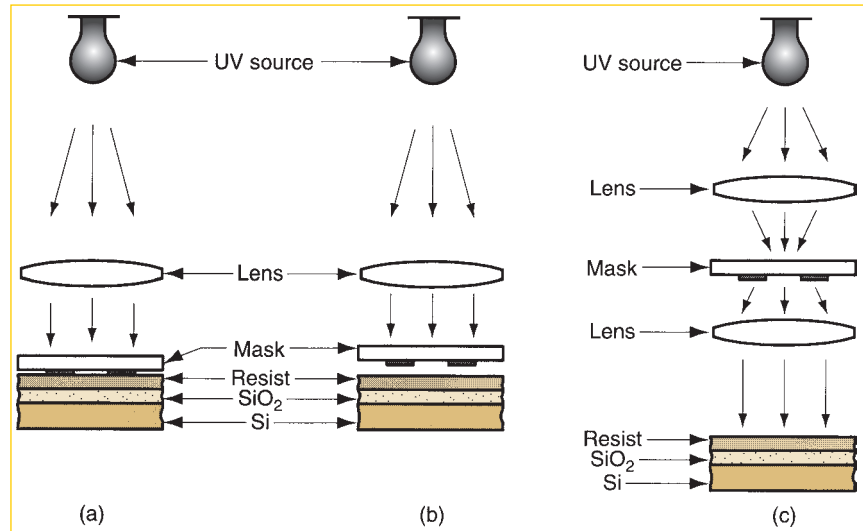
FIGURE 33.9 Application of (a) positive resist and (b) negative resist in optical lithography; for both types, the sequence shows: (1) exposure through the mask and (2) remaining resist after developing.

Photoresists A photoresist is an organic polymer that is sensitive to light radiation in a certain wavelength range; the sensitivity causes either an increase or decrease in solubility of the polymer to certain chemicals. Typical practice in semiconductor processing is to use photoresists that are sensitive to ultraviolet (UV) light. UV light has a short wavelength compared with visible light, permitting sharper imaging of microscopic circuit details on the wafer surface. It also permits the fabrication and photoresist areas in the plant to be illuminated at low light levels above the UV band.

Two types of photoresists are available: positive and negative. A **positive resist** becomes more soluble in developing solutions after exposure to light. A **negative resist** becomes less soluble (the polymer cross-links and hardens) when exposed to light. Figure 33.9 illustrates the operation of both resist types. Negative resists have better adhesion to SiO_2 and metal surfaces and good etch resistance. However, positive resists achieve better resolution, which has made it the more widely used technique as IC feature sizes have become smaller and smaller.

Exposure Techniques The resists are exposed through the mask by one of three exposure techniques: (a) contact printing, (b) proximity printing, and (c) projection printing, illustrated in Figure 33.10. In **contact printing**, the mask is pressed against the resist coating during exposure. This results in high resolution of the pattern onto the wafer surface; an important disadvantage is that physical contact with the wafers gradually wears out the mask. In **proximity printing**, the mask is separated from the resist coating by a distance of 10 to 25 μm (0.0004–0.001 in). This eliminates mask wear, but resolution of the image is slightly reduced. **Projection printing** involves the use of a high-quality lens (or mirror) system to project the image through the

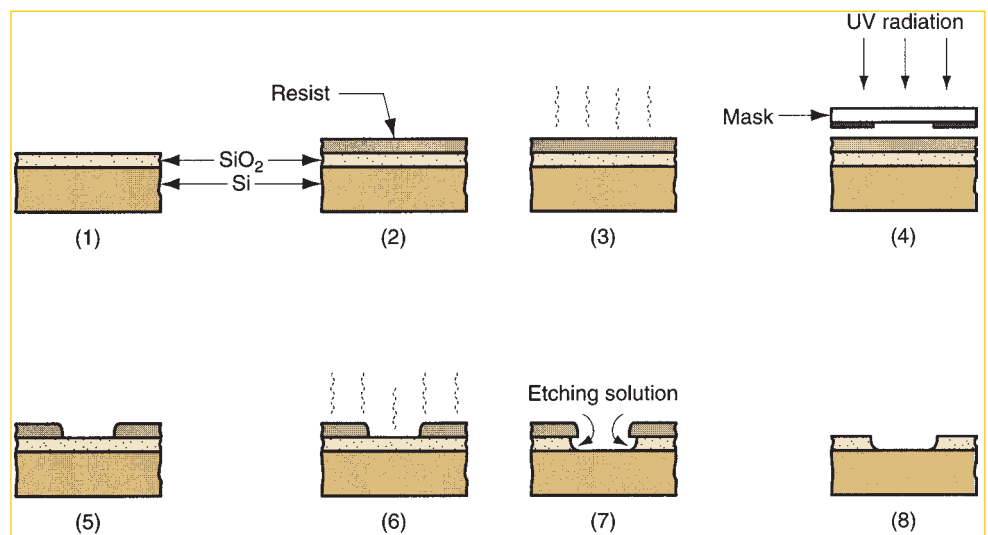
FIGURE 33.10 Optical lithography exposure techniques: (a) contact printing, (b) proximity printing, and (c) projection printing.



mask onto the wafer. This has become the preferred technique because it is non-contact (thus, no mask wear), and the mask pattern can be reduced through optical projection to obtain high resolution.

Processing Sequence in Optical lithography A typical processing sequence begins with surface of the silicon having just been oxidized to form a thin film of SiO₂ on the wafer. It is desired to remove the SiO₂ film in certain regions as defined by the mask pattern. The sequence for a positive resist proceeds as follows, illustrated in Figure 33.11. (1) **Prepare surface.** The wafer is properly cleaned to promote wetting and adhesion of the resist. (2) **Apply photoresist.** In semiconductor processing, photoresists are applied by feeding a metered amount of liquid resist onto the center of the wafer and then spinning the wafer to spread the liquid and achieve a uniform coating thickness. Desired thickness is around 1 μm (0.00004 in), which gives good resolution yet minimizes pinhole defects. (3) **Soft-bake.** The purpose of this

FIGURE 33.11 Optical lithography process applied to a silicon wafer: (1) prepare surface; (2) apply photoresist; (3) soft-bake; (4) align mask and expose; (5) develop resist; (6) hard-bake; (7) etch; (8) strip resist.



pre-exposure bake is to remove solvents, promote adhesion, and harden the resist. Typical soft-bake temperatures are around 90°C (190°F) for 10 to 20 min. (4) **Align mask and expose.** The pattern mask is aligned relative to the wafer and the resist is exposed through the mask by one of the methods described above. Alignment must be accomplished with very high precision, using optical-mechanical equipment designed specifically for the purpose. If the wafer has been previously processed by lithography so that a pattern has already been formed in the wafer, then subsequent masks must be accurately registered relative to the existing pattern. Exposure of the resist depends on the same basic rule as in photography—the exposure is a function of light intensity \times time. A mercury arc lamp or other source of UV light is used. (5) **Develop resist.** The exposed wafer is next immersed in a developing solution, or the solution is sprayed onto the wafer surface. For the positive resist in the example, the exposed areas are dissolved in the developer, thus leaving the SiO_2 surface uncovered in these areas. Development is usually followed by a rinse to stop development and remove residual chemicals. (6) **Hard-bake.** This baking step expels volatiles remaining from the developing solution and increases adhesion of the resist especially at the newly created edges of the resist film. (7) **Etch.** Etching removes the SiO_2 layer at selected regions where the resist has been removed. (8) **Strip resist.** After etching, the resist coating that remains on the surface must be removed. Stripping is accomplished using either wet or dry techniques. Wet stripping uses liquid chemicals; a mixture of sulfuric acid and hydrogen peroxide ($\text{H}_2\text{SO}_4 - \text{H}_2\text{O}_2$) is common. Dry stripping uses plasma etching with oxygen as the reactive gas.

Although the example describes the use of optical lithography to remove a thin film of SiO_2 from a silicon substrate, the same basic procedure is followed for other processing steps. The purpose of optical lithography in all of these steps is to expose specific regions beneath the photoresist layer so that some process can be performed on these exposed regions. In the processing of a given wafer, optical lithography is repeated as many times as needed to produce the desired integrated circuit, each time using a different mask to define the appropriate pattern. Figure 33.12 shows a silicon wafer that has been partially processed using optical lithography.

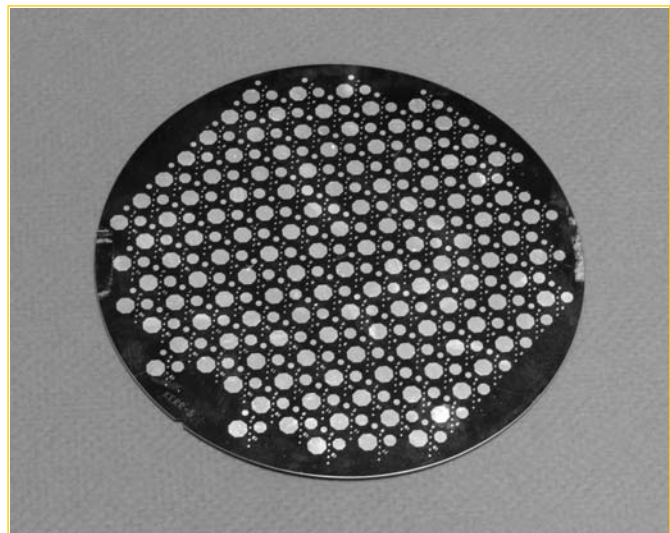


FIGURE 33.12 A partially processed silicon wafer after several optical lithography steps. (Courtesy of George E. Kane Manufacturing Technology Laboratory, Lehigh University.)

33.3.2 OTHER LITHOGRAPHY TECHNIQUES

As feature size in integrated circuits continues to decrease and conventional UV optical lithography becomes increasingly inadequate, other lithography techniques that offer higher resolution are growing in importance. These techniques are extreme ultraviolet lithography, electron beam lithography, X-ray lithography, and ion beam lithography. The following paragraphs provide brief descriptions of these alternatives. For each technique, special resist materials are required that react to the particular type of radiation.

Extreme ultraviolet lithography (EUV) represents a refinement of conventional UV lithography through the use of shorter wavelengths during exposure. The ultraviolet wavelength spectrum ranges from about 10 nm to 380 nm (nm = nanometer = 10^{-9} m), the upper end of which is close to the visible light range ($\sim 400\text{--}700$ nm wavelengths). EUV uses wavelengths in the range 10 nm to 14 nm, which permits the feature size of an integrated circuit to be reduced to about $0.05\text{ }\mu\text{m}$ (50 nm) [15]. This compares with about $0.1\text{ }\mu\text{m}$ (100 nm) using conventional UV exposure.

Compared with UV and EUV lithography, **electron-beam (E-Beam) lithography** virtually eliminates diffraction during exposure of the resist, thus permitting higher resolution of the image. Another potential advantage is that a scanning E-beam can be directed to expose only certain regions of the wafer surface, thus eliminating the need for a mask. Unfortunately, high-quality electron-beam systems are expensive. Also, because of the time-consuming sequential nature of the exposure method, production rates are low compared with the mask techniques of optical lithography. Accordingly, use of E-beam lithography tends to be limited to small production quantities. E-beam techniques are widely used for making the masks in UV lithography.

X-ray lithography has been under development since around 1972. As in E-beam lithography, the wavelengths of X-rays are much shorter than UV light (X-ray wavelength ranges from 0.005 nm to several dozen nm, overlapping the lower end of the UV range). Thus, they hold the promise of sharper imaging during exposure of the resist. X-rays are difficult to focus during lithography. Consequently, contact or proximity printing must be used, and a small X-ray source must be used at a relatively large distance from the wafer surface to achieve good image resolution through the mask.

Ion beam lithography systems divide into two categories: (1) focused ion beam systems, whose operation is similar to a scanning E-beam system and avoids the need for a mask; and (2) masked ion beam systems, which expose the resist through a mask by proximity printing. As with E-beam and X-ray systems, ion lithography produces higher image resolution than conventional UV optical lithography.

33.4 Layer Processes Used in IC Fabrication

The steps required to produce an integrated circuit consist of chemical and physical processes that add, alter, or remove regions on the silicon wafer that have been defined by optical lithography. These regions constitute the insulating, semiconducting, and conducting areas that form the devices and their interconnections in the integrated circuits. The layers are fabricated one at a time, step by step, each layer having a different configuration and each requiring a separate mask, until all of the microscopic details of the electronic devices and conducting paths have been constructed on the wafer surface.

This section considers the wafer processes used to add, alter, and subtract layers. Processes that add or alter layers to the surface include (1) thermal oxidation, used to grow a layer of silicon dioxide onto the silicon substrate; (2) chemical vapor deposition, a versatile process used to apply various types of layers in IC fabrication; (3) diffusion and ion implantation, used to alter the chemistry of an existing layer or substrate; and (4) various metallization processes that add metal layers to provide regions of electrical conduction on the wafer. Finally, (5) several etching processes are used to remove portions of the layers that have been added to achieve the desired details of the integrated circuit.

33.4.1 THERMAL OXIDATION

Oxidation of the silicon wafer may be performed multiple times during fabrication of an integrated circuit. Silicon dioxide (SiO_2) is an insulator, contrasted with the semiconducting properties of Si. The ease with which a thin film of SiO_2 can be produced on the surface of a silicon wafer is one of the attractive features of silicon as a semiconductor material.

Silicon dioxide serves a number of important functions in IC fabrication: (1) It is used as a mask to prevent diffusion or ion implantation of dopants into silicon. (2) It is used to isolate devices in the circuit. (3) It provides electrical insulation between levels in multilevel metallization systems.

Several processes are used to form SiO_2 in semiconductor manufacturing, depending on when during chip fabrication the oxide must be added. The most common process is thermal oxidation, appropriate for growing SiO_2 films on silicon substrates. In **thermal oxidation**, the wafer is exposed to an oxidizing atmosphere at elevated temperature; either oxygen or steam atmospheres are used, with the following reactions, respectively:

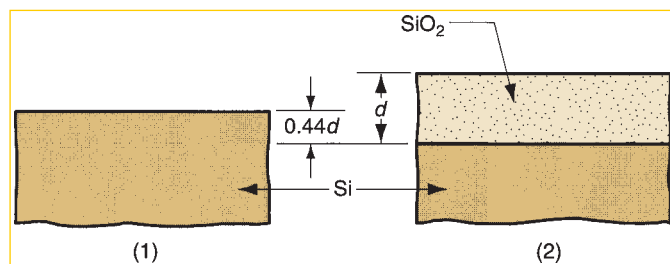


or



Typical temperatures used in thermal oxidation of silicon range from 900°C to 1200°C (1650°F to 2200°F). By controlling temperature and time, oxide films of predictable thicknesses can be obtained. The equations show that silicon at the surface of the wafer is consumed in the reaction, as seen in Figure 33.13. To grow a SiO_2 film of thickness d requires a layer of silicon that is $0.44d$ thick.

FIGURE 33.13 Growth of SiO_2 film on a silicon substrate by thermal oxidation, showing changes in thickness that occur: (1) before oxidation, and (2) after thermal oxidation.



When a silicon dioxide film must be applied to surfaces other than silicon, then direct thermal oxidation is not appropriate. An alternative process must be used, such as chemical vapor deposition.

33.4.2 CHEMICAL VAPOR DEPOSITION

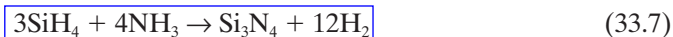
Chemical vapor deposition (CVD) involves growth of a thin film on the surface of a heated substrate by chemical reactions or decomposition of gases (Section 27.5.2). CVD is widely used in the processing of integrated circuit wafers to add layers of silicon, silicon dioxide, silicon nitride (Si_3N_4), and various metallization materials (discussed below). Plasma-enhanced CVD is often used because it permits the reactions to take place at lower temperatures.

Typical CVD Reactions in IC Fabrication In the case of silicon dioxide, if the surface of the wafer is only silicon (e.g., at the start of IC fabrication), then thermal oxidation is the appropriate process by which to form a layer of SiO_2 . If the oxide layer must be grown over materials other than silicon, such as aluminum or silicon nitride, then some alternative technique must be used, such as CVD. Chemical vapor deposition of SiO_2 is accomplished by reacting a silicon compound such as silane (SiH_4) with oxygen onto a heated substrate. The reaction is carried out at around 425°C ($\sim 800^\circ\text{F}$) and can be summarized:



The density of the silicon dioxide film and its bonding to the substrate is generally poorer than that achieved by thermal oxidation. Consequently, CVD is used only when the preferred process is not feasible; that is, when the substrate surface is not silicon, or when the high temperatures used in thermal oxidation cannot be tolerated. CVD can be used to deposit layers of doped SiO_2 , such as phosphorous-doped silicon dioxide (called P-glass).

Silicon nitride is used as a masking layer during oxidation of silicon. Si_3N_4 has a low oxidation rate compared with Si, so a nitride mask can be used to prevent oxidation in coated areas on the silicon surface. Silicon nitride is also used as a passivation layer (protecting against sodium diffusion and moisture). A conventional CVD process for coating Si_3N_4 onto a silicon wafer involves reaction of silane and ammonia (NH_3) at around 750°C ($\sim 1400^\circ\text{F}$):



Plasma-enhanced CVD is also used for basically the same coating reaction, the advantage being that it can be performed at much lower temperatures—around 300°C ($\sim 570^\circ\text{F}$).

Polycrystalline silicon (called **polysilicon** to distinguish it from silicon having a single crystal structure such as the boule and wafer) has a number of uses in IC fabrication, including [14]: conducting material for leads, gate electrodes in MOS devices, and contact material in shallow junction devices. Chemical vapor deposition to coat polysilicon onto a wafer involves reduction of silane at temperatures around 600°C ($\sim 1100^\circ\text{F}$):



Epitaxial Deposition A related process for growing a film onto a substrate is epitaxial deposition, in which the film has a crystalline structure that is an extension of the substrate's structure. If the film material is the same as the substrate (e.g., silicon on silicon), then its crystal lattice will be identical to and a continuation of the wafer crystal. Two primary techniques to perform epitaxial deposition are vapor-phase epitaxy and molecular-beam epitaxy.

Vapor-phase epitaxy is the more important in semiconductor processing and is based on chemical vapor deposition. In growing silicon on silicon, the process is accomplished under closely controlled conditions at higher temperatures than conventional CVD of Si, using diluted reacting gases to slow the process so that an epitaxial layer can be successfully formed. Various reactions are possible, including Equation (33.8), but the most widely used industrial process involves hydrogen reduction of silicon tetrachloride gas (SiCl_4) at around 1200°C ($\sim 2200^\circ\text{F}$) as follows:



The melting point of silicon is 1410°C (2570°F), so the preceding reaction is carried out at temperatures below T_m for Si, considered an advantage for vapor-phase epitaxy.

Molecular-beam epitaxy uses a vacuum evaporation process (Section 27.5.1), in which silicon together with one or more dopants are vaporized and transported to the substrate in a vacuum chamber. Its advantage is that it can be carried out at lower temperatures than CVD; processing temperatures are 400°C to 900°C ($\sim 750^\circ\text{F}$ – $\sim 1650^\circ\text{F}$). However, throughput is relatively low and the equipment is expensive.

33.4.3 INTRODUCTION OF IMPURITIES INTO SILICON

IC technology relies on the ability to alter the electrical properties of silicon by introducing impurities into selected regions at the surface. Adding impurities into the silicon surface is called **doping**. The doped regions are used to create p-n junctions that form the transistors, diodes, and other devices in the circuit. A silicon-dioxide mask produced by thermal oxidation and optical lithography is used to isolate the silicon regions that are to be doped. Common elements used as impurities are boron (B) which forms electron acceptor regions in the silicon substrate (p-type regions); and phosphorous (P), arsenic (As), and antimony (Sb), which form electron donor regions (n-type regions). The predominant technique by which silicon is doped with these elements is ion implantation.

In ion implantation (Section 27.2.2), vaporized ions of the impurity element are accelerated by an electric field and directed at the silicon substrate surface. The atoms penetrate into the surface, losing energy and finally stopping at some depth in the crystal structure, the average depth being determined by the mass of the ion and the acceleration voltage. Higher voltages produce greater depths of penetration, typically several hundred Angstroms ($1 \text{ Angstrom} = 10^{-4} \text{ mm} = 10^{-1} \text{ nm}$). Advantages of ion implantation are that it can be accomplished at room temperature and it provides exact doping density.

The problem with ion implantation is that the ion collisions disrupt and damage the crystal lattice structure. Very high energy collisions can transform the starting crystalline material into an amorphous structure. This problem is solved by annealing at temperatures between 500°C and 900°C (930°F and 1650°F), which allows the lattice structure to repair itself and return to its crystal state.

33.4.4 METALLIZATION

Conductive materials must be deposited onto the wafer during processing to serve several functions: (1) forming certain components (e.g., gates) of devices in the IC; (2) providing conduction paths between devices on the chip; and (3) connecting the chip to external circuits. To satisfy these functions the conducting materials must be formed into very fine patterns. The process of fabricating these patterns is known as **metallization**, and it combines various thin film deposition technologies with optical lithography. This section is concerned with the materials and processes used in metallization. Connecting the chip to external circuitry involves IC packaging, which is described in Section 33.6.

Metallization Materials Materials used in the metallization of silicon-based integrated circuits must have certain desirable properties, some of which relate to electrical function whereas others relate to manufacturing processing. The desirable properties of a metallization material are [5], [14]: (1) low resistivity; (2) low-contact resistance with silicon, (3) good adherence to the underlying material, usually Si or SiO_2 ; (4) ease of deposition, compatible with optical lithography; (5) chemical stability—noncorroding, nonreactive, and noncontaminating; (6) physical stability during temperatures encountered in processing; and (7) good lifetime stability.

Although no material meets all of these requirements perfectly, aluminum satisfies most of them either well or adequately, and it has been the most widely used metallization material. Other metallization materials include titanium, titanium nitride, and copper [15]. Aluminum is usually alloyed with small amounts of (1) silicon to reduce reactivity with silicon in the substrate, and (2) copper to inhibit electromigration of Al atoms caused by current flow when the IC is in service. Other materials used for metallization in integrated circuits include polysilicon (Si); gold (Au); refractory metals (e.g., W, Mo); silicides (e.g., WSi_2 , MoSi_2 , TaSi_2); and nitrides (e.g., TiN, TaN, and ZrN). These other materials are generally used in applications such as gates and contacts. Aluminum is generally favored for device interconnections and top level connections to external circuitry.

Metallization Processes Several processes are available to accomplish metallization in IC fabrication: physical vapor deposition, chemical vapor deposition, and electroplating. Among PVD processes, vacuum evaporation and sputtering are applicable (Section 27.5.1). **Vacuum evaporation** can be applied for aluminum metallization. Vaporization is usually accomplished by resistance heating or electron beam evaporation. Evaporation is difficult or impossible for depositing refractory metals and compounds. **Sputtering** can be used for depositing aluminum as well as refractory metals and certain metallizing compounds. It achieves better step coverage than evaporation, often important after many processing cycles when the surface contour has become irregular. However, deposition rates are lower and equipment is more expensive.

Chemical vapor deposition is also applicable as a metallization technique. Its processing advantages include excellent step coverage and good deposition rates. Materials suited to CVD include tungsten, molybdenum, titanium nitride, and most of the silicides used in semiconductor metallization. CVD for metallization in semiconductor processing is less common than PVD. Finally, **electroplating** (Section 27.3.1) is occasionally used in IC fabrication to increase the thickness of thin films.

33.4.5 ETCHING

All of the preceding processes in this section involve addition of material to the wafer surface, either in the form of a thin film or the doping of the surface with an impurity element. Certain steps in IC manufacturing require material removal from the surface; this is accomplished by etching away the unwanted material. Etching is usually done selectively, by coating surface areas that are to be protected and leaving other areas exposed for etching. The coating may be an etch-resistant photoresist, or it may be a previously applied layer of material such as silicon dioxide. Etching was briefly encountered in the previous discussion of optical lithography. This section gives some of the technical details of this step in IC fabrication.

The two main categories of etching process in semiconductor processing are wet chemical etching and dry plasma etching. Wet chemical etching is the older of the two processes and is easier to use. However, there are certain disadvantages that have resulted in growing use of dry plasma etching.

Wet Chemical Etching Wet chemical etching involves the use of an aqueous solution, usually an acid, to etch away a target material. The etching solution is selected because it chemically attacks the specific material to be removed and not the protective layer used as a mask. Some of the common etchants used to remove materials in wafer processing are listed in Table 33.2.

In its simplest form, the process can be accomplished by immersing the masked wafers in an appropriate etchant for a specified time and then immediately transferring them to a thorough rinsing procedure to stop the etching. Process variables such as immersion time, etchant concentration, and temperature are important in determining the amount of material removed. A properly etched layer will have a profile as shown in Figure 33.14. Note that the etching reaction is **isotropic** (it proceeds equally in all directions), resulting in an undercut below the protective mask. In general, wet chemical etching is isotropic, and so the mask pattern must be sized to compensate for this effect, just as in chemical machining (Section 25.4).

Note also that the etchant does not attack the layer below the target material in this illustration. In the ideal case, an etching solution can be formulated that will react only with the target material and not with other materials in contact with it. In practical cases, the other materials exposed to the etchant may be attacked but to a lesser degree than the target material. The **etch selectivity** of the etchant is the ratio of etching rates between the target material and some other material, such as a mask or substrate. For example, etch selectivity of hydrofluoric acid for SiO_2 over Si is infinite.

TABLE 33.2 Some common chemical etchants used in semiconductor processing.

Material to Be Removed	Etchant (usually in aqueous solution)
Aluminum (Al)	Mixture of phosphoric acid (H_3PO_4), nitric acid (HNO_3), and acetic acid (CH_3COOH).
Silicon (Si)	Mixture of nitric acid (HNO_3) and hydrofluoric acid (HF)
Silicon dioxide (SiO_2)	Hydrofluoric acid (HF)
Silicon nitride (Si_3N_4)	Hot phosphoric acid (H_3PO_4)

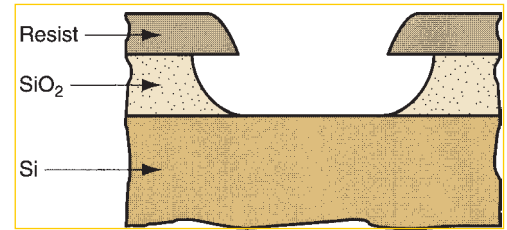


FIGURE 33.14 Profile of a properly etched layer.

If process control is inadequate, either under-etching or over-etching can occur, as in Figure 33.15. Underetching, in which the target layer is not completely removed, results when the etching time is too short and/or the etching solution is too weak. Over-etching involves too much of the target material being removed, resulting in loss of pattern definition and possible damage to the layer beneath the target layer. Over-etching is caused by overexposure to the etchant.

Dry Plasma Etching This etching process uses an ionized gas to etch a target material. The ionized gas is created by introducing an appropriate gas mixture into a vacuum chamber and using radio frequency (RF) electrical energy to ionize a portion of the gas, thus creating a plasma. The high-energy plasma reacts with the target surface, vaporizing the material to remove it. There are several ways in which a plasma can be used to etch a material; the two principal processes in IC fabrication are plasma etching and reactive ion etching.

In **plasma etching**, the function of the ionized gas is to generate atoms or molecules that are chemically very reactive, so that the target surface is chemically etched upon exposure. The plasma etchants are usually based on fluorine or chlorine gases. Etch selectivity is generally more of a problem in plasma etching than in wet chemical etching. For example, etch selectivity for SiO_2 over Si in a typical plasma etching process is 15 at best [4], compared with infinity when HF chemical etching is used.

An alternative function of the ionized gas can be to physically bombard the target material, causing atoms to be ejected from the surface. This is the process of sputtering, one of the techniques in physical vapor deposition. When used for etching, the process is called **sputter etching**. Although this form of etching has been applied in semiconductor processing, it is much more common to combine sputtering with plasma

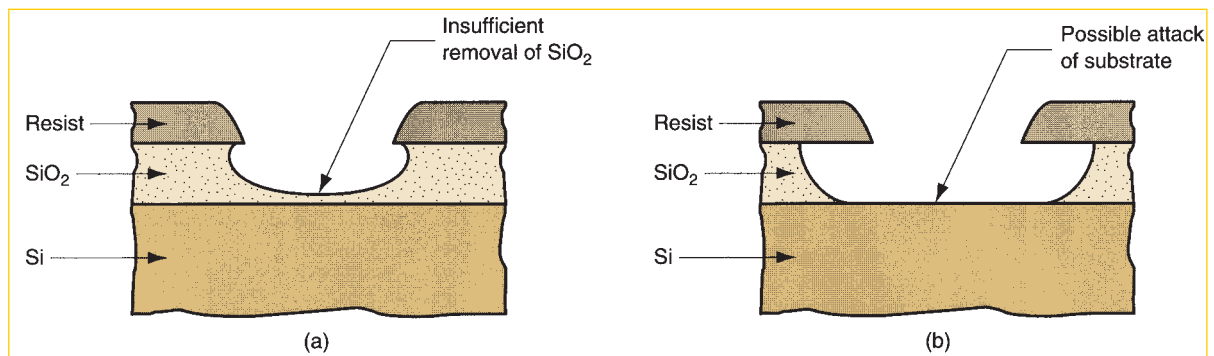
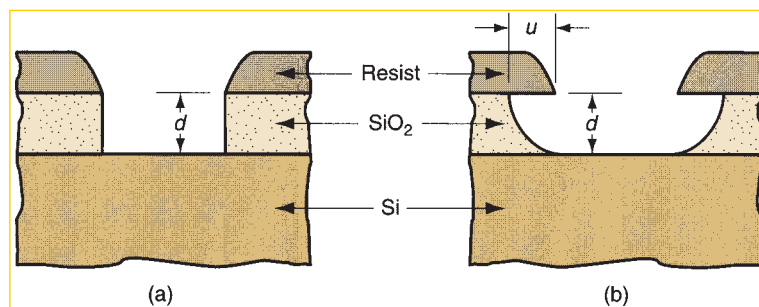


FIGURE 33.15 Two problems in etching: (a) under-etching and (b) over-etching.

FIGURE 33.16

(a) A fully anisotropic etch, with $A = \infty$, and (b) a partially anisotropic etch, with $A \approx 1.3$.



etching as described previously, which results in the process known as **reactive ion etching**. This produces both chemical and physical etching of the target surface.

The advantage of the plasma etching processes over wet chemical etching is that they are much more **anisotropic**. This property can be readily defined with reference to Figure 33.16. In (a), a fully anisotropic etch is shown; the undercut is zero. The degree to which an etching process is anisotropic is defined as the ratio:

$$A = \frac{d}{u} \quad (33.10)$$

where A = degree of anisotropy; d = depth of etch, which in most cases will be the thickness of the etched layer; and u = the undercut dimension, as illustrated in Figure 33.16(b). Wet chemical etching usually yields A values around 1.0, indicating isotropic etching. In sputter etching, ion bombardment of the surface is nearly perpendicular, resulting in A values approaching infinity—almost fully anisotropic. Plasma etching and reactive ion etching have high degrees of anisotropy, but below those achieved in sputter etching. As IC feature sizes continue to shrink, anisotropy becomes increasingly important for achieving the required dimensional tolerances.

33.5 Integrating the Fabrication Steps

Sections 33.3 and 33.4 examined the individual processing technologies used in IC fabrication. This section shows how these technologies are combined into the sequence of steps to produce an integrated circuit.

The planar processing sequence consists of fabricating a series of layers of various materials in selected areas on a silicon substrate. The layers form insulating, semi-conducting, or conducting regions on the substrate to create the particular electronic devices required in the integrated circuit. The layers might also serve the temporary function of masking certain areas so that a particular process is only applied to desired portions of the surface. The masks are subsequently removed.

The layers are formed by thermal oxidation, epitaxial growth, deposition techniques (CVD and PVD), diffusion, and ion implantation. Table 33.3 summarizes the processes typically used to add or alter a layer of a given material type. The use of lithography to apply a particular process only to selected regions of the surface is illustrated in Figure 33.17.

An example will be useful here to show the process integration in IC fabrication. An n-channel metal oxide semiconductor (NMOS) logic device will be used

TABLE • 33.3 Layer materials added or altered in IC fabrication and associated processes.

Layer Material (function)	Typical Fabrication Processes
Si, polysilicon (semiconductor)	CVD
Si, epitaxial (semiconductor)	Vapor phase epitaxy
Si doping (n-type or p-type)	Ion implantation, diffusion
SiO ₂ (insulator, mask)	Thermal oxidation, CVD
Si ₃ N ₄ (mask)	CVD
Al (conductor)	PVD, CVD
P-glass (protection)	CVD

to illustrate the processing sequence. The sequence for NMOS integrated circuits is less complex than for CMOS or bipolar technologies, although the processes for these IC categories are basically similar. The device to be fabricated is illustrated in Figure 33.1.

The starting substrate is a lightly doped p-type silicon wafer, which will form the base of the n-channel transistor. The processing steps are illustrated in Figure 33.18 and described here (some details have been simplified, and the metallization process for interconnecting devices has been omitted): (1) A layer of Si₃N₄ is deposited by CVD onto the Si substrate using optical lithography to define the regions. This layer of Si₃N₄ will serve as a mask for the thermal oxidation process in the next step. (2) SiO₂ is grown in the exposed regions of the surface by thermal oxidation. The SiO₂ regions are insulating and will become the means by which this device is isolated from other devices in the circuit. (3) The Si₃N₄ mask is stripped by etching. (4) Another thermal oxidation is done to add a thin gate oxide layer to previously uncoated surfaces and increase the thickness of the previous SiO₂ layer. (5) Polysilicon is deposited by CVD onto the surface and then doped n-type using ion implantation. (6) The polysilicon is selectively etched using optical lithography to leave the gate electrode of the transistor. (7) The source and drain regions (n⁺) are formed by ion implantation of arsenic (As) into the substrate. An implantation energy level is selected that will penetrate the thin SiO₂ layer but not the polysilicon gate or the thicker SiO₂ isolation layer. (8) Phosphosilicate glass (P-glass) is deposited onto the surface by CVD to protect the circuitry beneath.

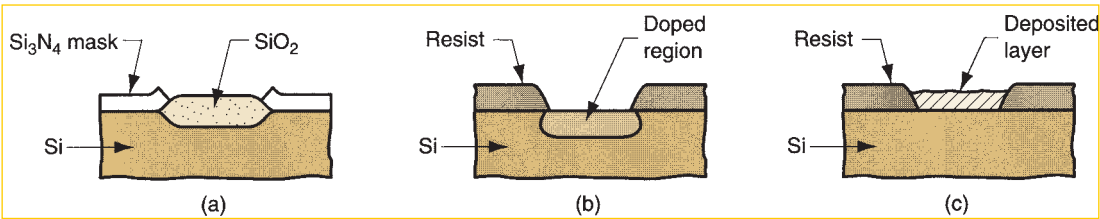


FIGURE 33.17 Formation of layers selectively through the use of masks: (a) thermal oxidation of silicon, (b) selective doping, and (c) deposition of a material onto a substrate.

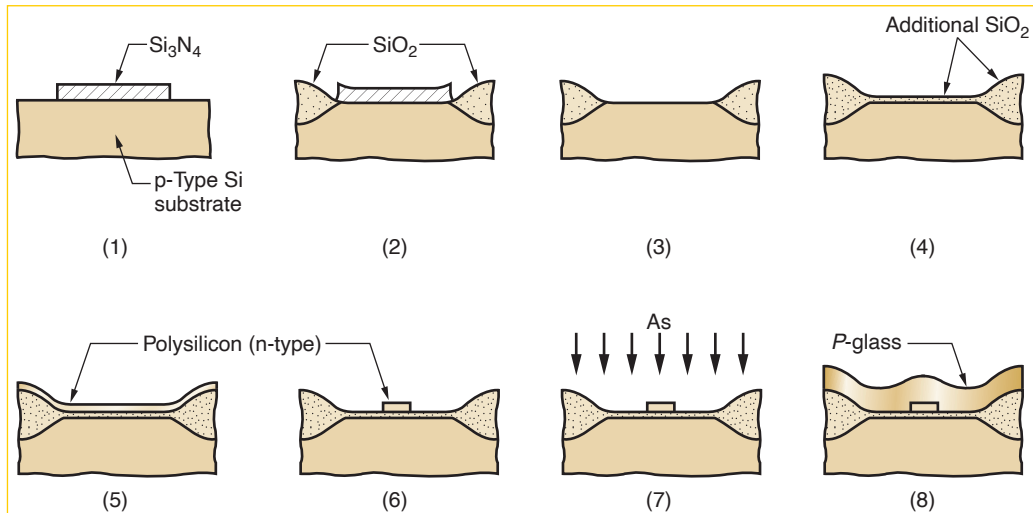


FIGURE 33.18 IC fabrication sequence: (1) Si_3N_4 mask is deposited by CVD on Si substrate; (2) SiO_2 is grown by thermal oxidation in unmasked regions; (3) the Si_3N_4 mask is stripped; (4) a thin layer of SiO_2 is grown by thermal oxidation; (5) polysilicon is deposited by CVD and doped n^+ using ion implantation; (6) the poly-Si is selectively etched using optical lithography to define the gate electrode; (7) source and drain regions are formed by doping n^+ in the substrate; (8) P-glass is deposited onto the surface for protection.

33.6 IC Packaging

After all of the processing steps on the wafer have been completed, a final series of operations must be accomplished to transform the wafer into individual chips, ready to connect to external circuits and prepared to withstand the harsh environment of the world outside the clean room. These final steps are referred to as IC packaging.

Packaging of integrated circuits is concerned with design issues such as (1) electrical connections to external circuits; (2) materials to encase the chip and protect it from the environment (humidity, corrosion, temperature, vibration, mechanical shock); (3) heat dissipation; (4) performance, reliability, and service life; and (5) cost.

There are also manufacturing issues in packaging, including: (1) chip separation—cutting the wafer into individual chips, (2) connecting it to the package, (3) encapsulating the chip, and (4) circuit testing. The manufacturing issues are the ones of greatest interest in this section. Although most of the design issues are properly left to other texts [8], [11], [13], and [15], some of the engineering aspects of IC packages, also known as **chip carriers**, and the types of packages are discussed here, before describing the processing steps to make them.

33.6.1 IC PACKAGE DESIGN

This section considers three topics related to the design of an integrated circuit package: (1) the number of input/output terminals required for an IC of a given size, (2) the materials used in IC packages, and (3) package styles.

Determining the Number of Input/Output Terminals The basic engineering problem in IC packaging is to connect the many internal circuits to input/output (I/O) terminals so that the appropriate electrical signals can be communicated

TABLE • 33.4 Typical values of Rent's Rule parameters C and m for several types of integrated circuits.

Integrated Circuit Type and Concise Definition ^a	Typical values ^b	
	Constant C	Exponent m
Microprocessor. An IC that functions as the central processing unit (CPU) of a computer. It is programmable and operates on instructions stored in internal memory to process digital data input and provide digital output.	0.89	0.45
Gate array. A partially-completed IC in which transistors and other devices are located at predefined locations so they can be interconnected by means of one or more final layering processes to provide specific functionality to the IC. Also known as an uncommitted logic array.	1.9	0.50
Static random access memory (SRAM). An IC memory device that stores data bits as long as power is supplied. It does not need to be periodically refreshed as in dynamic random access memory.	6.9	0.12
Dynamic random access memory (DRAM). An IC memory device that uses capacitors to store data. The capacitors require periodic recharging; hence the term dynamic. DRAM is a slower but less expensive data storage device compared to SRAM.	7.8	0.07

^aDefinitions were compiled from [19], [20], [22], and [24].^bValues compiled from data at [25].

between the IC and the outside world. As the number of devices in an IC increases, the required number of I/O terminals (leads) also increases. The problem is of course aggravated by trends in semiconductor technology that have led to decreases in device size and increases in the number of devices that can be packed into an IC. Fortunately, the number of I/O terminals does not have to equal the number of devices in the IC. The dependency between the two values is given by Rent's rule, named after the IBM engineer (E. F. Rent) who defined the following relationship around 1960:

$$n_{io} = C n_{ic}^m \quad (33.11)$$

where n_{io} = the required number of input/output terminals or external signal connections; n_{ic} = the number of internal circuits in the IC, usually taken to be the number of logic gates; and C and m are parameters in the equation.

The parameters in Rent's rule depend on the type of circuit and the design of the IC. Memory devices require far fewer I/O terminals than microprocessors due to their column and row structures. Typical values of the Constant C and exponent m for several common integrated circuit devices are listed in Table 33.4.

Example 33.2 Rent's rule

The 18-mm square chips from Example 33.1 have a processable area of 17 mm by 17 mm. The density of circuits (e.g., transistors) within each chip's processable area is 500 circuits per mm². Determine (a) the number of circuits (transistors) that can be placed on each chip and (b) the number of input/output terminals that would be required in the packaged microprocessor using Rent's rule. (c) If the packaged IC were a static random access memory device instead of a microprocessor, how many input/output terminals would be required? Refer to Table 33.4 for appropriate values of the parameters in Rent's rule.

Solution: (a) The processable area of each chip $= (17)^2 = 289 \text{ mm}^2$.

Number of circuits $n_{ic} = 289(500) = 144,500$.

(b) Rents rule with $C = 0.89$ and $m = 0.45$.

$n_{io} = Cn_{ic}^m = 0.89(144,500)^{0.45} = 187$ input/output terminals.

(c) Rents rule with $C = 6.9$ and $m = 0.12$.

$n_{io} = Cn_{ic}^m = 6.9(144,500)^{0.12} = 29$ input/output terminals.

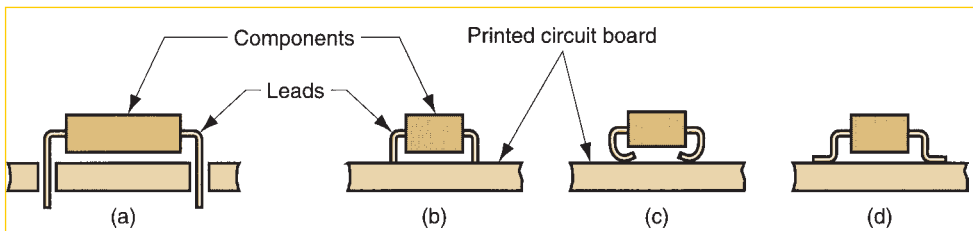
IC Package Materials Package sealing involves encapsulating the IC chip in an appropriate packaging material. Two material types dominate current packaging technology: ceramic and plastic. Metal was used in early packaging designs but is today no longer important, except for lead frames.

The common ceramic packaging material is alumina (Al_2O_3). Advantages of ceramic packaging include hermetic sealing of the IC chip and the fact that highly complex packages can be produced. Disadvantages include poor dimensional control because of shrinkage during firing and the high dielectric constant of alumina.

Plastic IC packages are not hermetically sealed, but their cost is lower than ceramic. They are generally used for mass produced ICs, where very high reliability is not required. Plastics used in IC packaging include epoxies, polyimides, and silicones.

IC Package Styles A wide variety of IC package styles is available to meet the input/output requirements indicated above. In nearly all applications, the IC is a component in a larger electronic system and must be attached to a printed circuit board (PCB). There are two broad categories of component mounting to a PCB, shown in Figure 33.19: through-hole and surface mount. In **through-hole technology**, also known as **pin-in-hole** (PIH) technology, the IC package and other electronic components (e.g., discrete resistors, capacitors) have leads that are inserted through holes in the board and are soldered on the underside. In **surface-mount technology** (SMT), the components are attached to the surface of the board (or in some cases, both top and bottom surfaces). Several lead configurations are available in SMT, as illustrated in (b), (c), and (d) of the figure. Today, most chip carriers are based on surface mount technology because it allows greater packing densities to be achieved in the circuit board assembly.

FIGURE 33.19 Types of component lead attachment on a printed circuit board: (a) through-hole, and several styles of surface-mount technology; (b) butt lead; (c) "J" lead; and (d) gull-wing.



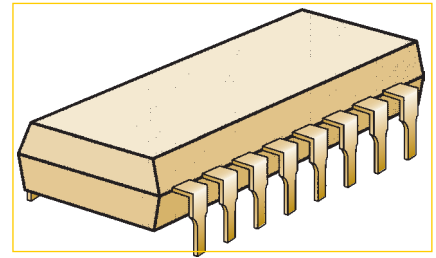


FIGURE 33.20 Dual in-line package with 16 terminals, shown as a through-hole configuration.

The major styles of IC packages include (1) dual in-line package, (2) square package, and (3) pin grid array. The **dual in-line package** (DIP) is a common form of IC package, available in both through-hole and surface-mount configurations. It has two rows of leads (terminals) on either side of a rectangular body, as in Figure 33.20. Spacing between leads (center-to-center distance) in the conventional through-hole DIP is 2.54 mm (0.1 in), and the number of leads ranges between 8 and 64. Hole spacing in the through-hole DIP style is limited by the ability to drill holes closely together in a printed circuit board. This limitation can be relaxed with surface-mount technology because the leads are not inserted into the board; standard lead spacing on surface-mount DIPs is 1.27 mm (0.05 in).

The number of terminals in a DIP is limited by its rectangular shape in which leads project only from two sides; that means that the number of leads on either side is $n_{io}/2$. For high values of n_{io} (between 48 and 64), differences in conducting lengths between leads in the middle of the DIP and those on the ends cause problems in high-speed electrical characteristics. Some of these problems are addressed with a square package, in which the leads are arranged around the periphery so that the number of terminals on a side is $n_{io}/4$. A surface-mount square package is illustrated in Figure 33.21.

Even with a square package, there is still a practical upper limit on terminal count dictated by the manner in which the leads in the package are linearly allocated. The number of leads on a package can be maximized by using a square matrix of leads. This chip carrier type is called a **pin grid array** (PGA) in through-hole technology and a **ball grid array** (BGA) in surface mount. In either case, the package consists of a two-dimensional array of terminals on the underside of a square chip enclosure. In the PGA, the terminals are pins that are inserted into through holes in the printed circuit board, whereas in the BGA, tiny balls of solder replace the pins and are soldered directly to copper pads on the PCB. Ideally, the entire bottom surface of the package is fully occupied by terminals, so that the terminal count in each direction is

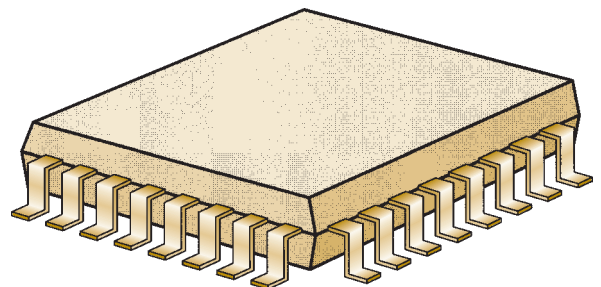


FIGURE 33.21 Square package for surface mounting with gull wing leads.

square root of n_i . However, as a practical matter, the center area of the package has no terminals because this region contains the IC chip.

33.6.2 PROCESSING STEPS IN IC PACKAGING

The packaging of an IC chip in manufacturing can be divided into the following steps: (1) wafer testing, (2) chip separation, (3) die bonding, (4) wire bonding, and (5) package sealing. After packaging, a final functional test is performed on each packaged IC.

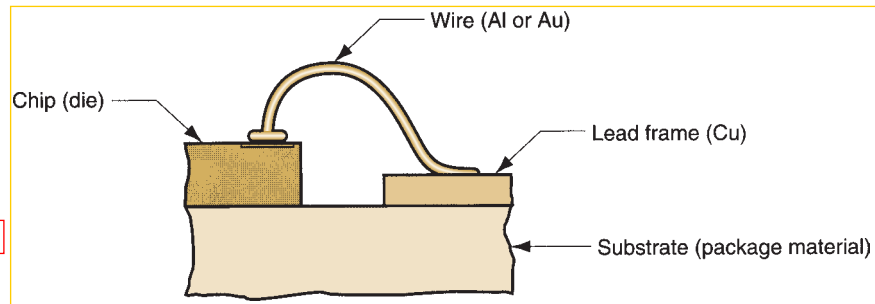
Wafer Testing Current semiconductor processing techniques provide many individual ICs per wafer. It is convenient to perform certain functional tests on the ICs while they are still together on the wafer—before chip separation. Testing is accomplished by computer-controlled test equipment that uses a set of needle probes configured to match the connecting pads on the surface of the chip; **multi-probe** is the term used for this testing procedure. When the probes contact the pads, a series of DC tests are carried out to indicate short circuits and other faults; this is followed by a functional test of the IC. Chips that fail the test are marked with an ink dot; these defects are not packaged. Each IC is positioned in its turn beneath the probes for testing, using a high precision x - y table to index the wafer from one chip site to the next.

Chip Separation The next step after testing is to cut the wafer into individual chips (dice). A thin diamond-impregnated saw blade is used to perform the cutting operation. The sawing machine is highly automatic and its alignment with the “streets” between circuits is very accurate. The wafer is attached to a piece of adhesive tape that is mounted in a frame. The adhesive tape holds the individual chips in place during and after sawing; the frame is a convenience in subsequent handling of the chips. Chips with ink dots are now discarded.

Die Bonding The individual chips must next be attached to their individual packages, a procedure called die bonding. Owing to the miniature size of the chips, automated handling systems are used to pick the separated chips from the tape frame and place them for die bonding. Various techniques have been developed to bond the chip to the packaging substrate; two methods are described here: eutectic die bonding and epoxy die bonding. **Eutectic die bonding**, used for ceramic packages, consists of (1) depositing a thin film of gold on the bottom surface of the chip; (2) heating the base of the ceramic package to a temperature above 370°C ($\sim 700^{\circ}\text{F}$), the eutectic temperature of the Au—Si system; and (3) bonding the chip to the metallization pattern on the heated base. In **epoxy die bonding**, used for plastic VLSI packaging, a small amount of epoxy is dispensed on the package base (the lead frame), and the chip is positioned on the epoxy; the epoxy is then cured, bonding the chip to the surface.

Wire Bonding After the die is bonded to the package, electrical connections are made between the contact pads on the chip surface and the package leads. The connections are generally made using small-diameter wires of aluminum or gold, as illustrated in Figure 33.22. Typical wire diameters for aluminum are 0.05 mm (0.002 in), and gold wire diameters are about half that diameter (Au has higher

FIGURE 33.22 Typical wire connection between chip contact pad and lead.



electrical conductivity than Al, but is more expensive). Aluminum wires are bonded by ultrasonic bonding, whereas gold wires are bonded by thermocompression, thermosonic, or ultrasonic means. **Ultrasonic bonding** uses ultrasonic energy to weld the wire to the pad surface. **Thermocompression bonding** involves heating the end of the wire to form a molten ball, and then the ball is pressed into the pad to form the bond. **Thermosonic bonding** combines ultrasonic and thermal energies to form the bond. Automatic wire bonding machines are used to perform these operations at rates up to 200 bonds per minute.

Package Sealing As mentioned, the two common packaging materials are ceramic and plastic. The processing methods are different for the two materials. **Ceramic packages** are made from a dispersion of ceramic powder (Al_2O_3 is most common) in a liquid binder (e.g., polymer and solvent). The mix is first formed into thin sheets and dried, and then cut to size. Holes are punched for interconnections. The required wiring paths are then fabricated onto each sheet, and metal is filled into the holes. The sheets are then laminated by pressing and sintering to form a monolithic (single stone) body.

Two types of **plastic package** are available, postmolded and premolded. In **post-molded packages**, an epoxy plastic is transfer molded around the assembled chip and lead frame (after wire bonding), in effect transforming the pieces into one solid body. However, the molding process can be harsh on the delicate bond wires, and premolded packages are an alternative. In **premolded packaging**, an enclosure base is molded before encapsulation and then the chip and lead frame are connected to it, adding a solid lid or other material to provide protection.

Final Testing Upon completion of the packaging sequence, each IC must undergo a final test, the purpose of which is to (1) determine which units, if any, have been damaged during packaging; and (2) measure performance characteristics of each device.

Burn-in test procedures sometimes include elevated temperature testing, in which the packaged IC is placed in an oven at temperatures around 125°C (250°F) for 24 hours and then tested. A device that fails such a test would have been likely to have failed early during service. If the device is intended for environments in which wide temperature variations occur, a temperature cycle test is appropriate. This test subjects each device to a series of temperature reversals, between values around -50°C (-60°F) on the lower side and 125°C (250°F) on the upper side. Additional tests for devices requiring high reliability might include mechanical vibration tests and hermetic (leak) tests.

33.7 Yields in IC Processing

The fabrication of integrated circuits consists of many processing steps performed in sequence. In wafer processing in particular, there may be hundreds of distinct operations through which the wafer passes. At each step, there is a chance that something may go wrong, resulting in the loss of the wafer or portions of it corresponding to individual chips. A simple probability model to predict the final yield of good product is

$$Y = Y_1 Y_2 \dots Y_n$$

where Y = final yield; Y_1, Y_2, \dots, Y_n are the yields of each processing step; and n = total number of steps in the processing sequence.

As a practical matter, this model, although perfectly valid, is difficult to use because of the large number of steps involved and the variability of yields for each step. It is more convenient to divide the processing sequence into major phases (see Figure 33.3) and to define the yields for each phase. The first phase involves growth of the single-crystal boule. The term **crystal yield** Y_c refers to the amount of single-crystal material in the boule compared with the starting amount of electronic grade silicon. The typical crystal yield is about 50%. After crystal growing, the boule is sliced into wafers, the yield for which is described as the **crystal-to-slice yield** Y_s . This depends on the amount of material lost during grinding of the boule, the width of the saw blade relative to the wafer thickness during slicing, and other losses. A typical value might be 50%, although much of the lost silicon during grinding and slicing is recyclable.

The next phase is wafer processing to fabricate the individual ICs. From a yield viewpoint, this can be divided into wafer yield and multiprobe yield. **Wafer yield** Y_w refers to the number of wafers that survive processing compared to the starting quantity. Certain wafers are designated as test pieces or similar uses and therefore result in losses and a reduction in yield; in other cases, wafers are broken or processing conditions go awry. Typical values of wafer yield are around 70% if testing losses are included. For wafers that come through processing and are multiprobe tested, only a certain proportion pass the test, called the **multiprobe yield** Y_m . Multiprobe yield is highly variable and can range from very low values (< 10%) to relatively high values (> 90%), depending on IC complexity and worker skill in the processing areas.

Following packaging, final testing of the IC is performed. This invariably produces additional losses, resulting in a **final test yield** Y_t in the range 90% to 95%. If the five phase yields are combined, the final yield can be estimated by

$$Y = Y_c Y_s Y_w Y_m Y_t \quad (33.12)$$

Given the typical values at each step, the final yield compared with the starting amount of silicon is quite low.

The heart of IC fabrication is wafer processing, the yield from which is measured in multiprobe testing Y_m . Yields in the other areas are fairly predictable, but not in wafer-fab. Two types of processing defects can be distinguished in wafer processing: (1) area defects and (2) point defects. **Area defects** are those that affect major areas

of the wafer, possibly the entire surface. These are caused by variations or incorrect settings in process parameters. Examples include added layers that are too thin or too thick, insufficient diffusion depths in doping, and over- or under-etching. In general these defects are correctable by improved process control or development of alternative processes that are superior. For example, doping by ion implantation has largely replaced diffusion, and dry plasma etching has been substituted for wet chemical etching to better control feature dimensions.

Point defects occur at very localized areas on the wafer surface, affecting only one or a limited number of ICs in a particular area. They are commonly caused by dust particles either on the wafer surface or the lithographic masks. Point defects also include dislocations in the crystal lattice structure (Section 2.3.2). These point defects are distributed in some way over the surface of the wafer, resulting in a yield that is a function of the density of the defects, their distribution over the surface, and the processed area of the wafer. If the area defects are assumed negligible, and the point defects are assumed uniform over the surface area of the wafer, the resulting yield can be modeled by the equation

$$Y_m = \frac{1}{1 + AD} \quad (33.13)$$

where Y_m = the yield of good chips as determined in multiprobe; A = the area processed, cm^2 (in^2); and D = density of point defects, defects/ cm^2 (defects/ in^2). This equation is based on **Bose-Einstein** statistics and has been found to be a good predictor of wafer processing performance, especially for highly integrated chips (VLSI and beyond).

Example 33.3 Yield in wafer processing

A silicon wafer with a diameter of 200 mm is processed over a circular area whose diameter = 190 mm. The chips to be fabricated are square with 10 mm on a side. From previous experience, the density of point defects in the surface area is 0.002 defects/ cm^2 . Determine an estimate of the number of good chips using the Bose-Einstein yield computation.

Solution: $n_c = 0.34(190/10)^{2.25} = 0.34(19)^{2.25} = 256$ chips.

Processable wafer area $A = \pi(190)^2/4 = 28,353 \text{ mm}^2 = 283.53 \text{ cm}^2$.

$$Y_m = \frac{1}{1 + 283.53(0.002)} = \frac{1}{1 + 0.567} = 0.638 = 63.8\% \text{ yield}$$

Number of good chips = $0.638(256) = 163.4$ rounded down to **163 good chips**.

Wafer processing is the key to successful fabrication of integrated circuits. For an IC producer to be profitable, high yields must be achieved during this phase of manufacturing. This is accomplished using the purest possible starting materials, the latest equipment technologies, good process control over the individual processing steps, maintenance of clean room conditions, and efficient and effective inspection and testing procedures.