6.6 The partial pressure of a material being deposited in a vacuum system must be well above the residual background gas pressure if reasonable deposition rates are to be achieved. What must the partial pressure of aluminum be to achieve a deposition rate of 100 nm/min? Assume close packing of spheres with a diameter of 5 Å, 100% adhesion of the impinging aluminum, and a temperature of 300 K.

6.7 A wafer 100 mm in diameter is mounted in an electron-beam evaporation system in which the spherical radius is 40 cm. Use Eq. (6.7) to estimate the worst-case variation in film thickness between the center and edges of the wafer for an evaporated aluminum film 1 μm thick.

6.8 Repeat Problem 6.7 for a 200-mm wafer mounted 50 cm from the e-beam source.

6.9 Electron-beam evaporation is going to be used to deposit a 0.6-μm-thick layer of aluminum on a 300-mm-diameter wafer. The thickness variation between the center and edges of the wafer is desired to be less than 0.05 μm. How far should the wafer be from the source?

6.10 An MBE system must operate under ultrahigh-vacuum conditions to prevent the formation of undesired atomic layers on the surface of the substrate. What pressure of oxygen can be permitted at 300 K if formation of a monolayer of contamination can be permitted after the sample has been in the chamber for no less than 4 hr?

6.11 **(a)** Calculate the growth rate of a silicon layer from an $SiCl_4$ source at 1200 °C. Use $h_g =$ 1 cm/sec, $k_s = 2 \times 10^6 \exp(-1.9/kT)$ cm/sec, and $N_g = 3 \times 10^{16}$ atoms/cm^3. (For silicon, $N = 5 \times 10^{22}$/cm^3.)

(b) What is the change in growth rate if the temperature is increased by 25 °C?

(c) At what temperature does $k_s = h_g$? What is the growth rate at this temperature?

(d) What is the value of E_A in Fig. 6.10?

6.12 Use Eqs. (6.31) and (6.32) to model the case of a 10-μm *n*-type epitaxial layer ($N_E = 1 \times 10^{16}$/cm^3) grown on a *p*-type substrate ($N_S = 1 \times 10^{18}$/cm^3). Plot the impurity profile in the epitaxial layer and substrate assuming that the layer was grown at a rate of 0.2 μm/min at a temperature of 1200 °C. Assume that boron and phosphorus are the impurities. Find the location of the *pn* junction.

6.13 Compare and discuss the advantages and disadvantages of evaporation, sputtering, and chemical vapor deposition.

6.14 **(a)** A 1-kg source of aluminum is used in an E-beam evaporation system. How many 100-mm wafers can be coated with a 1-μm Al film before the source material is exhausted? Assume that 15% of the evaporated aluminum actually coats a wafer. (The rest is deposited on the inside of the electron-beam system.)

(b) Repeat the process for a 300-mm wafer.

6.15 **(a)** A silicon wafer 100 mm in diameter is centered 200 mm above a small planar evaporation source. Calculate the ratio of thickness between the center and edges of the wafer using Eq. (6.7), following a 1-μm film deposition.

(b) Repeat the process for 200- and 300-mm wafers centered 40 cm above the small planar source.

6.16 Advanced CMOS processes often use lightly doped epitaxial layers grown on heavily doped substrates. Use Eq. (6.31) to predict the dopant profile in the epilayer if an intrinsic silicon layer is grown on top of a substrate that has a uniform concentration of 10^{20} As atoms/cm^3. Assume the layer thickness is 1 μm and that it is grown in $SiCl_4$ at 1,100 °C.

6.17 Repeat Problem 6.16 for the case where the intrinsic layer is grown on a substrate that has a uniform concentration of 2×10^{20} B atoms/cm^3. Assume the layer thickness is 2 μm and that it is grown in SiH_4Cl_2 at 950 °C.

CHAPTER 7

Interconnections and Contacts

The previous six chapters focused on the various processes used to fabricate semiconductor devices in the silicon substrate. To complete the formation of an integrated circuit, one must interconnect the devices and finally get connections to the world outside the silicon chip. Until the 1970s, integrated circuits had two possible levels of interconnection: diffusions and metallization. The use of polysilicon as a gate material in MOS devices added a third level useful for interconnecting devices and circuits.

In this chapter, we discuss the various forms of interconnections and the problems associated with making good contacts between metal and silicon. Refractory metal silicides and multilevel metallization used in VLSI processes are discussed, and an additional method for depositing patterned films, called *liftoff*, is also introduced.

Copper has been introduced in IC processing because of its lower resistivity. The CMP process, introduced in Chapter 3, is combined with standard electroplating techniques to achieve highly planar, inlaid copper interconnections referred to as Damascene technology. Low dielectric constant interlevel films are used to reduce the capacitance of the interconnection levels.

7.1 INTERCONNECTIONS IN INTEGRATED CIRCUITS

As we found in previous chapters, aluminum, polysilicon, and diffused regions are all easily isolated from each other using an insulating layer of silicon dioxide. Thus, today's integrated circuits have three different materials that may cross over each other. To be useful as an interconnect, the materials must also provide as low a sheet resistance as possible in order to minimize voltage drops along the interconnect lines, as well as to minimize propagation delay caused by the resistance and capacitance of the line. Finally, low-resistance "ohmic" contacts must be made between the materials, and the interconnection lines must be reliable throughout long-term operation.

Figure 7.1 shows a simple MOS logic circuit illustrating how polysilicon, metal, and diffused interconnections may cross over or contact each other. Aluminum is used

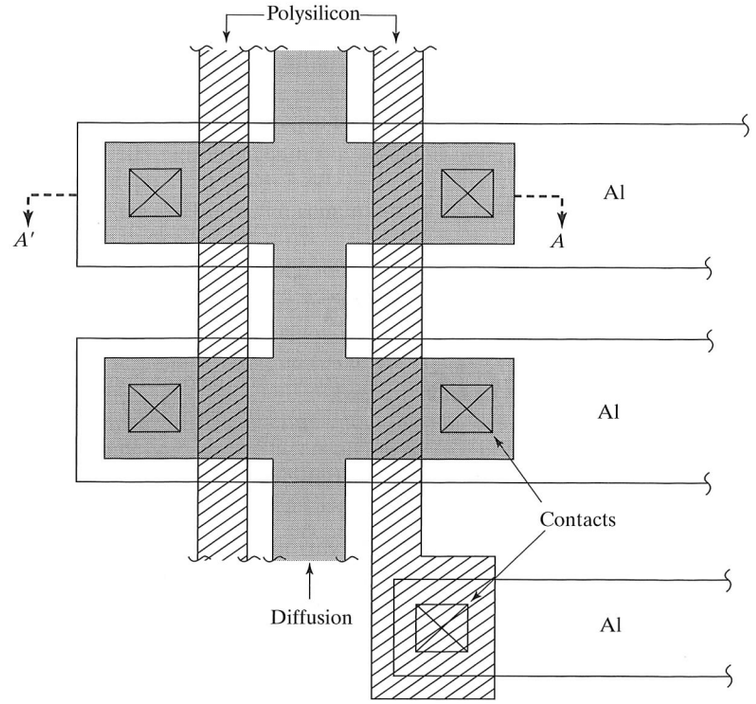


FIGURE 7.1

Portion of a MOS logic circuit showing the use of diffusion, polysilicon, and aluminum interconnections. (a) Top view; (b) cross section through A' -A.

to make contact to diffusions and polysilicon, and diffusions in various regions have been extended and merged together to form interconnections.

In this technology, polysilicon lines and diffused lines can only be connected together using the metal level. Improved circuit density can often be achieved by using "butted contacts" between polysilicon and diffusions or by changing the process to introduce "buried contacts" directly between the polysilicon and diffused layers. These two techniques will be examined later in this chapter.

7.2 METAL INTERCONNECTIONS AND CONTACT TECHNOLOGY

The requirement for low-resistivity materials leads one immediately to consider metals for use as interconnections, and the resistivities of common metals are compared in Table 7.1. Historically, aluminum and gold have been used with silicon IC processing. Gold requires the use of a multilayer sandwich involving other metals such as titanium or tungsten. Gold can be troublesome, because it is a rapid diffuser (see Fig. 4.5) in silicon and produces deep-level recombination centers in silicon that tend to significantly reduce the lifetime of free carriers. In addition, gold forms many problematic intermetallic compounds. Because of these various problems, the use of gold is most often restricted to chip packaging technologies.

Aluminum is compatible with silicon IC processing and is the most common material in use today. It is relatively inexpensive, adheres well to silicon dioxide, and has a bulk resistivity of 2.7 μohm-cm. However, care must be exercised to avoid a number of problems associated with the formation of good aluminum contacts to silicon.

Advanced multilevel metallization systems employ copper, because of its improved resistivity relative to aluminum. Copper has an even larger diffusion coefficient in silicon than gold and also causes lifetime reduction and leakage in silicon. Therefore, copper is generally not introduced into the fabrication sequence until one or two levels of aluminum metallization and interlevel dielectric levels have been formed above the semiconductor devices. These metallization and dielectric layers act as passivation layers to protect the active devices below.

7.2.1 Ohmic Contact Formation

We desire to form "ohmic" contacts between the metal and semiconductor. True ohmic contacts would exhibit a straight-line I–V characteristic with a low value of resistance (see Fig. 7.2(a)), as opposed to the I–V characteristic of a rectifying contact shown in Fig. 7.2(b). Figure 7.2(c) shows an I–V characteristic more representative of a practical ohmic contact to silicon. Although nonlinear near the origin, it develops only a small voltage across the contact at normal current levels.

Figure 7.3 shows a number of ways in which aluminum may contact semiconductor regions during device fabrication. Aluminum contact to p-type silicon normally results in

TABLE 7.1 Bulk Resistivity of Metals (μΩ-cm)

Ag: Silver	1.6
Al: Aluminum	2.65
Au: Gold	2.2
Co: Cobalt	6
Cu: Copper	1.7
Mo: Molybdenum	5
Ni: Nickel	7
Pd: Paladium	10
Pt: Platinum	10.6
Ti: Titanium	50
W: Tungsten	5

Source: WebElements (http://www.webelements.com)

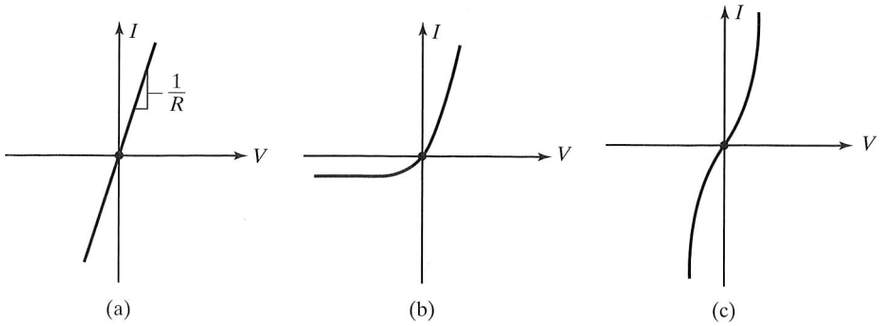


FIGURE 7.2

I-V characteristics of contacts between integrated-circuit materials. (a) Ideal ohmic contact; (b) rectifying contact; (c) practical nonlinear "ohmic" contact.

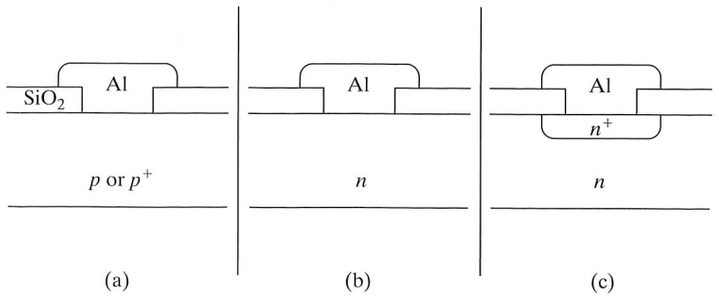


FIGURE 7.3

Three possible types of aluminum contacts to silicon. (a) Aluminum to *p*-type silicon forms an ohmic contact with an *I-V* characteristic approximating that in Fig. 7.2a; (b) aluminum to *n*-type silicon can form a rectifying contact (Schottky barrier diode) like that in Fig. 7.2b; (c) aluminum to n^+ silicon yields a contact similar to that in Fig. 7.2c.

a good ohmic contact for doping levels exceeding $10^{16}/cm^3$. However, a problem arises in trying to contact *n*-type silicon, as shown in Fig. 7.3(b). For lightly doped *n*-type material, aluminum can form a metal-semiconductor "Schottky-barrier" diode rather than an ohmic contact. To prevent this rectifying contact from forming, an n^+ diffusion is placed between the aluminum and any lightly doped *n*-type regions, as in Fig. 7.3(c). The resulting contact has an *I–V* characteristic similar to that in Fig. 7.2(c). This technique was used in forming the collector contact in the bipolar process shown in Fig. 1.6.

7.2.2 Aluminum–Silicon Eutectic Behavior

Silicon melts at a temperature of 1412 °C, and pure aluminum melts at 660 °C. However, aluminum and silicon together exhibit "eutectic" characteristics in which a mixture of the two materials lowers the melting point of the composite material to below that of either element. Figure 7.4 shows the phase diagram of the aluminum–silicon system at a pressure of 1 atm. The minimum melting temperature, or *eutectic temperature*, is 577 °C and corresponds to an 88.7% Al, 11.3% Si composition. Because of the relatively low eutectic

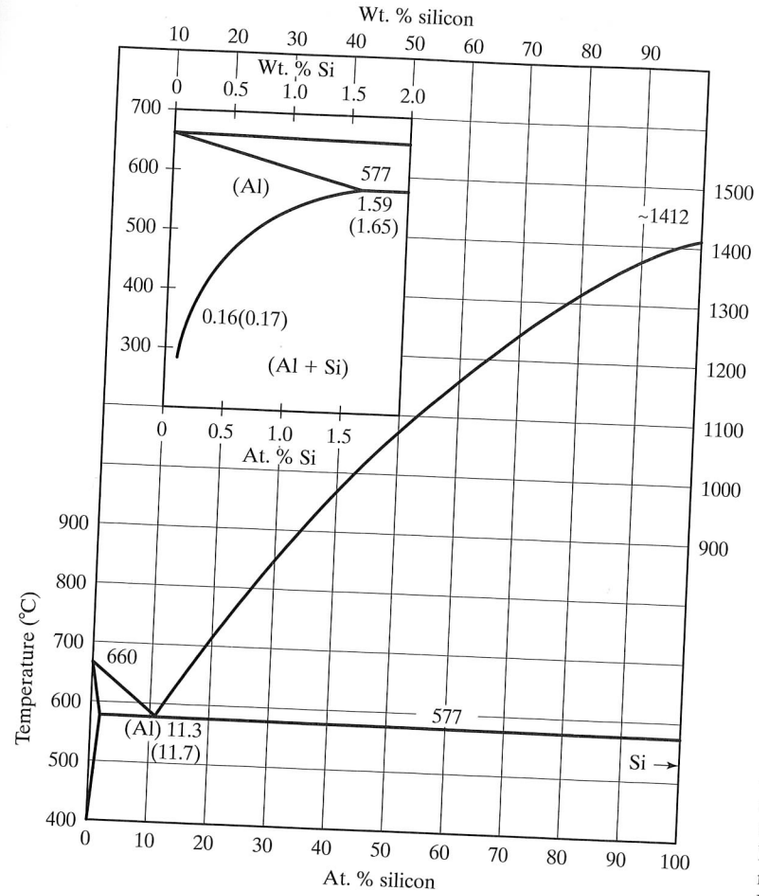


FIGURE 7.4

Phase diagram of the aluminum–silicon system. The silicon–aluminum eutectic point occurs at a temperature of 577 °C. At contact-alloying temperatures between 450 and 500 °C, aluminum will absorb from 0.5 to 1% silicon. Copyright 1958 McGraw-Hill Book Company, reprinted with permission from Ref. [1].

temperature of the Al–Si system, aluminum must be introduced into the IC process sequence after all high-temperature processing has been completed.

7.2.3 Aluminum Spiking and Junction Penetration

To ensure good contact formation, aluminum is normally annealed in an inert atmosphere at a temperature of 450 to 500 °C following deposition and patterning. Although this temperature is well below the eutectic temperature for silicon and aluminum, silicon still diffuses into the aluminum. The diffusion leads to a major problem associated with the formation of aluminum contacts to silicon, particularly for shallow junctions.

Anywhere a contact is made between aluminum and silicon, silicon will be absorbed by the aluminum during the annealing process. The amount of silicon absorbed will depend on the time and temperature involved in the annealing process, as well as the area of the contact. (See Problem 7.4.) To make matters worse, the silicon is not absorbed uniformly from the contact region. Instead, it tends to be supplied from a few points. As

the silicon is dissolved, spikes of aluminum form and penetrate the silicon contact region. If the contact is to a shallow junction, the spike may cause a junction short, as in Fig. 7.5.

The inset in Fig. 7.4 gives the solubility of silicon in aluminum. Between 400 °C and the eutectic temperature, the solubility of silicon in aluminum ranges from 0.25 to 1.5% by weight. To solve the spiking problem, silicon may be added to the aluminum film during deposition by coevaporation from two targets, or sputter deposition can be used with an aluminum target containing approximately 1% silicon. Both of these techniques deposit a layer in which the aluminum demand for silicon is satisfied, and the metallization does not absorb silicon from the substrate during subsequent annealing steps.

The junction penetration problem becomes particularly acute in high-density VLSI processes with extremely shallow junctions; another way to prevent spiking is to place a barrier material between the aluminum and silicon, as shown in Fig. 7.5. One possibility is to deposit a thin layer of polysilicon prior to aluminum deposition. The polysilicon will then supply the silicon needed to saturate the aluminum. Another alternative is to use a metal as a barrier. The metal must form a low-resistance contact with silicon, not react with aluminum, and be compatible with other process steps. Various semiconductor manufacturers have used a number of metals, including platinum, palladium, titanium, and tungsten.

7.2.4 Contact Resistance

There is a small resistance associated with an ohmic contact between two materials. To a first approximation, the *contact resistance* R_c is inversely proportional to the area of the contact:

$$R_c = \rho_c / A \tag{7.1}$$

where ρ_c is the specific contact resistivity in ohm-cm^2 and A is the area of the contact. For example, a 2×2 μm contact with $\rho_c = 1$ μohm-cm^2 yields a contact resistance of 25 ohms. Figure 7.6 shows the contact resistivity as a function of annealing temperature for several aluminum–silicon systems. It is evident why the 450 °C annealing process is used following aluminum deposition. Also note that the use of polysilicon under aluminum to prevent junction spiking yields a much poorer value of ρ_c.

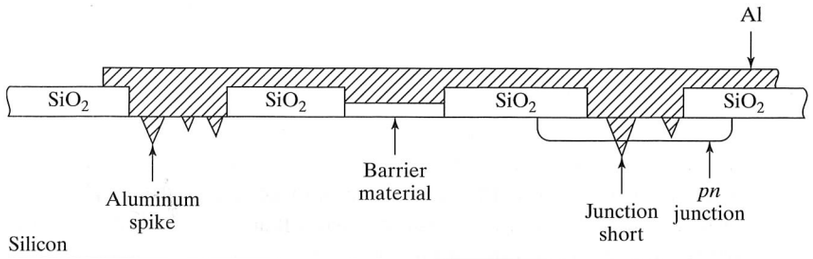


FIGURE 7.5

Aluminum spiking which occurs during aluminum–silicon alloying. Aluminum spikes can cause shorts in shallow junctions. Aluminum containing 1% silicon is often used to eliminate spiking. A barrier material of polysilicon or a metal such as titanium can also be used to prevent spiking.

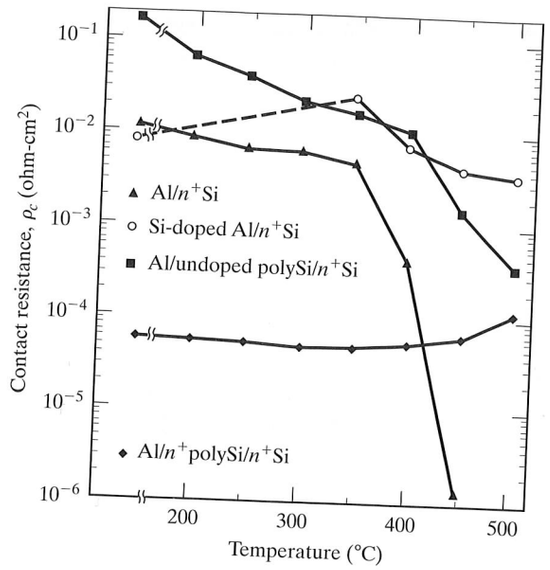


FIGURE 7.6

Contact resistivity of a variety of aluminum–silicon systems. An alloying temperature of 450 °C is typically used to obtain low-contact resistance for Al–Si contacts. Reprinted with permission from *Solid-State Electronics*, Vol. 23, pp. 255-262, M. Finetti et al., "Aluminum–Silicon Ohmic Contact on Shallow n$^+$/p Junctions" [2]. Copyright 1980, Pergamon Press, Ltd.

7.2.5 Electromigration

Metal interconnections in integrated circuits are operated at relatively high current densities, and a very interesting failure mechanism develops in aluminum and other conductors. *Electromigration* is the movement of atoms in a metal film due to momentum transfer from the electrons carrying the current. Under high-current-density conditions, metal–atom movement causes voids in some regions and metal pileup, or *hillocks*, in other regions, as shown in Fig. 7.7. Voids can eventually result in open circuits, and pileup can cause short circuits between closely spaced conductors.

The mean time to failure (MTF) of a conductor due to electromigration has been experimentally related to current density, J, and temperature by

$$\text{MTF} \propto (J^{-2}) \exp(E_A / kT) \tag{7.2}$$

where E_A is an activation energy with a typical value of 0.4 to 0.5 eV for aluminum.

The most common method of improving aluminum resistance to electromigration is to add a small percentage of a heavier metal such as copper. Targets composed of 95% Al, 4% Cu, and 1% Si are routinely used in sputter deposition systems. The aluminum–copper–silicon alloy films simultaneously provide electromigration resistance and eliminate aluminum spiking.

Pure copper interconections would be expected to exhibit a much higher electomigration resistance than aluminum, which is in fact the case, as shown in the results in Fig. 7.8, which compares electromigration performance of TiN clad copper lines with those formed of an AlCu alloy [6]. An order of magnitude in improvement is obtained.

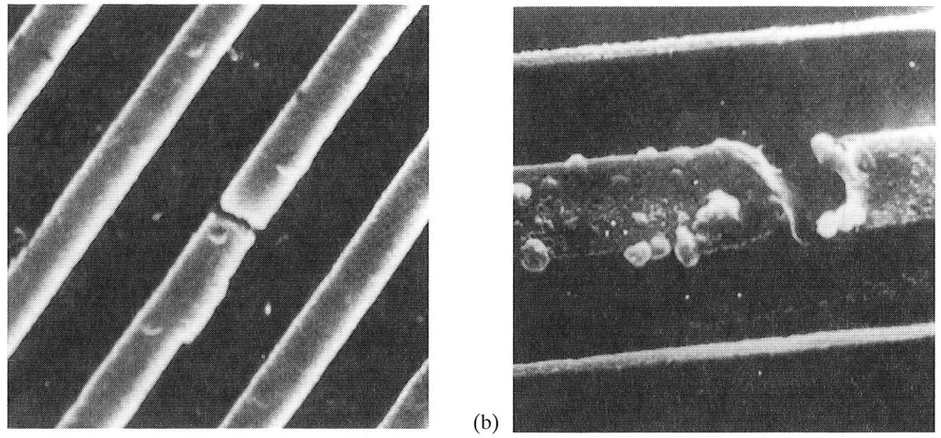


FIGURE 7.7

Scanning electron micrographs of aluminum interconnection failure caused by electromigration. (a) Sputtered aluminum with 0.5% copper; (b) evaporated aluminum with 0.5% copper. Copyright 1980, IEEE. Reprinted with permission from Ref. [3].

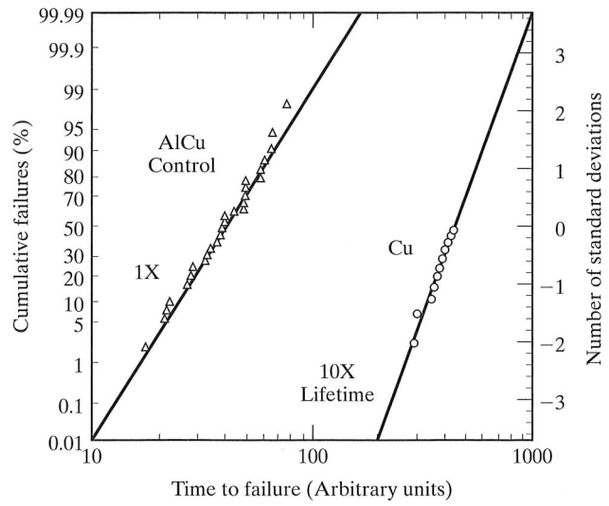


FIGURE 7.8

Electromigration performance improvement using copper metallization. Copyright 1997, IEEE. Reprinted with permission from Ref. [6].

7.3 DIFFUSED INTERCONNECTIONS

Diffused conductors with low sheet resistances represent the second available interconnect medium in basic IC technology. From Fig. 4.16, we can see that the minimum resistivity is approximately 1,000 μohm-cm. For shallow structures measuring about 1 μm, the minimum obtainable sheet resistance is typically between 10 and 20 ohms per square. Such sheet resistances are obviously much higher than that of metal, and one must be selective in the use of diffusions for signal or power distribution.

The diffused line must really be modeled as a distributed *RC* structure, as illustrated in Fig. 7.9, when signal propagation is considered. The resistance, *R*, of diffused regions was discussed in detail in Chapter 4, and *C* represents capacitance of the

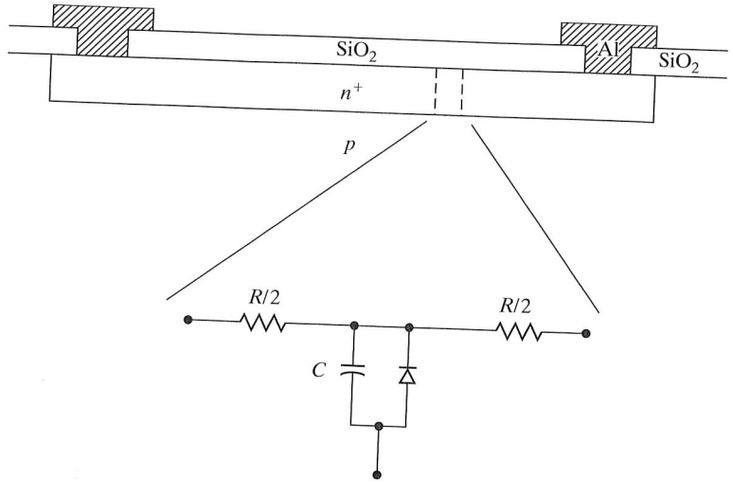


FIGURE 7.9

A lumped circuit model for a small section of an n^+ diffusion. The *RC* line delay limits the use of diffusions for high-speed signal distribution.

reverse-biased *pn* junction formed between the diffused region and the substrate. Heavily doped diffusions are normally used for interconnection purposes and can be approximated by a one-sided step junction in which the depletion layer extends predominantly into the substrate. The capacitance per unit area is given by

$$C = \sqrt{\frac{qN_sK_s\epsilon_o}{2(\phi_{bi} + V_R)}} \qquad \phi_{bi} = (kT/q)\ln\left(\frac{N_s}{n_i}\right) + 0.56\text{ V} \tag{7.3}$$

where N_s is the substrate doping, φ_{bi} is the built-in potential of the junction, and V_R is the reverse bias applied to the junction.

The relatively large *RC* product of long diffused lines results in substantial time delay for signals propagating down such a line. Hence, diffusions are more useful in interconnecting adjacent devices in integrated circuits. Figure 7.10 shows a three-input NMOS NOR-gate in which the source diffusions of the three input transistors are merged together as one diffusion. The three drains of the input devices, as well as the source of the depletion-mode load device, are also merged together as one diffusion. Figure 7.1 shows an example of the use of long diffused interconnection regions in a programmable-logic-array (PLA) structure.

7.4 POLYSILICON INTERCONNECTIONS AND BURIED CONTACTS

Heavily doped *n*-type polysilicon is the primary MOS transistor gate material in use today, and it provides an additional layer of interconnection that is easily insulated from other layers by thermal oxidation or insulator deposition. This extra level of interconnection greatly facilitates the layout of compact digital integrated circuits. Thin, heavily doped polysilicon layers have a minimum resistivity of approximately 300 μohm-cm, and they suffer from the same sheet-resistance problems associated with shallow diffused interconnections (typically 20 to 30 ohms per square). Polysilicon lines have substantial capacitance to the substrate and exhibit *RC* delay problems similar to those of diffused interconnections.

7.4.1 Buried Contacts

In the polysilicon-gate processes presented thus far, the polysilicon acts as a barrier material during ion implantation or diffusion. Thus, a diffusion can never pass beneath a polysilicon line. In addition, contact windows to the diffusions are not opened until after polysilicon deposition. It is therefore necessary to use a metal link to connect between polysilicon and diffusion, as in Fig. 7.11(a). Interconnecting the diffusion to polysilicon in this manner requires two contact windows and an intervening space, both of which are wasteful of area.

In memory arrays, where density is extremely important, an extra mask step can be introduced into the process to permit direct contact between polysilicon and silicon, as shown in Fig. 7.11(b). Prior to polysilicon deposition, windows are opened in the thin gate oxide, permitting the polysilicon to contact the silicon surface. Diffusion of the *n*-type dopant from the heavily doped n^+ polysilicon merges with the adjacent ion-implanted n^+ regions, and the result is called a *buried contact*. The edge of the contact exhibits the lowest resistance since the impurity concentration is greatest in that region.

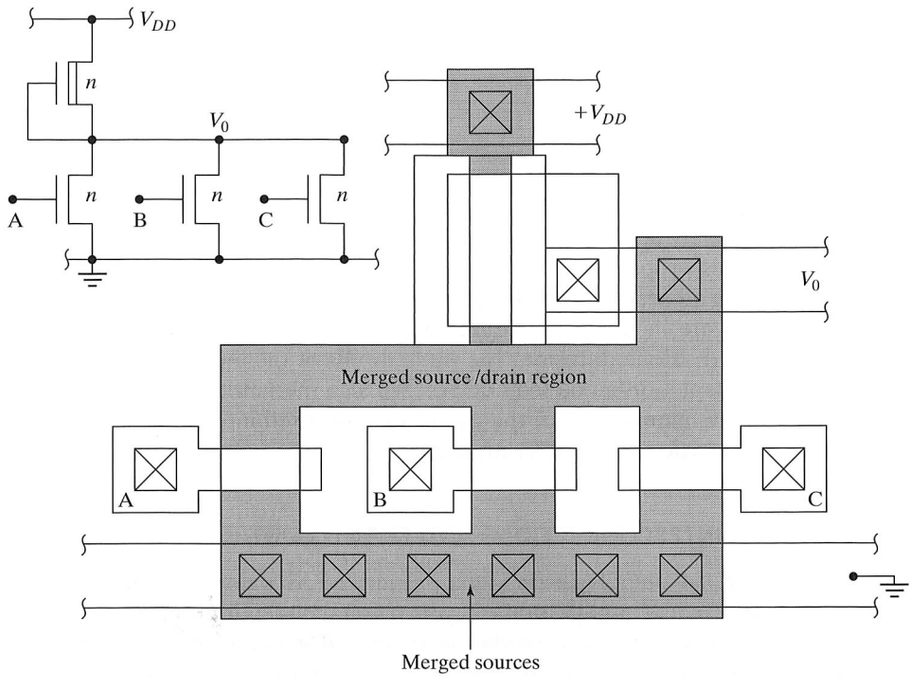


FIGURE 7.10

Layout of a three-input NMOS NOR-gate showing device interconnection through merging of adjacent source and drain diffusions.

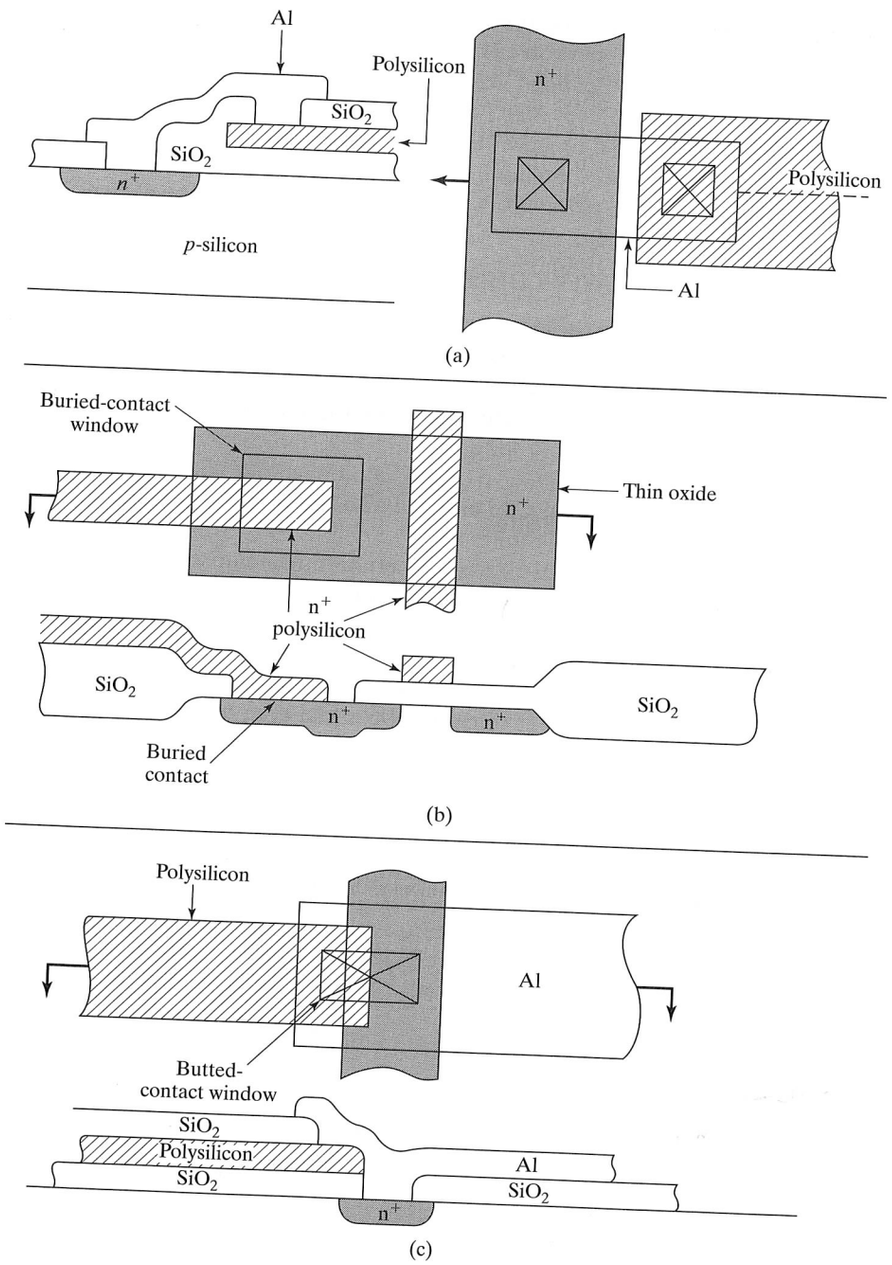


FIGURE 7.11

Three techniques for interconnecting polysilicon and n^+ diffusion. (a) Normal aluminum link requiring two contact regions and an intervening space; (b) buried-contact structure; (c) butted-contact structure.

7.4.2 Butted Contacts

Another method of conserving area is to form a "butted" contact as shown in Fig. 7.11(c). In this example, polysilicon is aligned with the edge of the diffusion contact window, and metal connects the diffusion and polysilicon together. The butted contact saves area by eliminating the space normally required between separate contact windows.

7.5 SILICIDES AND MULTILAYER-CONTACT TECHNOLOGY

The sheet resistance of both thin polysilicon and shallow diffusions cannot be reduced below 10 to 20 ohms per square, which greatly reduces their utility as an interconnection medium. Interconnect delays limit the speed of VLSI circuits, and as dice get larger and feature sizes get smaller, methods for improving these interconnections have had to be found.

7.5.1 Silicides, Polycides, and Salicides

A wide range of noble and refractory metals form compounds with silicon called *silicides*, and the sheet resistance of polysilicon and diffusion can be reduced by forming a low-resistivity, shunting silicide layer on their surfaces. A list of properties of possible silicides is given in Table 7.2. Several of the elements, including titanium, tungsten, platinum, and palladium, have been used in the formation of Schottky-barrier diodes in bipolar processes since the 1960s and are now used to form silicides for interconnection purposes.

TABLE 7.2 Properties of Some Silicides of Interest. Reprinted with permission of the American Institute of Physics from Ref [4].

Silicide	Starting Form	Sintering Temperature (°C)	Lowest Binary Eutectic Temperature (°C)	Specific Resistivity (μohm-cm)
$CoSi_2$	Metal on polysilicon	900	1195	18–25
	Cosputtered alloy	900		
$HfSi_2$	Metal on polysilicon	900	1300	45–50
$MoSi_2$	Cosputtered alloy	1000	1410	100
$NiSi_2$	Metal on polysilicon	900	966	50
	Cosputtered alloy	900		50–60
Pd_2Si	Metal on polysilicon	400	720	30–50
PtSi	Metal on polysilicon	600–800	830	28–35
$TaSi_2$	Metal on polysilicon	1000	1385	35–45
	Cosputtered alloy	1000		50–55
$TiSi_2$	Metal on polysilicon	900	1330	13–16
	Cosputtered alloy	900		25
WSi_2	Cosputtered alloy	1000	1440	70
$ZrSi_2$	Metal on polysilicon	900	1355	35–40

A structure with a silicide formed on top of the polysilicon gate, often called a *polycide*, is shown in Fig. 7.12. A layer of the desired metal is deposited using evaporation, sputtering, or CVD techniques. Upon heating of the structure to a temperature between 600 and 1000 °C, the metal reacts with the polysilicon to form the desired silicide. Coevaporation, cosputtering, or sputtering of a composite target may be used to simultaneously deposit both silicon and metal onto the polysilicon surface prior to the thermal treatment or "sintering" step. Silicides have resistivities in the range of 15 to 50 μohm-cm.

Another feature of silicide layers is the ability to oxidize the surface following silicide formation. At high temperatures, silicon diffuses readily through the silicide layer and will combine with oxygen at the silicide surface to form an SiO_2 insulating layer.

The eutectic temperature of the silicide and silicon will limit the temperature of further processing steps, as in the case of aluminum. However, many silicides are stable at temperatures exceeding 1000 °C. Exceptions include the silicides of nickel (900 °C), platinum (800 °C), and palladium (700 °C).

Silicides are also used to reduce the effective sheet resistance of diffused interconnections. Figure 7.13 outlines a process for simultaneous formation of silicides on both the gate and source–drain regions of an MOS transistor. An oxide spacer is used to prevent silicide formation on the side of the gate, because such formation could cause a short between the gate and diffusions. The spacer is formed by first coating the surface with a CVD oxide, followed by a reactive-ion etching step. The oxide along the edge of the gate is thicker than over other regions, and some oxide is left on the side of the gate at the point when the oxide is completely removed from the source and drain regions and the top of the gate. Next, metal is deposited over the wafer. During sintering, silicide forms only in the regions where metal touches silicon or polysilicon. Unreacted metal may be removed with a selective etch that does not attack the silicide. The result is a silicide that is automatically self-aligned to the gate and source–drain regions. *Self-aligned silicides* are often called *salicides*.

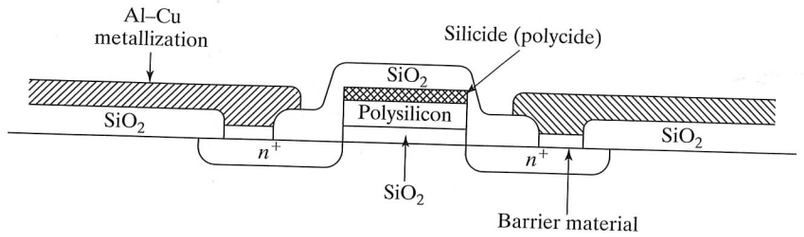


FIGURE 7.12

MOS structure showing the use of a "polycide" to reduce the sheet resistance of the polysilicon gate material and a barrier material to prevent aluminum spiking through shallow source–drain junctions.

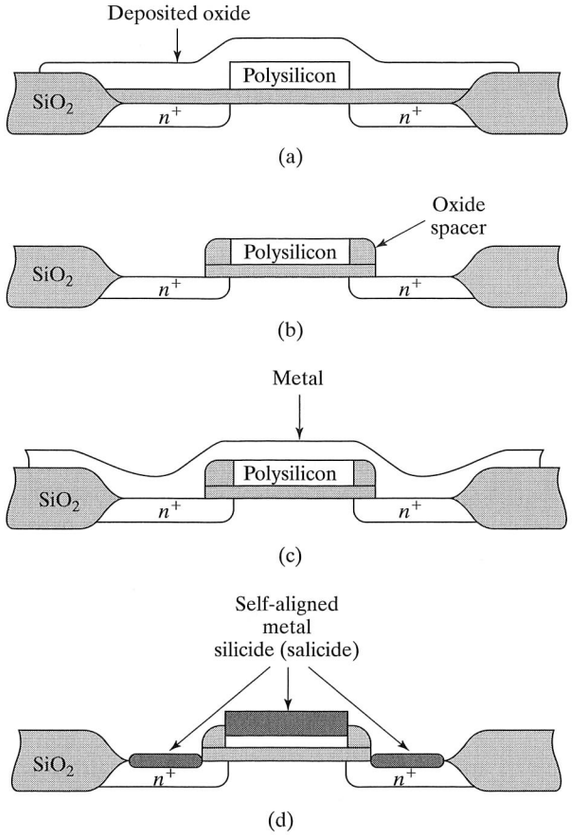


FIGURE 7.13

Use of self-aligned silicide ("salicide") in the formation of an MOS device. (a) Oxide is deposited over the normal MOS structure following polysilicon definition; (b) structure after reactive-ion etching leaving a sidewall oxide spacer; (c) metal is deposited over the structure and heated to form silicides; (d) unreacted metal is readily etched away, leaving silicide automatically aligned to gate and source–drain regions.

7.5.2 Barrier Metals and Multilayer Contacts

Aluminum contacts to silicides suffer from the same pitting and spiking problems associated with direct contact to silicon. To circumvent these problems, an intermediate layer of metal is used that prevents silicon diffusion. Figure 7.14 shows the application of titanium–tungsten (TiW) as a barrier metal over the silicides in the contact regions of both bipolar and MOS technologies. The final contact consists of a sandwich of a silicide over the diffusion, followed by the TiW diffusion barrier, and completed with aluminum–copper interconnection metallization. Multilayer contact structures are common in advanced, high-performance MOS, and bipolar technologies.

7.6 THE LIFTOFF PROCESS

The pattern definition processes which have been discussed previously have been "subtractive" processes, as illustrated in Fig. 7.15(a). The wafer is completely covered with a thin film layer, which is selectively protected with a masking layer such as photoresist. Wet or dry etching then removes the thin film material from the unprotected areas.

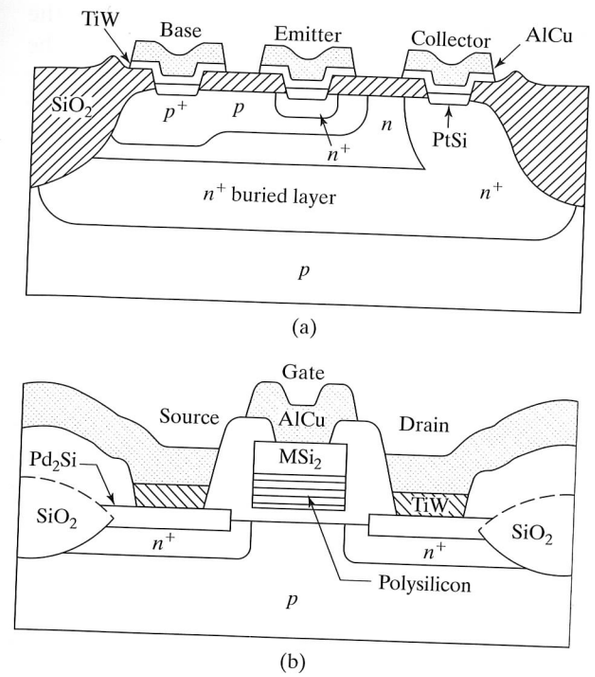


FIGURE 7.14

Device cross sections showing the use of silicide contacts in (a) bipolar and (b) MOS devices. Reprinted with permission from *Semiconductor International* magazine, August 1985[5]. Copyright 1985 by Cahners Publishing Co., Des Plaines, IL.

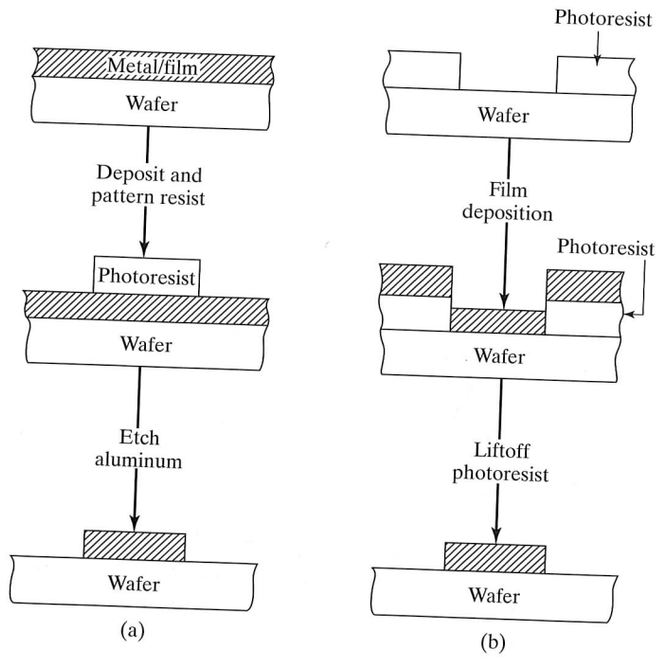


FIGURE 7.15

A comparison of interconnection formation by (a) subtractive etching and (b) additive metal liftoff.

The additive or *liftoff* process shown in Fig. 7.15(b) can also be used, in which the substrate is first covered with a photoresist layer patterned with openings where the final material is to appear. The thin film layer is deposited over the surface of the wafer. Any material deposited on top of the photoresist layer will be removed with the resist, leaving the patterned material on the substrate. For liftoff to work properly, there must be a very thin region or a gap between the upper and lower films. Otherwise, tearing and incomplete liftoff will occur.

The masking patterns for the liftoff and subtractive processes are the negatives of each other. This can be achieved by changing the mask from dark field to light field or by changing from negative to positive photoresist.

7.7 MULTILEVEL METALLIZATION

A single level of metal simply does not provide sufficient capability to fully interconnect complex VLSI chips. Many processes now use two or three levels of polysilicon, as well as several levels of metallization, in order to ensure wirability and provide adequate power distribution.

7.7.1 Basic Multilevel Metallization

A multilevel metal system is shown in Fig. 7.16. Standard processing is used through the deposition and patterning of the first level of metal. An interlevel dielectric, consisting

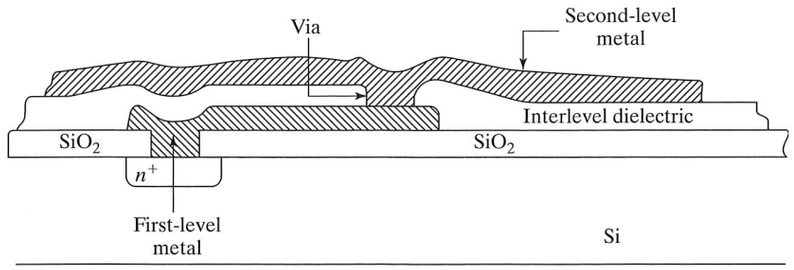


FIGURE 7.16

(a) Basic two-level metallization process may use polyimide, oxide, or nitride as an interlevel dielectric; (b) additional process steps may be added to fill the vias with metal prior to each metal deposition in order to achieve a more planar structure.

of CVD or sputtered SiO_2, or a plastic-like material called *polyimide*, is then deposited over the first metal layer. The dielectric layer must provide good step coverage and should help smooth the topology. In addition, the layer must be free of pinholes and be a good insulator. Next, vias are opened in the dielectric layer, and the second level of metallization is deposited and patterned.

7.7.2 Planarized Metallization

The topology that results from the simple multilayer interconnect process of Fig. 7.16(a) simply cannot be utilized in submicron processeses because of the depth-of-field limitations in the lithographic processes. The CMP process introduced in Chapter 3 is used to achieve highly planar layers. In the process flow in Fig. 7.16(b), a via filling technique is used to form the vias between metal layers. Tungsten is commonly used as the via metallization. The dielectric deposition, metallization, and CMP processes are repeated until the desired number of levels of interconnection is achieved. Integrated circuits with six levels of metal have been successfully fabricated using similar processes. An example of a planarized multilevel metal system employing aluminum metallization and tungsten "plugs" appears in Fig. 7.17 [7].

7.7.3 Low Dielectric Constant Interlevel Dielelctrics

Propagation delay associated with interconnections is a critical issue in high-performance microprocessors, as well as other integrated circuits. The *RC* product associated with these interconnections can be decreased by reducing either the resistance or capacitance or both. Copper is being used to reduce the resistance term and is described in

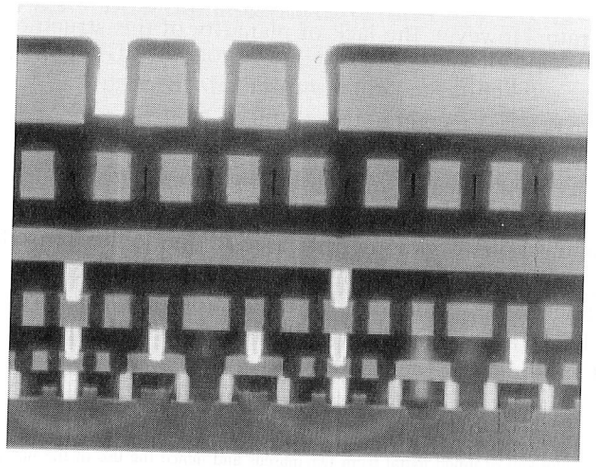


FIGURE 7.17

Multilevel aluminum metallization with tungsten plugs. Copyright 1998 IEEE. Reprinted with permission from Ref. [7].

detail in Section 7.8. Silicon dioxide, the most common interlevel dielectric, has a relative dielectric constant of 3.9. This value is fairly high, although much less than that of silicon itself ($e_r = 11.7$). Air isolated interconnects, with $e_r = 1$, have been utilized in GaAs circuits, but have not been successfully applied to silicon integrated circuits. Research teams are presently trying to identify dielectric materials that are compatible with silicon IC technology and have e_r values in the 2.0–2.5 range. Fluorinated oxides, porous oxides, and many polymer materials are under investigation [11–13].

7.8 COPPER INTERCONNECTS AND DAMASCENE PROCESSES

Because of its lower resistivity (see Table 7.1), copper is being used in place of other metals in multilevel metal systems. Unfortunately, copper is a deep-level impurity and a very rapid diffuser in silicon (see Fig. 4.5), and so great care must be exercised to prevent it from contaminating the silicon substrate and devices. The metallization techniques discussed so far have been subtractive processes in which the metal is deposited everywhere and then etched away where not desired. Manufacturable dry etching processes have not been developed for the removal of copper, so additive plating techniques are used. The Damascene processes use chemical mechanical polishing, as discussed in Chapter 3, to produce highly planar layers that may be used for multiple layers of interconnect.

7.8.1 Electroplated Copper Interconnect

Two methods of electroplating copper interconnect lines are shown in Fig. 7.18 [8]. The first involves plating through a mask. In Fig. 7.18(a), a conductive seed layer must first be deposited on the wafer that may already be planarized utilizing a CMP step. The seed layer provides an electrical path that is needed for current during the plating process. A masking layer such as photoresist is then deposited on the wafer and lithographically patterned. The wafer is immersed in the plating system, and a dc bias is applied between the solution and the seed layer. Copper plating occurs wherever the seed layer is exposed to the plating solution. Following the plating operation, the masking layer is removed and the seed layer etched away, leaving a copper interconnection line on top of the substrate. However, the lack of planarity of this structure limits its application in today's ICs.

7.8.2 Damascene Plating

The Damascene process of Fig. 7.18(b) is ideally suited to IC interconnect structures as pointed out in [8]. An insulating layer such as silicon dioxide is deposited on the surface of the substrate, and standard photolithography is used to define the desired interconnect pattern. Following seed-layer deposition, the entire surface is electroplated, filling the interconnect regions as well as covering the rest of the surface with copper. The excess copper is polished away using a CMP process step. The final structure is highly planar with metal lines inlaid in the insulator.[1]

[1]The metal interconnect structures produced by Damascene processing are embedded or inlaid in the insulating background material. The inlaid structures are reminiscent of inlaid metal from Daemascus and hence the use of the name Damascene processes.

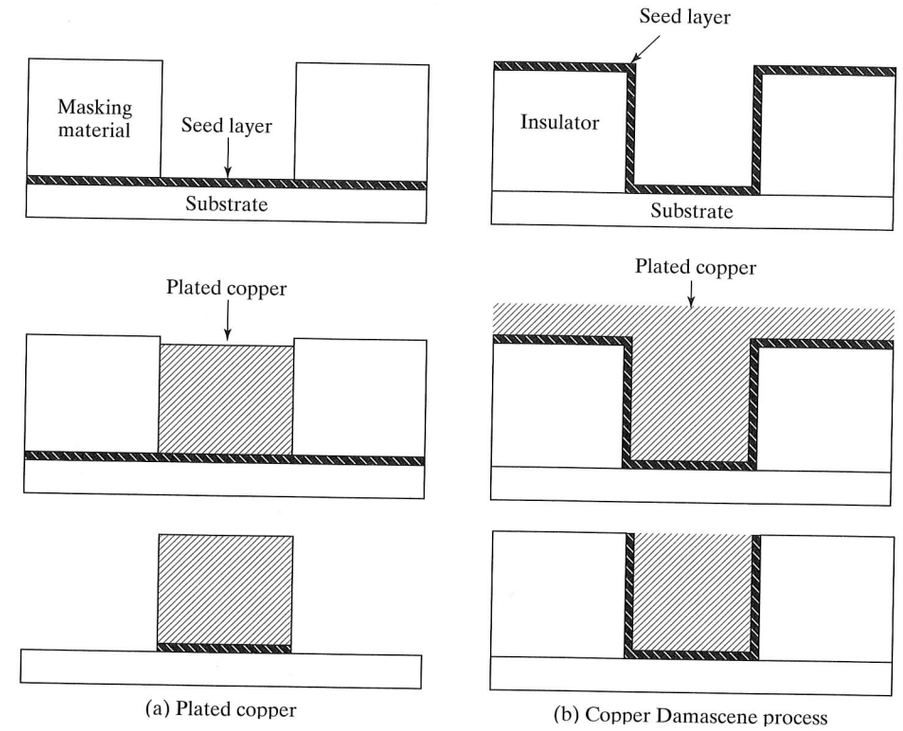


FIGURE 7.18

Plated copper and copper Damescence process steps. (a) Mask openings are defined over seed layer and copper is plated in the opening. A nonplanar structure results after plating mask removal. (b) A seed layer is deposited over the patterned insulator, and copper is plated over the entire structure. A planar surface results after the excess copper is lapped away.

7.8.3 Dual Damescence Structures

The full power of the Damescence technique is realized through the dual Damescence process, which forms interconnection lines and vias between interconnect levels at the same time, as illustrated in Fig. 7.19. The process begins with the substrate coated with a thin etch stop layer such as silicon nitride. Two layers of an insulator such as SiO_2 are deposited with a thin intervening etch stop layer. The insulator sandwich is capped with a final etch stop layer.

Windows are opened in the silicon nitride layer defining the via locations, and the insulator is etched away with the etch terminating on the silicon nitride layer. A new set of windows defining the interconnection lines are etched through the nitride. The nitride etch stop is also removed from the bottom of the via. The oxide is then etched simultaneously from the upper and lower levels of oxide. A barrier layer such as titanium nitride (TiN) may be deposited before the seed layer deposition to prevent interaction between the plated copper layer and the insulator. Electroplated copper then builds up the vias and the interconnect lines simultaneously. The structure is completed with CMP removal of the excess copper. The process can be

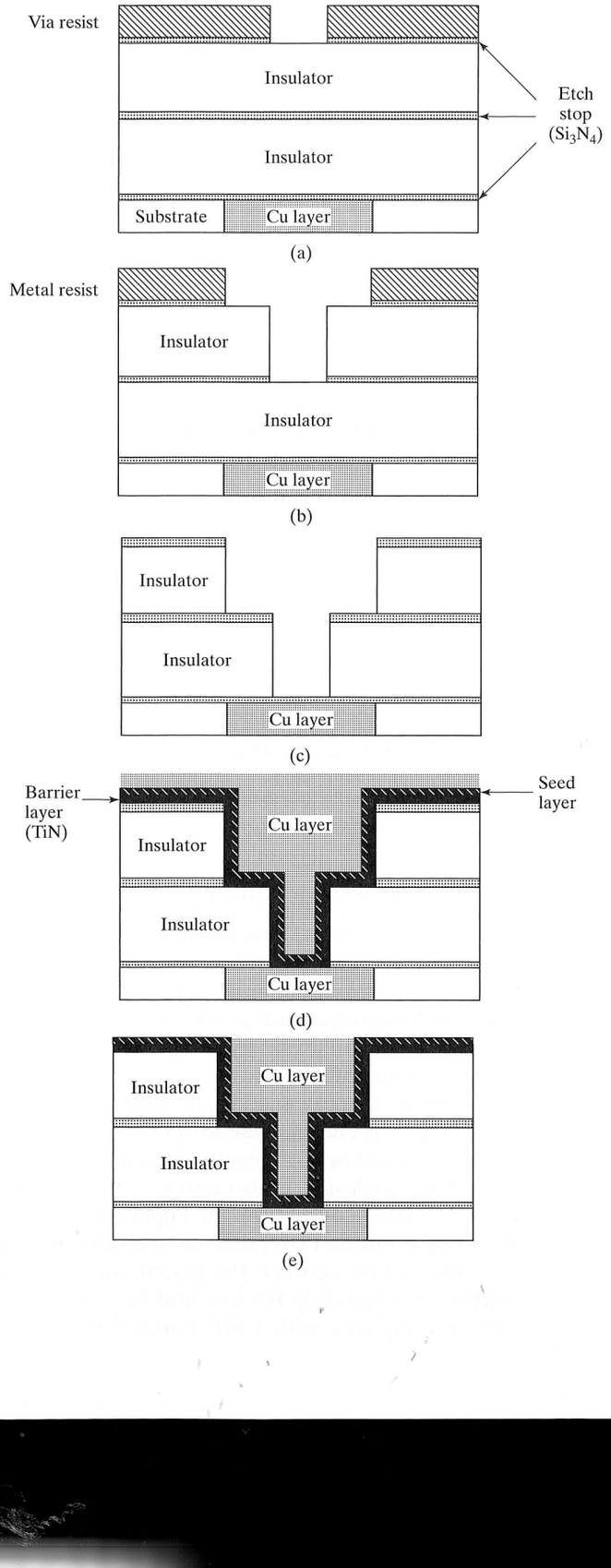
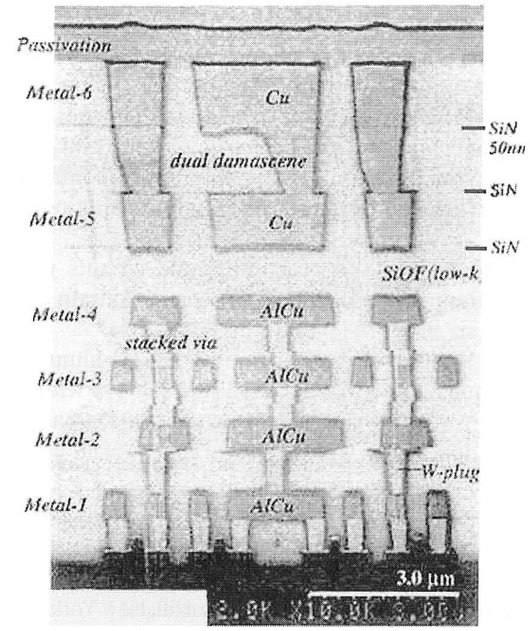


FIGURE 7.19

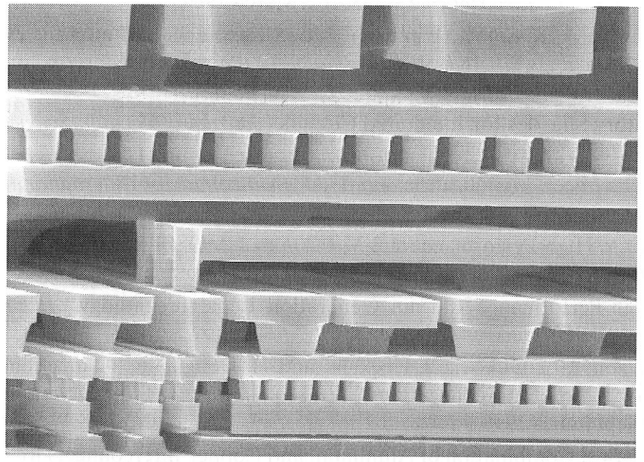
Dual Damascene process flow. (a) An insulator sandwich is first deposited and the upper nitride layer is patterned. The insulator layer is etched. The etch terminates on the silicon nitride etch stop. (b) The nitride layer is patterned and etched. (c) Following the next oxide etch step, two different width openings exist in the two oxide layers. (d) Barrier and seed layers are deposited and plated with copper. (e) Final structure following removal of excess copper.

repeated to build up additional layers of interconnect. The resulting copper interconnects are surrounded by a thin cladding layer of TiN.

Figure 7.20 shows two multilevel metal systems involving dual Damescence processing. The first [6], Fig. 7.20(a), uses four layers of aluminum copper interconnections with tungsten via plugs plus two levels of Damescence copper interconnections. The second [9], Fig. 7.20(b), shows the use of six levels of copper wiring.



(a)



(b)

FIGURE 7.20

Microphotographs of six-level metallization.
(a) Dual Damescene copper combined with aluminum–copper and tungsten plugs on the lower levels. (b) Dual Damescene copper. Note planarity of both structures. Copyright 1997 and 1998 IEEE. Reprinted with permission from Refs. [6] & [9].

SUMMARY

In this chapter, we have explored the various types of interconnections used in modern integrated circuits, including diffusion, polysilicon, and metal. Diffusion and polysilicon have a relatively high sheet resistance, which often restricts their use to local interconnections. The formation of metal silicides on the surface of polysilicon lines and diffusions can substantially reduce the sheet resistance of these interconnections.

Problems relating to the formation of good ohmic contacts between aluminum and silicon have also been discussed. An n^+ layer is required between aluminum and *n*-type silicon to prevent formation of a Schottky-barrier diode instead of an ohmic contact. Aluminum penetration into silicon is a serious problem in forming contacts to shallow junctions. Metals such as tungsten and titanium are often used as silicon diffusion barriers to prevent aluminum penetration into contacts to silicon or silicides.

At high current densities, a failure mechanism called *electromigration* can cause open and short circuits to form in the metallization layers. Aluminum containing approximately 1% silicon and 4% copper is used to minimize aluminum spiking and electromigration, respectively.

Multilevel metal processes have been developed for integrated circuits which require more than one level of metallization. Some of today's processes contain up to three levels of polysilicon, and others use six or more levels of metallization

The first successful approaches to multilevel metallization covered aluminum with layers of planarizing oxide or polyimidel. However the topology achieve with this approach has proven to be too rugged for deep submicron processes. Damascene processes combine electroplated copper deposition with chemical mechanical polishing to achieve highly planar multilevel metallization.

REFERENCES

[1] M. Hansen and A. Anderko, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.

[2] M. Finetti, P. Ostoja, S. Solmi, and G. Soncini, "Aluminum–Silicon Ohmic Contact on 'Shallow' n^+/p Junctions," *Solid-State Electronics*, *23*, 255–262, March 1980.

[3] S. Vaidya, D. B. Fraser, and A. K. Sinha, "Electromigration Resistance of Fine Line Al for VLSI Applications," *Proceedings of 18th IEEE Reliability Physics Symposium*, pp. 165–167, 1980.

[4] S. P. Murarka, "Refractory Silicides for Integrated Circuits," *Journal of Vacuum Science and Technology*, *17*, 775–792, July/August 1980.

[5] P. S. Ho, "VLSI Interconnect Metallization—Part 3," *Semiconductor International*, 128–133, August 1985.

[6] S. Venkatesen et al., "A High-Performance 1.8-V, 0.20-μm CMOS Technology with Copper Metallization," *IEEE IEDM Technical Digest*, pp. 769–772, December 1997.

[7] S. Yang et al., "A High-Performance 180 nm Generation Logic Technology," *IEEE IEDM Technical Digest*, pp. 197–200, December 1998.

[8] P. C. Andricacos, U. Uzoh, J. O. Dukovic, J. Horkans, and H. Deligianni, "Damascene Copper Electroplating for Chip Interconnections," *IBM Journal of Research and Development*, pp. 567–574, September 1998.

[9] M. Igarashi et al., "The Best Combination of Aluminum and Copper Interconnects for a High-Performance 0.18 μm CMOS Logic Device," *IEEE IEDM Technical Digest*, pp. 829–832, December 1998.

[10] D. Edelstein et al., "Full Copper Wiring in a sub–0.25 μm CMOS ULSI Technology," *IEEE IEDM Technical Digest*, pp. 773–776, December 1997.

[11] Y. Matsubara et al., "Low-k Fluorinated Amorphous Carbon Interlayer Technology for Quarter Micron Devices," *IEEE IEDM Technical Digest*, pp. 369–372, December 1996.

[12] E. M. Zielinski et al., "Damascene Integration of Copper and Ultra-low-k Xerogel for High-Performance Interconnects," *IEEE IEDM Technical Digest*, pp. 936–938, December 1997.

[13] Y. Matsubara et al., "RC Delay Reduction of 0.18 μm CMOS Technology Using Low Dielectric Constant Fluorinated Amorphous Carbon," *IEEE IEDM Technical Digest*, pp. 841–844, December 1998.

FURTHER READING

[1] P. B. Ghate, J. C. Blair, and C. R. Fuller, "Metallization in Microelectronics," *Thin Solid Films*, *45*, 69–84, August 15, 1977.

[2] G. L. Schnable and R.S. Keen, "Aluminum Metallization—Advantages and Limitations for Integrated Circuit Applications," *Proceedings of the IEEE*, *57*, 1570–1580, September 1969.

[3] J. Black, "Physics of Electromigration," *Proceedings of the 12th IEEE Reliability Physics Symposium*, p. 142, 1974.

[4] C. Y. Ting, "Silicide for Contacts and Interconnects," *IEEE IEDM Technical Digest*, pp. 110–113, December 1984.

[5] S. P. Murarka, "Recent Advances in Silicide Technology," *Solid-State Technology*, *28*, 181–185, September 1985.

[6] S. Sachdev and R. Castellano, "CVD Tungsten and Tungsten Silicide for VLSI Applications," *Semiconductor International*, pp. 306–310, May 1985.

[7] P. B. Ghate, J. C. Blair, C. R. Fuller, and G. E. McGuire, "Application of Ti:W Barrier Metallization for Integrated Circuits," *Thin Solid Films*, *53*, 117–128, September 1, 1978.

[8] R. A. M. Wolters and A. J. M. Nellissen, "Properties of Reactive Sputtered TiW," *Solid-State Technology, 29*, 131–136, February 1986.

[9] T. Sakurai and T. Serikawa, "Lift-Off Metallization of Sputtered Al Alloy Films," *Journal of the Electrochemical Society, 126*, 1257–1260, July 1979.

[10] T. Batchelder, "A Simple Metal Lift-Off Process," *Solid-State Technology, 25*, 111–114, February 1982.

[11] S. A. Evans, S. A. Morris, L. A. Arledge, Jr., J. O. Englade, and C. R. Fuller, "A 1-μm Bipolar VLSI Technology," *IEEE Transactions on Electron Devices, ED-27*, 1373–1379, August 1980.

[12] Y. Sasaki, O. Ozawa, and S. Kameyama, "Application of $MoSi_2$ to the Double-Level Interconnections of I^2L Circuits," *IEEE Transactions on Electron Devices, ED-27*, 1385–1389, August 1980.

[13] R. A. Larsen, "A Silicon and Aluminum Dynamic Memory Technology," *IBM Journal of Research and Development, 24*, 268–282, May 1980.

[14] J. M. Mikkelson, L. A. Hall, A. K. Malhotra, S. D. Seccombe, and M. S. Wilson, "An NMOS VLSI Process for Fabrication of a 32b CPU Chip," *IEEE ISSCC Digest, 24*, 106–107, February 1981.

[15] P. B. Ghate, "Multilevel Interconnection Technology," *IEEE IEDM Digest*, 126–129, December 1984.

PROBLEMS

7.1 **(a)** What is the sheet resistance of a 1-μm-thick aluminum-copper-silicon line with a resistivity of 3.2 μohm-cm?

(b) What would be the resistance of a line 500 μm long and 10 μm wide?

(c) What is the capacitance of this line to the substrate if it is on an oxide which is 1 μm thick? (Assume that you can use the parallel-plate capacitance formula.)

(d) What is the RC product associated with this 500-μm line?

7.2 **(a)** Repeat Problem 7.1 for a polysilicon line with a resistivity of 500 μohm-cm.

(b) Repeat Problem 7.1 for a titanium silicide line with a resistivity of 25 μohm-cm.

(c) Repeat Problem 7.1 for a copper line with a resistivity of 1.7 μohm-cm.

7.3 **(a)** Compute estimates of the sheet resistance of shallow arsenic and boron diffusions by assuming uniformly doped rectangular regions with the maximum achievable electrically active impurity concentrations. (See Fig. 4.6.) Use hole and electron mobilities of 75 and 100 cm^2/V-sec, respectively, and a depth of 0.25 μm.

(b) Compare your answers with those for diffused lines obtained from Figs. 4.6 and 4.16. Use the maximum possible electrically active concentration for the boron and arsenic surface concentrations.

7.4 Suppose that a 500×15 μm aluminum line makes contact with silicon through a 10×10 μm contact window as shown in Fig. P7.4. The aluminum is 1 μm thick and is annealed at 450 °C for 30 min. Assume that the silicon will saturate the aluminum up to a distance $\sqrt{Dt}$ from the contact. D is the diffusion coefficient of silicon in aluminum which follows an Arrhenius relationship with $D = 0.04$ cm^2/sec and $E_A = 0.92$ eV. Assume that silicon is absorbed uniformly through the contact and that the density of aluminum and silicon is the same. How deep will the aluminum penetration into the silicon be?

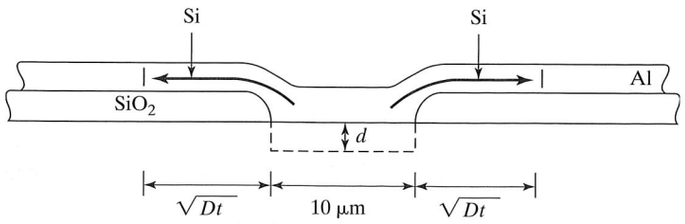


FIGURE P7.4

7.5 A certain process forms aluminum contacts to n^+ silicon through a 1×1 μm contact window resulting in a contact resistance of 0.5 ohms.

(a) What is the specific contact resistivity for this contact?

(b) What will the contact resistance be if the contact windows are reduced to 0.1×0.1 μm? Does this seem acceptable for a VLSI process?

7.6 Electromigration failures depend exponentially on temperature.

(a) What is the ratio of the MTFs of identical aluminum conductors operating at the same current density at 300 K and 400 K?

(b) At 77 K, (liquid-nitrogen temperature) and 400 K? Use $E_A = 0.5$ eV.

7.7 **(a)** What is the mean time to failure for the AlCu line in Fig. 7.8?

(b) How about for the copper line in the same figure?

7.8 An n^+ diffusion is used for interconnection. The surface concentration of the diffusion is $4 \times 10^{19}/cm^3$ and the junction depth is 4 μm. The diffusion is formed in a p-type wafer with a background concentration of $1 \times 10^{15}/cm^3$.

(a) What is the sheet resistance of this diffusion?

(b) Estimate the capacitance per unit length if the diffusion is 15 μm wide. Assume the rectangular geometry shown in Fig. P7.8 and use the step-junction-capacitance formula.

7.9 What is the maximum current that may be allowed to flow in an aluminum conductor 1 μm thick and 4 μm wide if the current density must not exceed 5×10^5 A/cm^2?

7.10 What is the maximum current that may be permitted to flow in an aluminum conductor 0.25 μm thick and 0.5 μm wide if the current density cannot exceed 10^6 A/cm^2?

7.11 What is the resistance of a 0.25×0.25 μm tungsten plug that is 1 μm high?

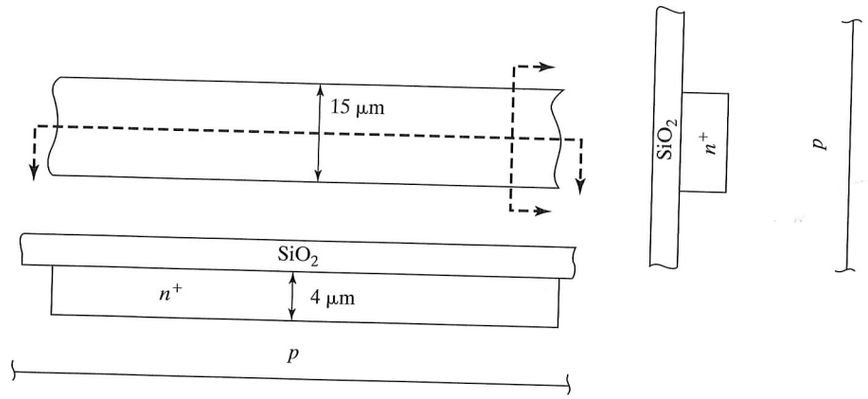


FIGURE P7.8

7.12 **(a)** What is the sheet resistance of a 0.5-μm-thick copper line with a resistivity of 1.7 μohm-cm?

(b) What is the resistance of a line that is 50 μm long and 0.5 μm wide?

(c) What is the capacitance of this line to the substrate if a 1-μm thick "low-K" dielectric is utilized with $\varepsilon = 2\varepsilon_o$?

(d) What is the *RC* product associated with the 50-μm line?

CHAPTER 8

Packaging and Yield

The low cost normally associated with integrated circuits results from mass production in which many wafers, each containing a large number of IC dice, are all processed together. There may be tens to thousands of dice per wafer and 25 to 200 wafers per lot. After wafer processing is completed, however, the dice must be tested, separated, and assembled in packages that are easy to handle and to mount in electronic systems. The testing and assembly operations substantially increase the cost of the final product.

In this chapter, we first present an overview of testing and die separation. Then we discuss IC assembly, including die attachment, wire bonding, and a survey of the various types of packages used with integrated circuits.

The ultimate cost of the integrated circuit is related to the total yield of assembled and tested devices. In the early stages in the development of a new process or circuit, we are lucky if a few functional dice are found per wafer. Late in the life of a process with a mature circuit design, yields of 60 to 80% are not uncommon. A discussion of the dependence of yield on defect density and die size concludes this chapter.

8.1 TESTING

Following metallization and passivation-layer processing, each die on the wafer is tested for functionality. Special parametric test dice are placed at a number of sites on the wafer. At this stage, dc tests are used to verify that basic process parameters fall within acceptable limits. To perform the tests, a probe station lowers a ring of very fine, needle-sharp probes into contact with the pads on the test die. Test equipment is connected to the circuit through the probes and controlled by a computer system. If the wafer-screening operation shows that basic process and device parameters are within specification, functional testing of each die begins.

Under computer control, the probe station automatically steps across the wafer, performing functional testing at each die site. Defective dice are marked with a drop of ink. Later, when the dice are separated from the wafer, any die with an ink spot is discarded. It has become impossible to exhaustively test complex VLSI devices such as