CHAPTER 8

Metallization

R. Liu

8.1 INTRODUCTION

Metallization is the process that connects individual devices together by means of microscopic wires to form circuits. These wires are necessary for a circuit to function, but they introduce parasitic resistance and capacitance, which degrade the performance of devices. Often these wires carry electrical currents in the milliampere range, resulting in very high current density (10^5 A/cm^2) that threatens to break the submicron-size interconnects. Thus, as devices scale down with ever-improved output, the loss within the metal interconnects dissipates a greater portion of the improved performance. Therefore, the essence of metallization practice is to minimize the undesirable effects.

In this chapter metallization issues and techniques will be discussed. The introductory section discusses the major issues. Sections 8.2, 8.3, and 8.4 examine the key metallization processes. Section 8.5 concentrates on the two key techniques for multilevel metallization—interlayer dielectric and global planarization. Section 8.6 covers some aspects of metallization reliability not included in Chapter 12, and Section 8.7 presents a summary and discusses the future materials and processing needs.

8.1.1 Parasitics in Device Performance

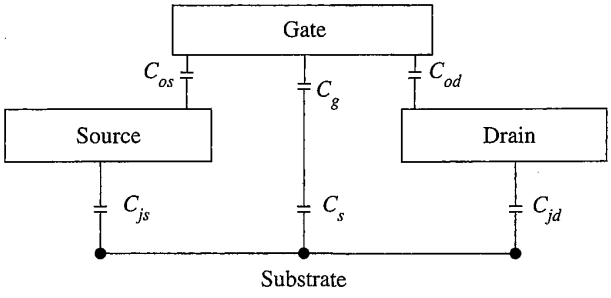
The switching speed of a MOSFET is determined by how fast the gate capacitor, plus all the parasitic capacitors, can be charged. Since $Q = CV$ and $i = dQ/dt = C(dV/dt)$, the rate of charging is dominated by (1) the transconductance, which determines the current output of the MOSFET, and (2) the total capacitance. To improve switching speed one has to reduce the total capacitance and maximize the transconductance of the MOSFET.

Micron Ex. 1029, p. 400
Micron v. YMTC
IPR2024-00911

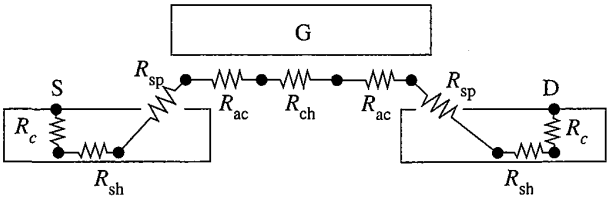
Figure 1*a* shows the principal parasitics of a MOSFET. C_g is the gate capacitance, C_{os} and C_{od} the overlap capacitance of the source and the drain with the gate, and C_{js} and C_{jd} the junction capacitance of the source and the drain, respectively. Note that C_{os} and C_{od} are part of the physical structure of C_g, but because they do not contribute to the channel current, they are purely parasitic. The effective gate capacitance can be written as $C_g' = C_g + C_{os} + C_{od}$. For a device with a 0.35 μm gate length $C_g' \approx 1.15C_g$, or C_{os} and C_{od} together contribute about 15% of the gate delay.

Junction capacitance is determined by the substrate doping concentration and the junction area. This capacitance is easily estimated using $C_j = \epsilon A/W$, where W is the depletion width given by $W^2 = 2\epsilon V/qN_d$; $\epsilon = 1.05 \times 10^{-12}$ F/cm, the dielectric permittivity of Si; and N_d, the substrate doping (Chapter 9). For $N_d = 10^{17}$ cm^{-3}, one can estimate a $C_j \approx 3 \times 10^{-16}$ F/(μm)^2, or 0.3 fF/(μm)^2. Although this looks like a very small capacitance, it is not insignificant compared to the gate capacitance C_g. For example, for a MOSFET with a gate length of 0.35 μm and a width of 1 μm, $C_g \approx 1.5$ fF. In this case, the junction capacitance of the source and the drain each contribute a parasitic of about 20% of C_g. The larger the total gate capacitance C_g' and the larger the junction capacitance C_j, the slower the MOSFET can switch on and off. The gate capacitance C_g, however, cannot be reduced because the output current of the MOSFET is proportional to C_g. To improve the switching speed, the parasitic capacitance must be minimized.

The transconductance of the MOSFET can be improved by reducing the total resistance. Figure 1*b* illustrates the resistive components in a MOSFET. The largest resistive component is the channel resistance, R_{ch}, which is similar to the case of capacitive components. The channel resistance is determined by the mobility of the



(*a*)



(*b*)

FIGURE 1
Parasitics in a MOSFET: (*a*) Capacitive parasitics; (*b*) resistive parasitics.

Micron Ex. 1029, p. 401
Micron v. YMTC
IPR2024-00911

carrier in the channel and is part of the device design optimization. All other resistive components are parasitics. The most important are[1] accumulation and spreading resistance $R_{ac} + R_{sp}$, the diffusion sheet resistance R_{sh}, and the contact resistance R_c. The accumulation resistance R_{ac} and the spreading resistance R_{sp} together represent the resistance to the current flowing from the source into the channel inversion layer and then spreading out into the drain. Combined, they make a large contribution to the parasitic resistance, but they are not affected by metallization. R_{sh} and R_c, on the other hand, are affected by junction design and contact metallization. For MOSFETs with gate length L larger than 0.35 μm, R_c seldom reaches 10% of channel resistance. As devices continue to shrink, however, R_c increases because the contact area shrinks. Consequently, at $L < 0.25$ μm, R_c can contribute significant resistance (Section 8.1.3).

Without parasitics, the intrinsic switching speed of a MOSFET can be estimated by $\tau \approx C_g V/I_{Dsat}$, where C_g is the gate capacitance; V, the power supply voltage; and I_{Dsat}, the saturation current. For an nMOSFET with $L = 0.35$ μm operating at 3.3 V, $\tau \approx 12$ ps, corresponding to a cutoff frequency $f_T \approx 50$ GHz. Adding the (mostly capacitive) parasitics, the speed degrades by about 60% to 20 ps.

Further scaling of devices will decrease C_g, C_{os}, and C_{od}, as well as C_{js} and C_{jd} by about $1/S$, where S is the scaling factor. Parasitic resistance, however, will increase as contact size shrinks and decreases the transistor current drive. In addition, the power supply voltage scales down and further reduces the output current. Therefore, without reducing the parasitic capacitance and resistance, higher switching speed will be difficult to achieve.

8.1.2 Parasitics in Circuit Performance

The switching speed of the transistor sets the upper limit of the circuit speed. The speed of a circuit, however, is determined by the capacitive and resistive loads on the transistors and is thus only proportional to the switching speed of the transistors. In the simplest case, a CMOS inverter is connected to another identical inverter, as shown in Fig. 2*a*. Figure 2*a*(*i*), (*ii*), (*iii*) shows a simple inverter circuit, two inverters connected by a wire of length L, and the resistance and capacitance of each component.

The switching delay of a single inverter can be estimated from the total capacitance of the circuit. The width of the pMOS is usually about twice that of the nMOS so as to match the output current. Therefore, the total capacitance is $\sim 3C_n$, where $C_n = C_{os} + C_{od} + C_{js} + C_{jd} + C_g$ is the capacitance of an nMOS. The switching delay of the inverter is then $t_{in} \approx 3\tau$, or about three times the switching speed of the nMOSFET (with parasitics). The capacitance of the wire adds another delay to the signal before it reaches the next inverter. This delay is given by $t_w = C_w V/i$, where C_w is the wire capacitance. The wire capacitance can be calculated using $C_w = \epsilon_{ox} Lwk/d$, where L, w are the length and the width of the wire, respectively; d is the thickness of the oxide separating the wire from the ground plane; $k \approx 3$, a geometry factor accounting for fringe fields; and $\epsilon_{ox} = 0.345$ pF/cm, the dielectric permittivity for SiO_2. For $w = d(= 0.5$ μm), we can calculate $C_w \approx 1000$ fF/cm.

Micron Ex. 1029, p. 402
Micron v. YMTC
IPR2024-00911

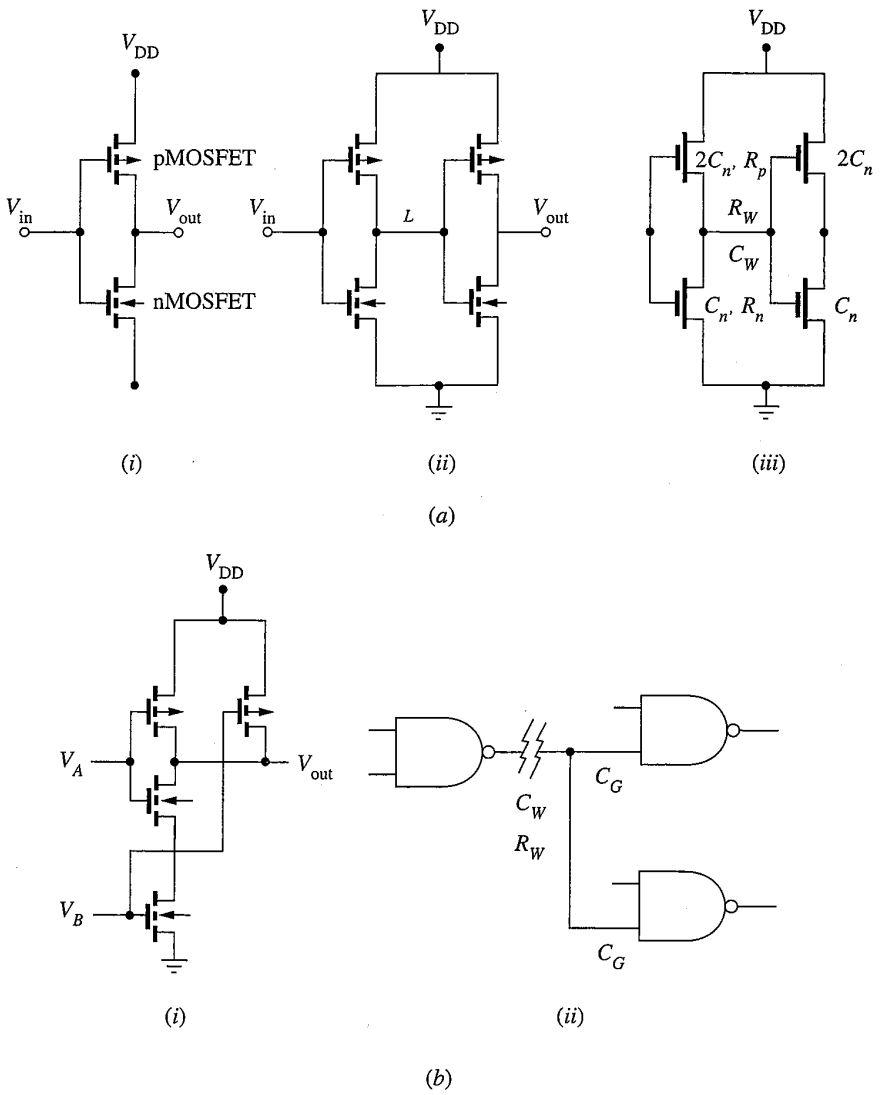


FIGURE 2
Parasitics in simple circuits: (*a*) CMOS inverters: (*i*) a CMOS inverter; (*ii*) two inverters connected by a wire of length *L*; (*iii*) capacitive and resistive components. (*b*) NAND gates: (*i*) a two-input NAND gate; (*ii*) a NAND gate connected through a long wire to two more NAND gates (fan-out 2).

The total capacitance of the inverter and the wire is $C = C_w + 3C_n$. For a minimum dimension nMOS the transistor width is about 0.5 μm, which gives a $C_n \approx 1.2$ fF. If the wire is 10 μm in length, then $C_w \approx 1$ fF, and the total capacitance is 4.6 fF. The current output of the nMOS is about 0.2 mA at 3.3 V, which gives a switching delay of $t = CV/i = 77$ ps. About 17 ps of the 77-ps delay is due to the capacitance load of the wire. The output of the second invert is further delayed by its capacitance. The total delay of the two inverters is then 77 ps + 60 ps = 137 ps. Without the wire capacitance, the delay would be 120 ps. Therefore, the capacitance of a 10-μm wire contributes about a 10% delay.

Micron Ex. 1029, p. 403
Micron v. YMTC
IPR2024-00911

If the wire length L connecting the two inverters is 1 mm instead of 10 μm, then $C_w = 100$ fF, and the total capacitance becomes 107.2 fF. The capacitive load is now dominated by the wire capacitance, and the switching delay becomes $\Delta t = CV/i = 1769$ ps $= 1.769$ ns. It is clear that to drive a large capacitive load, more current from the driving circuit is needed to obtain high-speed performance.

Figure 2*b* shows the parasitic capacitance and resistance seen by a logic gate. The two-input NAND gate requires two nMOSFETs and two pMOSFETs and thus has a total capacitance of 6 C_n. As this logic gate drives other logic gates at the end of a long wire (1 to 2 mm in length), the switching speed of the gate is much reduced. The current from the driver gate not only has to charge up the equivalent capacitance of 12 C_n but also has to charge up the extra capacitance of the long wire. Figure 3 shows the effects of *RC* delay and *IR* drop on the switching speed. The switching speed as a function of wire length without considering the effect of *IR* drop in the line is plotted in Fig. 3*a*. Note that even when the *RC* load on the wire is ignored, the switching speed of the logic gate is still about a factor of 10 slower than that for a MOSFET. This is unavoidable, since the extra load from the fan-out gates requires the charging of a larger total capacitance.

With the presence of a long wire, both a capacitive load C and a resistive load R_L are added. The resistive load R_L, however, is usually much smaller than the channel and the parasitic resistance of the nMOS and pMOS. For example, for a 1-mm Al wire 0.5 μm × 0.5 μm in cross section, $R_w \approx 120\ \Omega$, whereas the resistance of a 1-μm wide pMOSFET is > 5 kΩ. Therefore, R_L has little effect on circuit delay.

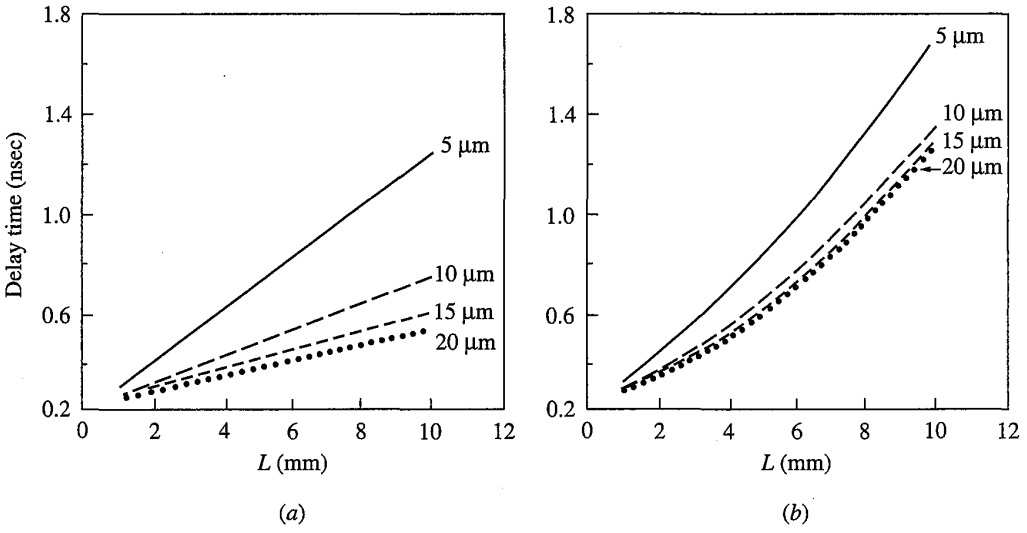


FIGURE 3
RC delay time for a CMOS inverter driving a metal line with length L and three fan-outs. The metal line width = 0.5 μm, and the dielectric thickness = 0.5 μm. Labels are for the transistor width of nMOSFETs. The width of pMOSFETs is 2× that of nMOS. (*a*) Delay time when no line resistance is considered. (*b*) Delay time when the 60 mΩ/□ Al line resistance is considered. Note that the benefit of wide transistors is diminished owing to the *IR* drop. (*Courtesy of H. I. Cong.*)

Micron Ex. 1029, p. 404
Micron v. YMTC
IPR2024-00911

When only the capacitive load of the long wire is considered, the switching delay increases linearly with the length of the wire, since the capacitance increases linearly with the wire length. To maintain a fast switching speed, one can send more current down the wire by increasing the width of the driving transistor. Thus, when the *IR* drop along the wire is ignored, it appears that the *RC* delay can be compensated by a higher current from wide transistors.

Consider the situation when the *IR* drop along the wire cannot be ignored. The *IR* drop along the wire adds a nonlinear component to the curve shown in Fig. 3*a*, now replotted in Fig. 3*b*. The reason for this nonlinearity is that the *IR* drop depends linearly on the wire length and, as a result, the delay time now increases quadratically with the wire length. Even more significant is the diminishing benefit of wide transistors. The higher current provided by the wide transistor now brings a larger *IR* drop along the wire, which results in a drop in the drain voltage seen by the driver transistor. With a large enough I, the *IR* drop reduces the drain voltage of the driver transistor sufficiently so that it operates in the linear region and behaves like a resistor. Once this condition is reached, a further increase in transistor width only keeps the transistor operating further in the linear region with no increase in its current output. Thus, when the wire resistance becomes significantly high, the *IR* drop imposes a self-limit to circuit speed.

8.1.3 Contact Issues

Ideally, the metal wires make contact with the semiconductor device without adding parasitic resistance. In practice, there is always some contact resistance and the best one can achieve is to minimize it. A common practice is to dope the Si or polysilicon heavily to produce a low-resistance ohmic contact. The heavy doping in Si reduces the thickness of the Schottky barrier and allows carriers to tunnel through the barrier.[2]

The quality of a metal-semiconductor contact is measured by the specific contact resistance, ρ_{sc}, which has dimensions of Ω-cm^2. Figure 4 shows the specific contact resistance on Si as a function of Si doping.[2] The resistance of a metal wire contacting the source, drain, or gate of a MOS device is simply ρ_{sc}/A, where A is the contact area. The specific contact resistance of a metal to n^+ Si may reach $\sim 1\times10^{-7}$ Ω-cm^2, whereas that to p^+ Si seldom goes below 2×10^{-7} Ω-cm^2. Thus, for a contact window of 1 μm to n^+ Si the contact resistance is $\sim$ 13 Ω. This is insignificant compared with the channel resistance of several kΩ.

Consider, however, a MOSFET with a channel length of 0.18 μm and with contact windows of 0.2 μm in diameter. The channel resistance is about 1 kΩ/μm, but now the contact resistance is $\sim$ 300 Ω. With these design rules the contact resistance is a significant parasitic that will reduce the performance of the MOSFET.

An innovative way to reduce the contact resistance is to increase the metal-semiconductor contact area without increasing the contact window size. This is done by metallizing the entire source/drain and gate area with a self-aligned silicide, or *salicide* (Section 8.3). The metal wiring now contacts the silicide instead of Si. Because a metal-to-metal contact usually has low specific contact resistance, $< 1 \times 10^{-8}$ Ω-cm^2, the contact resistance can be reduced by more than an order of

Micron Ex. 1029, p. 405
Micron v. YMTC
IPR2024-00911

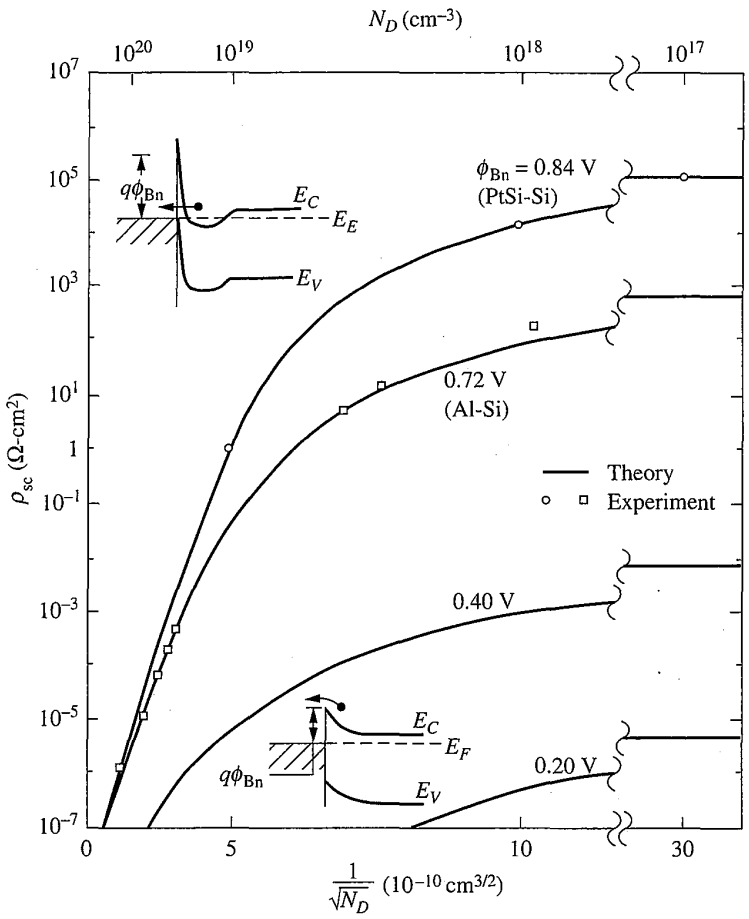


FIGURE 4
Specific contact resistance to Si. (*From Ref. 2 with permission.*)

magnitude without increasing the contact size. The use of silicides, however, causes other complications, which will be discussed in Section 8.3.

The specific contact resistance of metal-semiconductor interfaces represents the lower limit of contact resistance. In practice, both etching residues in the contact windows and damage of the Si by contact window etching may cause contact resistance 10 to 100 times higher than this limit. To minimize contact resistance, details of contact reactive ion etching (RIE) and contact cleaning are painstakingly designed to alleviate the etching damage and eliminate residues.

Vias, which connect different levels of metal wirings, also are affected by etching damage and residues but to a lesser degree. However, the tolerance for contact resistance in the via level is much smaller (usually $< 1\ \Omega$). Consequently, the via contact presents a challenge as difficult as that of the Si contact.

8.1.4 Multilevel Metallization

Nearly all MSI (medium-scale integration), LSI (large-scale integration), and earlier generations of VLSI circuits used only one level of Al wiring. Today, as well as in

Micron Ex. 1029, p. 406
Micron v. YMTC
IPR2024-00911

the foreseeable future, virtually all VLSI and ULSI circuits are made with two to five levels of Al or W wirings. Multilevel metallization provides greater flexibility in circuit design and, even more importantly, a substantial reduction in die size and thus chip cost. Curiously, multilevel metallization has been introduced reluctantly; many innovative ideas have been practiced to circumvent the use of more metal wiring layers.

The reluctance to adopt multilevel wiring is readily traced to the delayed development of processing and design tools. The development of a reliable, low-temperature, plasma-enhanced, interlayer dielectric (ILD) process, and the development of ILD planarization, prove to be the bottlenecks even today. In addition, the capability and user friendliness of CAD tools for logic and custom circuits continue to lag behind the need for faster multilevel wiring design.

Figure 5 compares a two-level metal and a three-level metal design of a segment of a standard cell ASIC (application-specific integrated circuit). The details of the standard cells are irrelevent to this discussion and are not shown. A significant advantage of the third level of metal is to reduce the routing channels between cells because routing in the horizontal direction can now be done over the cell. Other types of nonmemory circuits can, in general, realize a similar advantage with the additional level of wiring. The reduction in chip area is 20 to 35%, depending on the particular design.

The performance of the circuit is increased because of the smaller die size and the added design flexibility. The economics of an additional level of metal, however,

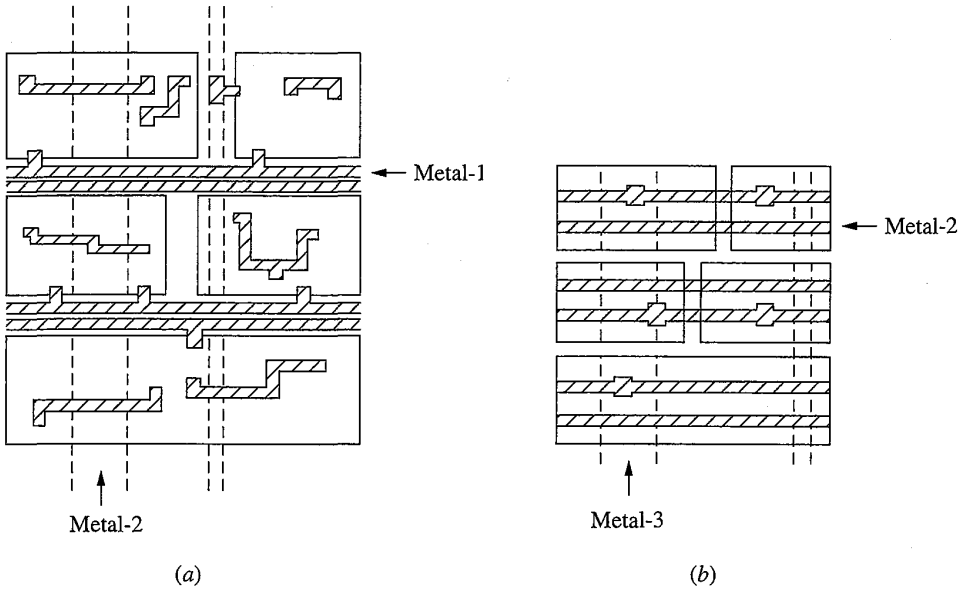


FIGURE 5
Two-level metal and three-level metal routing of wires for a segment of standard cell design: (*a*) 2LM requires routing channels between functional cells; (*b*) 3LM can route over cells and reduce area by ~ 30%. Metal-1 (not shown) is used only for local interconnection within the cell.

Micron Ex. 1029, p. 407
Micron v. YMTC
IPR2024-00911

are not always simple. For a simple CMOS process with, for example, 12 masks using two levels of metal, the addition of the third level of metal adds approximately 15% to the processing cost. If the new metal level does not result in lower yield, then the chip yield will increase because of die size reduction. This will more than compensate for the increased processing cost.

Each added level of metal, however, creates more processing difficulty. If there is a significant loss in yield, or expensive new processes are needed to achieve high yield, then the economic benefit is not ensured. Therefore, to utilize multilevel metallization both design and processing engineering are vital to achieve minimum die size, low cost, and high-yield processing.

8.1.5 Metallization Reliability

With each new level of integration the current output for the devices increases, yet the cross section of the Al wire decreases. In addition, these thinner wires are buried under more layers of other wiring and dielectric materials and are subjected to increased mechanical stress. The reliability of the Al wiring therefore becomes an important concern. The fundamental mechanism and parameters that affect electromigration and stress migration are discussed in Chapter 12. In Section 8.6 we will examine the impact of scaling and multilevel interconnect on the reliability.

8.2 METAL DEPOSITION TECHNIQUES

Metals are deposited by a variety of techniques but principally by physical vapor deposition (PVD) or by chemical vapor deposition (CVD). These techniques and their applications will be discussed in the following sections.

8.2.1 Physical Vapor Deposition

The most common forms of PVD are evaporation, e-beam evaporation, plasma spray deposition, and sputtering. Evaporation and e-beam evaporation were used extensively in earlier generations of MSI and LSI but have since been replaced by sputtering. Today, all VLSI and ULSI devices are fabricated by sputtered metal, and we will focus only on this method. Details of evaporation principles and techniques can be found in Refs. 3 and 4.

Sputtering prevailed because of (1) the high deposition rate afforded by modern cathode and target design, (2) the capability to deposit and maintain complex alloy compositions, (3) the capability to deposit high-temperature and refractory metals, (4) the capability to maintain well-controlled, uniform deposition on large (200 mm diameter and larger) wafers, and (5) the capability, in multichamber systems, to clean the contact before depositing metal. No other deposition technique can offer all of these advantages, and for Al alloys and contact metals, sputtering has been and probably will continue to be the technique of choice.

Micron Ex. 1029, p. 408
Micron v. YMTC
IPR2024-00911

Fundamentals of the sputtering process

The basic sputtering process employs a plasma of Ar gas. The Ar plasma can easily be generated by a glow discharge, as shown in Fig. 6*a*. A glow discharge typically consists of several different dark and glowing regions. Near the cathode is a very thin layer of ion space charge, followed by a dark space or gap (the Crookes dark space) where most of the potential between the cathode and the anode drops, followed by a negative glow (the glow discharge), a negative dark space (the Faraday dark space), and positive columns at the anode. The Ar plasma (the negative glow region) is sustained by the secondary electrons generated from ion bombardment of the cathode. These electrons accelerate away from the cathode, acquiring nearly the full energy of the dc potential across the electrodes when crossing the cathode dark space, and collide with and excite the Ar atoms in the glow discharge region, creating the plasma. After several collisions, the electrons lose their energy and continue to

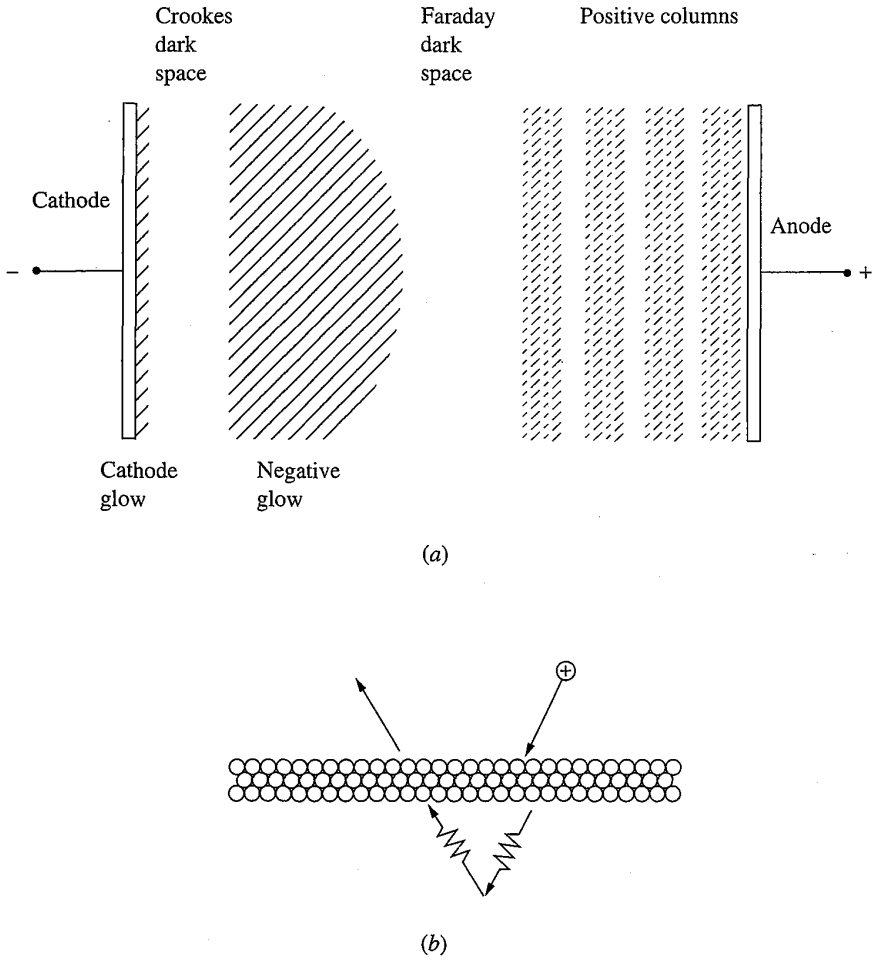


FIGURE 6
Sputtering process: (*a*) Characteristic regions of a glow discharge; (*b*) momentum transfer through phonon interaction.

Micron Ex. 1029, p. 409
Micron v. YMTC
IPR2024-00911

drift toward the anode. Since they cannot excite any Ar atoms, nor can they create new electron-ion pairs, this region is dark (the Faraday dark region). After that, the electrons are gradually accelerated again by the anode, creating columns of faint glows and dark regions called the positive column. The alternating faintly glowing and dark regions are created by electrons acquiring enough energy to excite Ar atoms and then falling below the minimum energy level again.

Since most of the potential drop in a glow discharge is across the Crookes dark space, if one moves the anode forward beyond the position of the positive column the negative glow is virtually unaffected. As a result, virtually all sputtering systems are designed with the anode very close to the negative charge. The closer the target is to the wafer, the higher the deposition rate. The details of basic sputtering techniques can be found in Refs. 5 and 6.

The mechanism of sputtering is quite simple. The charged ions in the Ar plasma diffuse into the Crookes dark zone, acquire virtually all the energy from the voltage drop, and hit the cathode (target) surface. The momentum of the Ar ion transfers to the target material through phonon interactions. The result is the ejection of one or more atoms from the surface of the target, as shown in Fig. 6*b*. The ejected atom, being neutral, flies through the plasma (with a small chance of being ionized on the way) and lands on the wafer. The angular distribution of sputtered materials has been studied[6] and found to follow the cosine law. Therefore, like evaporation processes, sputter deposition occurs essentially along a line-of-sight path with a cosine distribution, resulting in poor step coverage if the surface topography of the wafer is abrupt.

The sputtering yield from the target varies widely with the material and with the energy and the incident angle of the ions. The sputtering yield for a number of materials peaks at incident angles of 45 to 60°. Although this does not have a direct bearing on the discussion of PVD, it provides the basis of the use of Ar sputter-etch, for example, for ILD processing (Section 8.5). Sputtering yield as a function of target material and ion energy can be found in Ref. 6.

In sputter deposition, the most important concern is to increase the ion bombardment rate on the cathode so that a reasonable deposition rate can be achieved. The glow discharge, on the other hand, relies on the secondary electrons from the target and is intrinsically inefficient. Consequently, means to increase the secondary electron production, as well as to increase the efficiency of ionization, are designed to improve the sputtering rate. One common design is the use of an $\mathbf{E}\times\mathbf{B}$ field to enhance the trapping of electrons. The cathode design using a magnetic field is called *magnetron.* The invention of magnetron sputtering greatly improved the sputter deposition rate and made sputtering the leading technique of PVD deposition.

Sputtering can also be achieved by applying an rf field instead of a dc field to create the plasma. In an rf field, both electrons and ions receive an ac field, but because of the high frequency only electrons can respond to the rf field. Consequently, the ions act as if the field were dc. The benefit of rf sputtering is that the electrons are oscillating with the field, resulting in no charge accumulation at the target even if it is an insulator. Therefore, rf sputtering is used for depositing insulating or semiconducting materials.

Figure 7 shows a magnetron sputtering system. The particular material to be sputtered is made into a disc that is thermally bonded to the cathode. Quite a large

Micron Ex. 1029, p. 410
Micron v. YMTC
IPR2024-00911

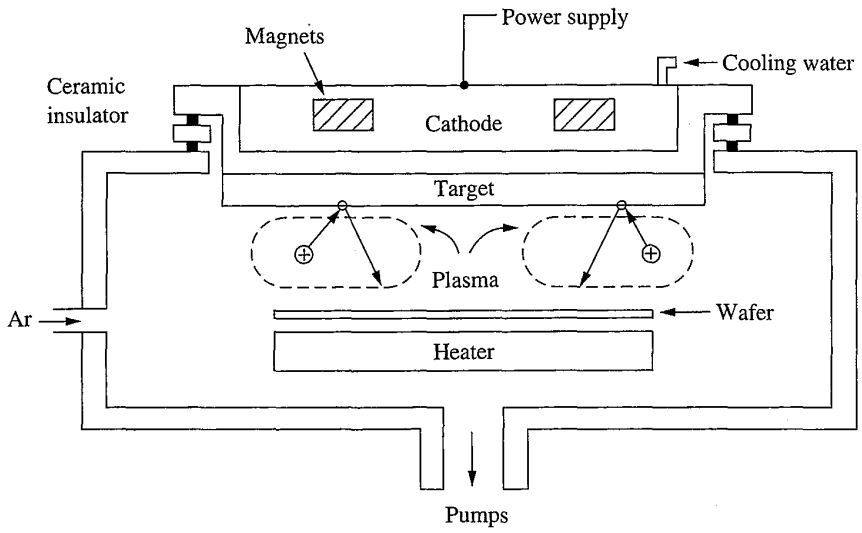


FIGURE 7
A dc-magnetron sputtering system.

amount of power is supplied to the Ar plasma to maximize the sputtering rate (in the range of 3 kW to 20 kW). Most of this power is absorbed by the sputtering target, which needs to be cooled through thermal contact with the cathode. Usually, the cathode is water-cooled. Apart from the high rate, another critical criterion for sputter deposition is the uniformity of the deposited film, especially on large wafers. To achieve high rate and uniformity, new sputtering cathodes are designed with rotating, rare earth, high-strength permanent magnets. This improved magnetron cathode is the basis for modern high-rate, single-wafer deposition systems.

Film properties

Sputtered film properties have been studied thoroughly.[7] The substrate temperature and the Ar pressure are the most important parameters. At low substrate temperature, the surface mobility of the sputtered atoms is low, resulting in porous films with poor physical properties. At higher temperatures with $T/T_m \approx 0.5$ to 0.7 (T_m is the melting temperature of the metal), surface diffusion occurs during sputtering and results in films with good physical properties comparable to the bulk. The films are columnar in this region because the grains grow unidirectionally. The grain size increases with increasing temperature because of lower density of nucleation at higher temperature.

At higher temperatures, $T/T_m > 0.7$, bulk diffusion occurs during deposition, and the film consists of equiaxially larger grains. This pattern is caused by recrystallization during film deposition. Although large-grained films are desirable (e.g., for electromigration resistance), in practice high-temperature sputtering has seldom been used until recently (see Section 8.4.3) because of equipment limitations. Most Al films are deposited between 200 and 300°C, which corresponds to $T/T_m \approx 0.5$ to 0.6. Good-quality Al films with grain size between 0.5 and 1.0 μm are achieved at these temperatures.

Micron Ex. 1029, p. 411
Micron v. YMTC
IPR2024-00911

The Ar pressure plays an important role in the film stress and electrical and physical properties. During high-rate sputtering, Ar gas can be incorporated in the sputtered film[8] to nearly 1 at% (atomic percent). Ar incorporation changes the film properties substantially. Even worse, since Ar is not tightly bonded with metal, it can condense during subsequent thermal treatment into bubbles, resulting in blistering or delamination of the film.

Generally, better film properties are achieved at a higher substrate temperature and a lower Ar pressure. In practice, the rate of deposition is a function of plasma density, which in turn is a function of the Ar pressure. Therefore, until the higher-plasma-density magnetron, which allows high-rate sputtering at low pressure, was invented, high-quality films were hard to achieve. Today, with even more efficient magnetron designs, good film properties for Al are achieved routinely.

Alloy sputtering

For electromigration (EM) resistance, Cu is alloyed to Al to enhance the EM lifetime. For contacts, the high solubility of Si in Al causes the Al to spike into the Si substrate, leading to source/drain junction leakage. This is caused by Si diffusion from the source/drain to the Al metal to satisfy the high solubility of Si in Al at low temperature.[9] To prevent Al spiking, Si is often alloyed into Al to reduce Si diffusion. Therefore, in practice, Al is always used in alloy forms and not as a pure metal.

Sputtering is especially suitable for alloy deposition. It is known that the sputtering yield is a strong function of the target material. This is also true for alloy sputtering. It is generally difficult to equalize the sputtering rate of various elements. However, when an alloy target is used, the differential sputtering rate for different elements in the alloy results in a composition that is enriched in the element with the lower yield. Consequently, the exposed target surface has a higher concentration of the element of low yield. This self-compensating effect is important because it ensures that the film deposited on the wafer has essentially the same alloy composition as the target material.

Reactive sputtering

The ions in the Ar plasma generally have energy near or higher than the binding energy of the target material. Consequently, refractory metals can be readily sputter-deposited. In addition, compounds such as SiC, TiN, etc., which are difficult to deposit by other PVD methods, can be sputtered readily.

Conductive compounds such as TiN are especially important for ULSI application as barrier and glue metal layers[10] (see Section 8.4 on CVD W plug) and as antireflective coatings over Al alloys. TiN can be sputtered by two techniques—by using a compound TiN target or by reactive sputtering. Both are used in ULSI fabrication.

Reactive sputtering of TiN is done by introducing nitrogen into the sputtering chamber in addition to the Ar plasma. The plasma provides enough energy to dissociate the nitrogen molecules into atomic nitrogen, which subsequently can interact with Ti to form TiN.[11] The film properties of TiN vary widely with deposition conditions[12] as shown in Fig. 8. With no N_2 flow, pure Ti is sputtered, as expected. With increasing N_2 flow, the film resistivity increases until a maximum is reached

Micron Ex. 1029, p. 412
Micron v. YMTC
IPR2024-00911

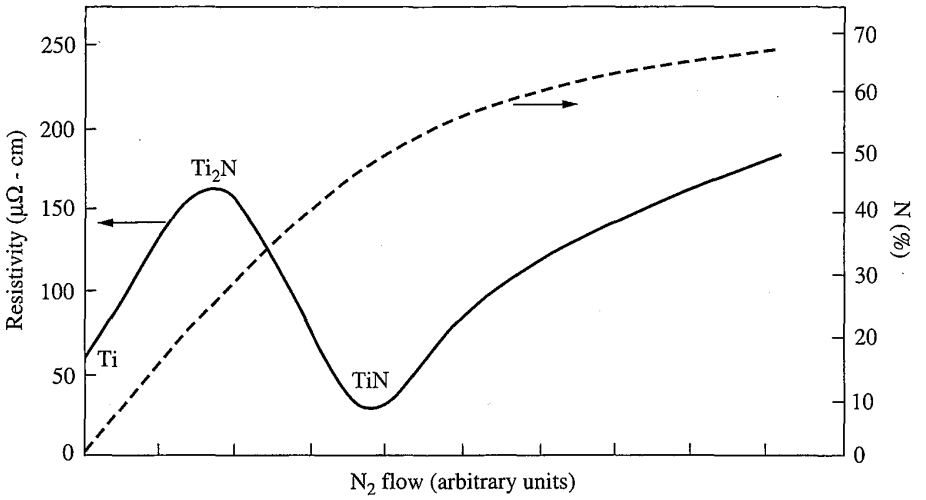


FIGURE 8
Properties of TiN as a function of N_2 flow rate.

at a Ti/N ratio of ~ 2 : 1. Then the resistivity decreases again and reaches a minimum when the Ti/N ratio is about 1. This is the condition when stoichiometric TiN is deposited. A further increase in the N_2 flow results in nitrogen-rich TiN with high resistivity.

TiN is a high-melting-point, hard metal. In addition, the thermal expansion coefficient of TiN is different from Si, and as a result TiN deposited at 300 to 400°C can have a high compressive stress in the range of 1000 MPa. This brittle, high-stress material naturally is a potential source for particle generation, e.g., from flaking from the walls of the sputtering chamber. Careful design of the sputtering equipment and a carefully controlled deposition process are the prerequisites for particle-free reactive TiN deposition.

Step coverage

During sputtering, the atoms ejected from the sputtering target follow the cosine distribution.[5] To avoid contamination from residue gas, sputtering requires high vacuum and high-purity Ar with a pressure of 1 to 10 mtorr. At this pressure the mean free path is on the order of several centimeters, or approximately the distance between the target and the wafer. Therefore, atoms ejected from the target experience few collisions with the ambient and fly to the wafer essentially following a line-of-sight path.

The line-of-sight trajectory of the atomic flux causes poor coverage on vertical or steeply slanting surfaces. For metal wiring this is undesirable, since the resistance of the wire is increased by the thin sections. An even more serious difficulty is deposition of metal into contact windows and vias. Because of the severe geometric restrictions, the sidewalls and the bottoms of contacts and vias receive only 10% or less of the metal deposited on the top surface. Consequently, for contacts with a high aspect ratio, the very thin metal in the contacts causes severe reliability problems.

The planar-type sputtering cathode improves the metal step coverage only marginally. Adding a collimator, thus artificially modifying the cosine distribution

Micron Ex. 1029, p. 413
Micron v. YMTC
IPR2024-00911

of the atomic flux, improves the bottom coverage in high-aspect-ratio contacts, but at the expense of worsened sidewall coverage. Increasing the Ar pressure, thus increasing the number of collisions, causes a deterioration of the film property. Higher substrate temperature improves the surface mobility of Al atoms and provides better step coverage, but the nucleation and growth steps need to be carefully controlled (see Section 8.4.3).

For critical contact application, the poor step coverage of sputtering eventually forces the adoption of CVD metal-plug processes.

8.2.2 Directional Enhancement—Collimation

The metal coverage at the bottom of a contact can be increased by employing directional enhancement. This is done simply by inserting a collimator between the sputtering cathode and the wafer.[13] The collimator replaces the Si wafer as the electrical ground for the plasma. Neutral species sputtered off the target at a high angle are then intercepted by the collimator and deposit on the side wall of the collimator instead of reaching the wafer. Consequently, only atoms that are ejected within a small angle to the normal of the target will reach the wafer. These atoms have a much higher chance of reaching the bottom of the contact hole than those with larger angles.

Figure 9*a* shows the bottom coverage obtained by collimated sputtering compared with that from conventional sputtering. Simulation results using a cosine source distribution and a line-of-sight path agree well with available experimental

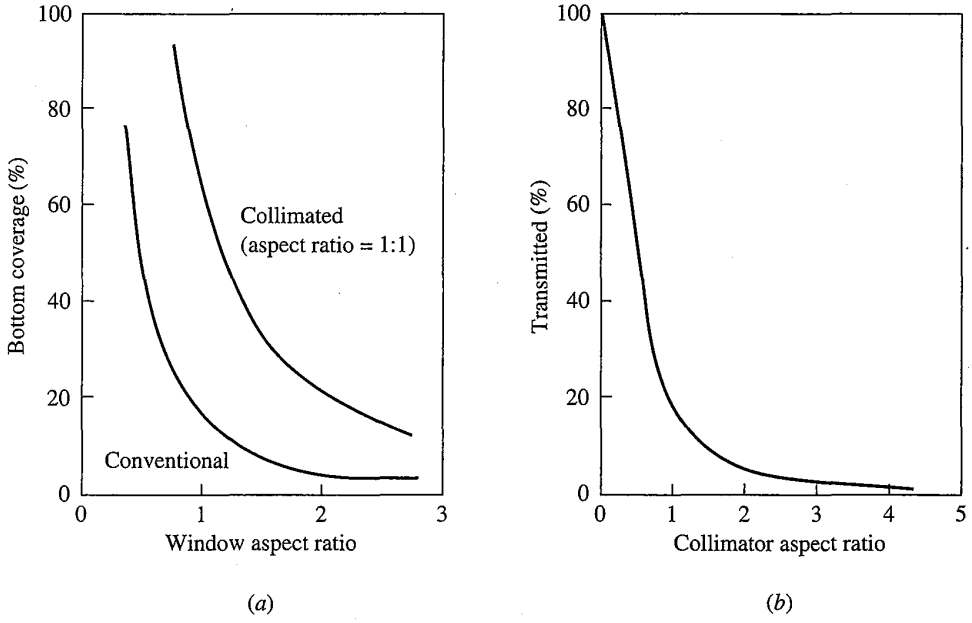


FIGURE 9
Properties of collimated sputtering: (*a*) Bottom coverage in windows; (*b*) transparency of the collimator as a function of its aspect ratio.

Micron Ex. 1029, p. 414
Micron v. YMTC
IPR2024-00911

data, indicating that the collimator works simply as an interceptor that reduces the flux from large-angle ejectors. The improvement in the bottom coverage is significant,[13] from $< 10\%$ to $> 25\%$ for a contact hole aspect ratio of $2 : 1$.[13,14]

The sidewall coverage, however, is degraded, as expected. The lack of sidewall coverage makes collimated sputtering undesirable for interconnect metal such as Al. But for contact and barrier material such as Ti and TiN, where bottom coverage is crucial, collimation provides a simple extension of an existing process and can be easily integrated into IC processing. As a result, collimated sputtering for Ti and TiN is quickly adopted in CMOS processing for 0.35-μm and 0.25-μm designs.

Collimation, however, also means discarding a large portion of sputtered material from the target, since only material ejected within a small angle passes through the collimator. Figure 9*b* shows the transparency of a collimator as a function of its aspect ratio. For a collimator aspect ratio (defined by length/diameter) of $1 : 1$, approximately 70% of sputtered material is intercepted by the collimator, and for an aspect ratio of $1.5 : 1$, nearly 90% is intercepted.[15] The low transparency of the collimator increases the cost of deposition because (1) the deposition rate is drastically reduced and (2) the target utilization is reduced and, therefore, both the material cost and the maintenance cost increase. Thus, even though collimation is easily retrofittable to IC processing, the higher cost makes this process less desirable if a low-cost alternative (such as CVD) can be developed.

8.2.3 Chemical Vapor Deposition

A number of metals and metal compounds can be deposited by chemical reaction or thermal decomposition of precursors. Techniques for depositing Al,[16–20] Cu,[21–24] WSi_2[25–27] TiN,[28–31] and W,[32–37] have all been developed and applied to devices and/or prototype circuits. However, except for WSi_2 and W, CVD metals are not widely used yet because of immature process and equipment development.

Metal CVD techniques

The nature of metal CVD is not different from Si or SiO_2 CVD. A precursor chemical containing the desired metal and a chemical reaction (including decomposition), which has the elemental metal or the metal compound as the final product, are the key ingredients. Usually the wafer needs to be heated to 100 to 800°C to provide the initial thermal energy to overcome the reaction barrier. The kinetics of the CVD process have been discussed in detail in Chapter 5, and the metal CVD follows the same general principle.

CVD metal can also be deposited in a plasma. The advantage of plasma deposition is that the energy required to overcome the initial reaction barrier can be supplied by the plasma, and the deposition temperature can be reduced. This reduction is especially important for metal or metal compounds used for vias and multilevel interconnects since their tolerance for high temperature is limited.

CVD W

The chemical vapor deposition of W and WSi_2 is widely used for circuit fabrication.[38] WSi_2 is used on top of gate polysilicon as a *polycide* structure and for

Micron Ex. 1029, p. 415
Micron v. YMTC
IPR2024-00911

local interconnect.[26,27] CVD W is used both as a contact plug and as a first-level metal.

The basic chemistry for CVD W is straightforward:

$$WF_6 + 3H_2 \rightarrow W + 6HF \tag{8.1}$$

$$2WF_6 + 3Si \rightarrow 2W + 3SiF_4 \tag{8.2}$$

$$WF_6 + SiH_4 \rightarrow W + SiF_4 + 2HF + H_2 \tag{8.3}$$

$$2WF_6 + 3SiH_4 \rightarrow 2W + 3SiF_4 + 6H_2 \tag{8.4}$$

In addition, WF_6 can be reduced by Al and Ti through

$$WF_6 + 2Al \rightarrow W + 2AlF_3 \tag{8.5}$$

$$2WF_6 + 3Ti \rightarrow 2W + 3TiF_4 \tag{8.6}$$

In practice, the process is more complex, with several intermediate reaction products such as WF_4 and WF_5 involved.[39]

During CVD W deposition, the wafer is held on a heated chuck that is between 400 and 500°C, opposite a shower head where the WF_6, H_2, or SiH_4 gases are injected. Usually a two- or three-step process is used. SiH_4 is introduced without any flow of WF_6 to initiate the deposition of a very thin layer (a few nanometers) of amorphous Si as a prenucleation layer. This step is followed by a $SiH_4 + WF_6$ silane reduction nucleation process, and then the high-rate $H_2 + WF_6$ hydrogen reduction deposition. At the nucleation stage less than 100 nm of W is deposited, and the bulk of the W deposition is by hydrogen reduction.

This complex procedure was developed to ensure that during the initial nucleation stage the Si from the source/drain area did not participate in the chemical reaction. As seen from Eqs. (8.2), (8.5), and (8.6), WF_6 reacts readily with Si, Al, and Ti. The free energies of the WF_6 reduction processes [Eqs. (8.1)–(8.6)] can be calculated from standard thermochemical tables.[40] The ΔG values of these reactions are shown in Table 1. It is clear that hydrogen reduction is the least favorable process, whereas reduction of WF_6 by Si, Al, and Ti are highly favorable. When WF_6 reacts with the Si from the source/drain region, it causes junction leakage.[41,42] Therefore, precautions must be taken (by introducing SiH_4 first) to avoid the consumption of Si from the substrate. WF_6 also reacts with lower-layer Al wiring and forms AlF_3 [Eq. (8.5)], which increases the via contact resistance.[43] The most common problem, however, is the reaction of WF_6 with the Ti glue layer [Eq.(8.6)]. This reaction causes the delamination of the TiN and "volcanoes." (See Section 8.4.)

TABLE 1
Free energy changes for CVD W deposition

Reaction	ΔG at 500°K (KJ/g-atom W)
$WF_6 + 3H_2 \rightarrow W + 6HF$	−113
$WF_6 + 1.5Si \rightarrow W + 1.5SiF_4$	−478
$WF_6 + 1.5SiH_4 \rightarrow W + 1.5SiF_4 + 3H_2$	−870
$WF_6 + 1.5Ti \rightarrow W + 1.5TiF_4$	−1037
$WF_6 + 2Al \rightarrow W + 2AlF_3$	−1162

Micron Ex. 1029, p. 416
Micron v. YMTC
IPR2024-00911

Even with elaborate deposition procedures, junction leakage caused by the local attack of WF_6 can still be a problem. "Wormholes" with diameters of only a few nanometers have been observed to penetrate several hundred nanometers into the Si.[44] Consequently, a barrier metal layer such as TiN or TiW is almost always used before CVD W deposition. The barrier metal layer serves more than one function. In addition to reducing the junction leakage, this metal layer also provides adhesion of the CVD W on SiO_2 and prevents the generation of particles from poorly grown CVD W on SiO_2.

W deposited by a CVD process has an electrical resistivity 1.5 to 2 times that of the bulk W. This increase is due mainly to the incorporation of F, which can reach > 1 atomic percent, in the W film. As a result, CVD W has a resistivity 3 to 4 times that of Al and is unsuitable for interconnect purposes except for short runs. CVD W also grows into ragged morphology with poor reflectivity, even though the grain size is not large (usually < 200 nm), because the exposed W surface consists of the as-grown crystal facets with no preferred orientation, resulting in random reflection of light. Therefore, for a CVD W thicker than 200 nm, patterning of submicron W lines is difficult.

CVD TiN

TiN is widely used as a barrier-metal layer for CVD W deposition and can be deposited by sputtering from a compound target or by reactive sputtering. CVD TiN can provide a better step coverage than PVD methods and is potentially more economical than collimated sputtering.

CVD TiN can be deposited[28,45,46] using $TiCl_4$ and NH_3; H_2/N_2; or NH_3/H_2:

$$6TiCl_4 + 8NH_3 \rightarrow 6TiN + 24HCl + N_2 \tag{8.7}$$

$$2TiCl_4 + 2NH_3 + H_2 \rightarrow 2TiN + 8HCl \tag{8.8}$$

$$2TiCl_4 + N_2 + 4H_2 \rightarrow 2TiN + 8HCl \tag{8.9}$$

The deposition temperature is 400 to 700°C for the NH_3 reduction and is > 700°C for the N_2/H_2 reaction. Generally, the higher the deposition temperature, the better the TiN film and the less Cl incorporated in the TiN. However, even the best TiN produced by this process contains $\sim 0.5\%$ of Cl. The lower-temperature processes result in even higher ($\sim 5\%$) Cl in the TiN,[45] causing concerns about the corrosion reliability of the Al wiring. CVD TiN by $TiCl_4$ processes at near 700°C has good electrical resistivity, $\sim 100\ \mu\Omega$-cm or less, but has high stress, ~ 2000 MPa. Because of the high deposition temperature ($\geq$ 650°C) these processes are unsuitable for via applications.

TiN can also be deposited by using metal-organic precursors.[47–50] These precursors are usually in the form of a Ti-alkylamine complex. One example is the use of tetrakis-(dimethylamido)-Ti [also written as TDMAT, or $Ti(NMe_2)_4$] and NH_3; the reaction[51] follows the $TiCl_4 + NH_3$ reaction closely:

$$6Ti[N(CH_3)_2]_4 + 8NH_3 \rightarrow 6TiN + 24HN(CH_3)_2 + N_2 \tag{8.10}$$

Similar reactions using tetrakis-(diethylamido)-Ti (TDEAT) have also been reported.[50]

TiN deposition by metal-organic precursors can be done at low temperature ($\leq$ 450°C) and imposes no danger of Cl incorporation. The low deposition temperature

Micron Ex. 1029, p. 417
Micron v. YMTC
IPR2024-00911

makes the metal-organic CVD (MOCVD) process suitable for use on all metal levels. However, the deposited films have lower densities and are unstable.[52] Because of carbon and oxygen inclusion the films have high resistivity, $\geq$ 500 $\mu\Omega$-cm, compared with the 40 to 100 $\mu\Omega$-cm for reactive sputtered TiN. In addition, the step coverage[52] for high-aspect-ratio vias is only marginal. As a result, low-temperature MOCVD TiN is not an ideal barrier material yet. TiN deposited by TDEAT at 425°C[50] has better step coverage and lower resistivity (180 $\mu\Omega$-cm). Recently, mixed amorphous/crystalline CVD TiN has been deposited[53] using TDMAT with good step coverage ($\sim$70%).

Low-temperature deposition of TiN using $TiCl_4$ and NH_3 with plasma assistance has also been reported.[29,54] Similar to plasma-enhanced dielectric processes, with an rf plasma TiN can be deposited at about 500 to 600°C. Plasma-enhanced CVD TiN contains lower Cl and has better physical and electrical properties than plain CVD. However, the Cl content is still in the 1% range, and the film properties degrade quickly below 500°C. Recently, high-density ECR (electron cyclotron resonance) plasma-enhanced deposition of TiN has been reported. By using a $TiCl_4$ precursor, good quality TiN with resistivity about 40 $\mu\Omega$-cm has been achieved at 400°C.[55] By using a TDMAT precursor, TiN has been deposited by downstream ECR plasma at $<$ 400°C.[56,57] TiN with a resistivity of 200 to 300 $\mu\Omega$-cm and low C and oxygen content has been reported.

CVD Ti

Although an excellent barrier material, TiN directly in contact with Si causes high contact resistance. A Ti layer below the TiN is usually used to ensure good contact resistance. CVD of Ti, however, is difficult because Ti is highly electronegative and precursors that can reduce to elemental Ti are thermodynamically unfavorable. Plasma-assisted CVD is necessary to impart the extra energy needed. Recently, ECR plasma-assisted CVD of Ti has been reported.[55,58] At 400°C the ECR-enhanced CVD Ti reacted directly with Si to form $TiSi_2$. Both good contact resistance and junction leakage were observed.

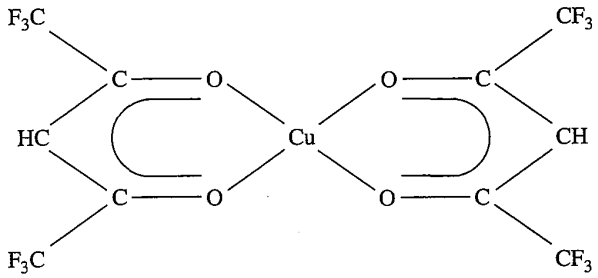
CVD Cu

Copper can be deposited by CVD readily. Because of its low resistivity and good electromigration resistance, Cu metallization has received much attention recently. There are two types of metal-organic precursors for Cu: the divalent Cu^{II} and the monovalent Cu^{I} compounds. The Cu^{II} precursors have the form of Cu^{II} (β-diketonate)$_2$. The most commonly used Cu^{II} precursor is bis-hexafluoroacetylacetonate-Cu^{II}, or Cu(hfac)$_2$. The structure $(CF_3COCHCF_3CO)_2Cu$ is shown in Fig. 10. Cu(hfac)$_2$ is a solid at room temperature and sublimes at low heat (35 to 130°C). In a CVD system, the Cu(hfac)$_2$ compound is heated, and a carrier gas (Ar or H_2) is passed through the bed of Cu(hfac)$_2$ powder to bring the vapor into the reaction chamber. Although Cu(hfac)$_2$ can decompose directly without a reaction with another chemical, the deposition rate is low and the Cu film is inferior. Higher deposition rates and better quality films are achieved by hydrogen reduction:[27,58,59]

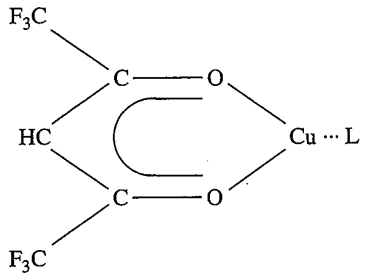
$$Cu(hfac)_2 + H_2 \rightarrow Cu + 2H(hfac) \tag{8.11}$$

Although the Cu(hfac)$_2$ + H_2 reaction starts at $\sim$175°C, it does not reach the surface

Micron Ex. 1029, p. 418
Micron v. YMTC
IPR2024-00911



(*a*)



(*b*)

FIGURE 10
Molecular structures of (*a*) $Cu^{II}(hfac)_2$; (*b*) $Cu^{I}(hfac)L$.

reaction-limited stage until 320°C.[59] To get lower resistivity Cu films, one must adjust the deposition temperature to 350 to 450°C[59,60,61] Because Cu is likely to be used with low-permittivity polymeric dielectric, the high deposition temperature is undesirable.

Copper from $Cu(hfac)_2$ CVD can be deposited as blanket films or selectively. Study of the kinetics of deposition[60] using $Cu(hfac)_2$ and H_2O discovered an incubation period for growth that differs on Pt and on SiO_2. The difference in nucleation time can be explained[60,61] by the catalytic action of Pt to help the chemisorption and decomposition of $Cu(hfac)_2$ through

$$Cu^{II}(hfac)_2 \rightarrow Cu^{I}(hfac) + hfac \tag{8.12}$$

$$Cu^{I}(hfac) \rightarrow Cu + hfac \tag{8.13}$$

$$hfac + H \rightarrow Hhfac \tag{8.14}$$

$$Hhfac + Hhfac \rightarrow (Hhfac)_2\uparrow \tag{8.15}$$

Equation 8.13 indicates that $Cu^{I}(hfac)$ is a more readily usable compound as a Cu precursor. However, $Cu^{I}(hfac)$ is not stable, and better precursors need to be found.

Various Cu^{I} precursors have been tested[22] in search of a better CVD process. The most promising is the Cu^{I} β-diketonates. This group of precursors has the form of $Cu^{I}(\beta$-diketonate)L where L is a neutral ligand weakly bonded to the Cu. Examples of L are 2-butyne, vinyltrimethylsilane (vtms), trimethylphosphine $(PMe)_3$, etc. Hexafluoroacetylacetonate (hfac) is the most commonly used β-diketonate. The structure of $Cu^{I}(hfac)L$ is similar to $Cu^{II}(hfac)_2$, with the L ligand replacing one of the hfac rings (Fig. 10*b*).

Micron Ex. 1029, p. 419
Micron v. YMTC
IPR2024-00911

Depending on the L ligand, CVD using Cu^I compounds may be selective, blanket, or both. $Cu^I(hfac)(PMe)_3$ was found to deposit Cu on metal selectively only,[62] but $Cu^I(hfac)$(2-butyne) deposits Cu only blanketly.[63] Both precursors can deposit clean Cu at high rate (200 to 1000 nm/min), but both are solids at room temperature. Recently, a liquid precursor $Cu^I(hfac)$(vtms) was synthesized,[64] and low-resistivity Cu[23,65,66] at $< 200°C$ was achieved.

$Cu^I(hfac)L$ deposits pure Cu through a disproportionation mechanism:[62,63]

$$2Cu^I(hfac)L \rightarrow Cu^0 + Cu^{II}(hfac)_2 + L \tag{8.16}$$

Both $Cu^{II}(hfac)_2$ and L are stable gaseous compounds and desorb from the surface readily at the deposition temperature. Detailed reaction kinetics for $Cu^I(hfac)$(vtms) has been studied.[66,67] Because this material is liquid at room temperature and good Cu films can be achieved below 200°C, it is a suitable precursor for CVD Cu.

Copper lacks a dense oxide and is thus vulnerable to corrosion. Also, like Au, Cu can diffuse through SiO_2 and cause deep impurity levels in Si, reducing bipolar gain and causing junction leakage. In addition, Cu has no volatile compounds at room temperature and thus cannot be etched by the familiar RIE process at moderate temperatures ($\leq$ 200°C). Therefore, the use of Cu as interconnect depends on issues well beyond the deposition techniques.

CVD Cu has resistivity $\sim$ 10 to 20% above the bulk. To prevent corrosion and Cu diffusion into Si, a cladding layer (e.g., TiN or Ta) is needed. The lower resistivity of Cu is therefore not totally realizable. In addition to better conductivity, Cu has better electromigration resistance than Al because of its higher atomic weight. Therefore, Cu metallization and CVD Cu are still being intensely studied for future ULSI application.

CVD Al

Al can be deposited by several CVD precursors; the most extensively studied is tri-isobutyl-Al, $(C_4H_9)_3Al$ or TIBA.[17] The chemistry is a three-step decomposition:

$$TIBA + H_2 \rightarrow DIBAH + C_4H_8 \tag{8.17}$$

$$DIBAH + H_2 \rightarrow AlH_3 + 2C_4H_8 \tag{8.18}$$

$$AlH_3 \rightarrow Al + \tfrac{3}{2}H_2 \tag{8.19}$$

where DIBAH is the abbreviation for di-isobutyl Al hydride, or $(C_4H_9)_2$ AlH. The first reaction is at 40 to 50°C, before the gas reaches the wafer; and the second reaction occurs on the wafer with the wafer held at 150 to 300°C. In practice, DIBAH is the chemical precursor that decomposes on the wafer surface. However, the direct use of DIBAH is impractical owing to its low vapor pressure and low deposition rate.

The Al deposited by this process contains virtually no hydrocarbon residue and has a resistivity of $\sim$ 2.8 $\mu\Omega$-cm, close to that of bulk Al. When deposited on a good nucleation layer such as TiN, (111)-textured Al with high reflectivity can be achieved.[68]

A problem with the application of CVD Al to ULSI circuits is that there is no Cu in the CVD Al to provide electromigration resistance. A simple solution to this problem is to deposit $\sim$ 50% of the total thickness of Al by CVD and the rest by

Micron Ex. 1029, p. 420
Micron v. YMTC
IPR2024-00911

sputtering an Al-Cu alloy.[69] A subsequent heating to 250 to 400°C allows the Cu in the alloy to redistribute to the entire Al wiring. The maximum Cu concentration is determined by the solubility of Cu in Al at the annealing temperature. At 300°C, a Cu content of ~ 0.35 to 0.40 at% has been reported.[69] Recently, deposition of CVD Al with simultaneous Cu doping using dimethyl aluminum hydride (DMAH) and cyclopentadienyl copper triethylphosphine (CpCuTEP) has been reported.[70] Cu content as high as 2.5% was achieved.

One disadvantage of CVD Al is that the utilization rate of the precursor is low since Al is a light element. In a TIBA molecule, less than 15% of the molecular weight is from the Al. In practice, only a fraction of the precursor reaches the wafer surface and, consequently, a sizable amount of used precursor must be reprocessed or disposed of, which adds to the cost of the process.

The fire and explosion hazard of the precursor material for CVD Al is another issue that must be addressed. TIBA is pyrophoric (self-igniting) in air and reacts violently with water. Other precursors are generally more hazardous than TIBA. One should note that Al alkyls have been in use by the polymer industry for a number of years, and standard industrial safety and hygiene procedures are readily available. However, semiconductor manufacturers do not have this experience yet.

Even though the process and the equipment development of CVD Al are immature compared with CVD W, the advantage of lower cost and better reliability still makes CVD Al desirable. CVD Al is not vulnerable to corrosion and does not require elaborate cladding. Al CVD can be deposited at low temperature ($\leq$ 250°C) and is compatible with low-permittivity dielectric materials. Thus, interest in CVD Al continues.[70,71,72]

CVD $TaSi_2$

Sputtered $TaSi_2$ is used for polycide gate structure.[73] CVD $TaSi_2$ is a lower-cost replacement but has not been used widely because of its complex CVD behavior. CVD $TaSi_2$ can be deposited at 600°C using the silane reduction of $TaCl_5$:[74]

$$TaCl_5 + 2SiH_4 \rightarrow TaSi_2 + 5HCl + 1.5H_2 \tag{8.20}$$

A competitive reaction to Eq. (8.20) is

$$5TaCl_5 + 3SiH_4 + 6.5H_2 \rightarrow Ta_5Si_3 + 25HCl \tag{8.21}$$

During LPCVD, Ta_5Si_3 is the more favored reaction. $TaSi_2$ is formed only at the $TaSi_x$/polysilicon interface where an extra supply of Si is available. In the presence of Si, the Ta-rich silicide converts to $TaSi_2$ through

$$Ta_5Si_3 + 7Si \rightarrow 5TaSi_2 \tag{8.22}$$

$TaCl_5$, however, reacts rapidly with Si through the displacement reaction

$$4TaCl_5 + 13Si \rightarrow 4TaSi_2 + 5SiCl_4 \tag{8.23}$$

This reaction is self-limiting at 150 to 250 nm when Si diffusion through the $TaSi_2$ is insufficient to support the reaction. The displacement reaction, however, deposits $TaSi_2$ nonuniformly, resulting in a jagged[74] silicide/polysilicon interface, which causes gate oxide breakdown. Therefore, in practice, a polysilicon layer of

Micron Ex. 1029, p. 421
Micron v. YMTC
IPR2024-00911

$\sim$ 150 nm is deposited first, followed by the silane reduction process. Finally, an anneal at $> 600°C$ is needed to convert the Ta-rich silicide to $TaSi_x$ with $x \approx 2.2$.

LPCVD of $TaSi_2$ can also be achieved by dichlorosilane reduction[75] at 650°C, that is,

$$TaCl_5 + 2SiH_2Cl_2 + 2.5H_2 \rightarrow TaSi_2 + 9HCl \tag{8.24}$$

This reaction suffers from the same Si displacement reaction as the silane reduction. Consequently, the same three-step process, Si deposition + reduction reaction + sintering, is needed.

CVD WSi_2

WSi_2 and $MoSi_2$ are used widely on top of gate polysilicon to form a low-resistance polycide gate. Both sputtered and CVD WSi_2 are used for ULSI device fabrication. CVD WSi_2 is readily deposited using the silane reduction of WF_6 at 300 to 400°C,[25,26,76,77]

$$WF_6 + 2SiH_4 \rightarrow WSi_2 + 6HF + H_2 \tag{8.25}$$

A comparison of Eq. (8.25) with Eqs. (8.3) and (8.4), however, immediately reveals the ambiguity of the chemical reactions since the reactants are the same. In a CVD reactor, the flow rates of WF_6 and SiH_4 control the outcome of the reaction. As can be expected, a higher SiH_4/WF_6 ratio results in WSi_2 deposition. WSi_2 does not form at a SiH_4/WF_6 ratio below 3. Instead, W with Si incorporated in the grain boundaries is deposited at low flow rates.[76] In practice, a SiH_4/WF_6 ratio greater than 10 is used to ensure the deposition of $WSi_x (x \approx 2.2–2.6)$. Although a displacement reaction similar to $TaSi_2$ [Eq. (8.23)] also occurs, it is self-limiting at < 20 nm because of the lower deposition temperature for WSi_2. Therefore, the differences in chemistry and the deposition temperature make CVD WSi_2 more desirable than $TaSi_2$.

Unlike CVD W, which is used for plug/interconnect, WSi_2 is used on gate polysilicon, which has to endure high temperatures (800 to 1000°C) during source/drain drive and reoxidation. CVD WSi_2 films have high stress ($\sim$ 1000 MPa),[77] and silane reduction produces CVD films with relatively poor step coverage. During high-temperature processes, peeling and cracking of the WSi_2 at step edges have been a problem. In addition, F incorporation at high concentrations of 10^{20} atoms/cm^3 is common. The F diffuses to the gate oxide and causes flat band shifts. Therefore, even though silane-reduced WSi_2 has been in manufacturing for some time, a better process is needed.

WSi_2 can be deposited by dichlorosilane (SiH_2Cl_2, or DCS) reduction at higher temperatures (500 to 600°C)

$$WF_6 + 3.5SiH_2Cl_2 \rightarrow WSi_2 + 1.5SiF_4 + 7HCl \tag{8.26}$$

$$WF_6 + 3.5SiH_2Cl_2 \rightarrow WSi_2 + 1.5SiCl_4 + 6HF + HCl \tag{8.27}$$

Depending on the surface reaction kinetics, SiF_xCl_y, Cl_2, and F_2 are all possible reaction products.

WSi_2 deposited by dichlorosilane reduction[78,79] contains much less F (10^{18} atoms/cm^3) than from silane reduction, and the Cl content is also low (10^{18} atoms/

Micron Ex. 1029, p. 422
Micron v. YMTC
IPR2024-00911

cm^3). The resistivity and the film stress are comparable to that from silane reduction. However, it has a deposition rate that is three to five times as high and a better step coverage. As a result the DCS process reduces the peeling and cracking problems at the step edge as well as the device instability problems. Therefore, DCS will gradually replace silane in CVD WSi_2.

CVD $TiSi_2$

$TiSi_2$ has a resistivity three to four times lower than WSi_2 and $TaSi_2$ and is the most desirable for polycide. Because of the advantage of self-aligned silicide (salicide), CVD $TiSi_2$ has not been widely regarded as important. However, self-aligned structures are reaching a limitation in thickness and thermal stability (see Section 8.3.2). The CVD process, especially the selective deposition on the source/drain and the gate, may become important for future ULSI devices. $TiSi_2$ can be deposited by LPCVD using $TiCl_4$ and silane reduction[80] at 650 to 700°C:

$$TiCl_4 + 2SiH_4 \rightarrow TiSi_2 + 4HCl + 2H_2 \tag{8.28}$$

Excessive Si substrate consumption (300 to 600 nm), however, occurs through the displacement reaction:

$$TiCl_4 + 2Si + 2H_2 \rightarrow TiSi_2 + 4HCl \tag{8.29}$$

The displacement reaction is controlled by lowering the deposition temperature to < 600°C (to slow the diffusion) and by increasing the silane flow. Lowering the deposition temperature, however, lowers the deposition rate too. In addition, film properties degrade at lower temperatures. By employing plasma enhancement (PECVD), low-resistivity (14 to 25 $\mu\Omega$-cm) $TiSi_2$ has been deposited[81] at 590°C. In PECVD, the reaction is probably more complex than suggested by Eqs. (8.28) and (8.29). Spectroscopic studies[82] detected HCl, SiH_3Cl, SiH_2Cl_2, and SiCl as by-products.

Selective deposition of $TiSi_2$ follows essentially the chemistry in Eq. (8.29), and observations[83] confirmed the Si consumption from the displacement reaction. Thermodynamic calculations[84] show that Si consumption can be minimized at the expense of lower selectivity. Recently, selective $TiSi_2$ deposition using a high silane flow[85] to reduce the Si consumption has been reported. In this approach, the displacement reaction [Eq. (8.29)] provides the selectivity but the silane reduction reaction [Eq. (8.28)] checks the rate. Consequently, $TiSi_2$ could be selectively deposited without excessive Si etching.

8.2.4 Electro- and Electroless Plating

Electro- and electroless plating have been used extensively in printed circuit board and multichip module applications but have not been applied in manufacturing ULSI chips. Plating provides a low-cost method for depositing thick metal layers over large areas and is ideal for printing circuit boards and for multilayer board fabrication. Because of their tolerance to higher temperature, on-chip interconnects have been made exclusively by PVD and CVD processes. For future ULSI devices, however, low-temperature, low-cost processes compatible with low-permittivity dielectric

Micron Ex. 1029, p. 423
Micron v. YMTC
IPR2024-00911

materials are important. Plating of metal (especially Cu and barrier materials) is an ideal low-cost process for this application.

Selective electroless Ni plating using nickel sulfate for submicrometer contact plugs by hypophosphite reduction has been reported:[86]

$$Ni^{2+} + (H_2PO_2)^- + H_2O \rightarrow Ni^o + 2H^+ + H(HPO_3)^- \qquad (8.30)$$

Electroless plating of Cu for on-chip interconnects has been reported.[87] A Ti underlayer was used to provide a conducting surface, and the Ti was activated by Pd deposition in $PdCl_2$ solution before the Cu deposition. Cu can be deposited using $CuSO_4$ and formaldehyde reduction:[88]

$$Cu^{2+} + 2HCHO + 4OH^- \rightarrow Cu^o + 2HCOO^- + 2H_2O + H_2 \qquad (8.31)$$

The electroless-deposited Cu has resistivity between 2.0 and 2.5 $\mu\Omega$-cm, which is lower than Al (2.7 $\mu\Omega$-cm). After a 400°C anneal, the resistivity drops[87] further by about 10%. Electroless Cu with resistivity close to bulk (< 1.9 $\mu\Omega$-cm compared to 1.7 $\mu\Omega$-cm) has been reported.[89]

8.3 SILICIDE PROCESS

Si forms many stable metallic and semiconducting compounds with metals. Several metal silicides show low resistivity and high thermal stability, making them suitable for ULSI application. The formation kinetics, film properties, and potential application have been thoroughly examined by Murarka[90] and by Nicolet and Lau.[91] In this section we address some specific applications and discuss the issues regarding the use and choice of silicides.

Silicides can be formed by several means: (1) sputtering and sinter, (2) CVD, and (3) self-alignment. The first process is not selective nor self-aligned, and is used only for gate silicide. The last process is self-aligned and can be used for source/drain/gate or for source/drain only.

8.3.1 Polycide Gate Structure

Polycide is the condensed name for polysilicon-silicide, which is widely used for gate structures that contain a non-self-aligned silicide. In this structure, a silicide is blanket-deposited over the polysilicon and the composite is patterned together. A well-designed polycide gate can achieve a gate sheet resistance $\leq 10\ \Omega/\square$ compared to that for polysilicon of $\geq 40\ \Omega/\square$. Because of the considerably lower resistivity, this poly structure can be used for local routing over short distances. This is especially beneficial when only one or two levels of metals are used in a circuit, since the poly routing can help to reduce the die size by as much as 20%. In fact, the invention of polycide structure effectively delayed the need for multilevel metallization by several years.

A typical polycide formation sequence is shown in Fig. 11*a*. For sputter deposition, a high-temperature, high-pressure sintered, high-purity compound target is

Micron Ex. 1029, p. 424
Micron v. YMTC
IPR2024-00911

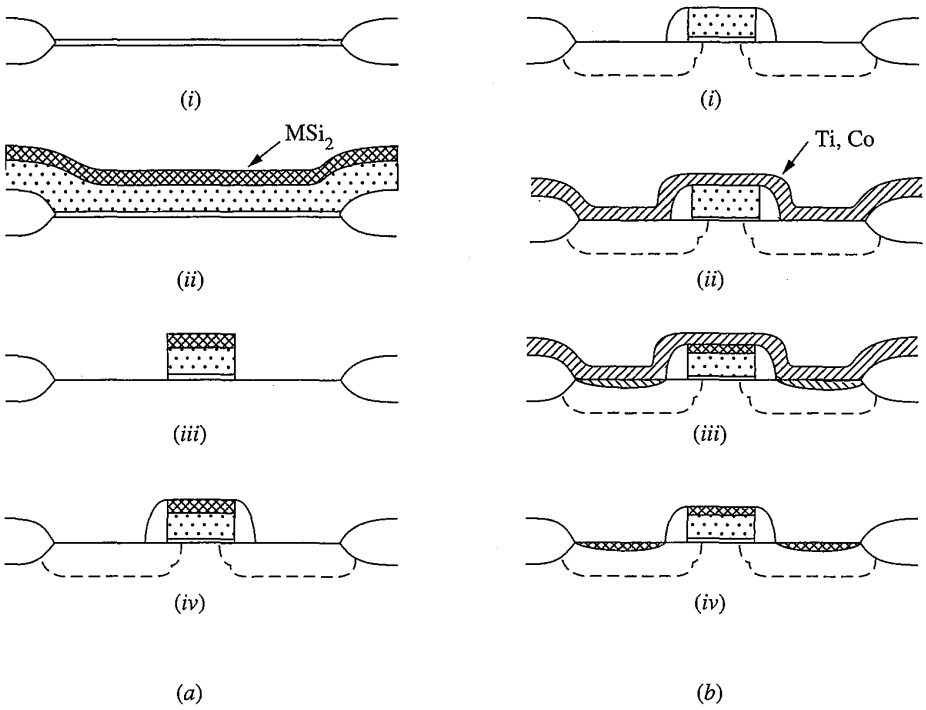


FIGURE 11
Polycide and salicide processes: (*a*) Polycide structure: (*i*) gate oxide; (*ii*) polysilicon and silicide deposition; (*iii*) pattern polycide; (*iv*) lightly doped drain (LDD) implant, sidewall formation, and S/D implant. (*b*) Salicide structure: (*i*) gate patterning (polysilicon only), LDD, sidewall, and S/D implant; (*ii*) metal (Ti, Co) deposition; (*iii*) anneal to form salicide; (*iv*) selective (wet) etch to remove unreacted metal.

used to ensure the quality of the silicide. The as-deposited film usually has a higher Si content in the form of MSi_x with $x \approx 2.4$ to 2.6 instead of 2.0. During the high-temperature sintering process the excess Si precipates out to the underlying poly-Si, reducing the silicide to a near stoichiometric composition with $x \approx 2.0$ to 2.2. Silicides with a lower Si content tend to oxidize badly during sintering, resulting in peeling and blistering problems. Silicides can also be deposited with a CVD process.

The most commonly used polycides are WSi_2,[26,92] $TaSi_2$,[73] and $MoSi_2$.[93] All are refractory, thermally stable, and resistant to processing chemicals. These properties make them better choices for polycide even though their electrical resistivity (50 to 100 $\mu\Omega$-cm) is higher than for $TiSi_2$ and $CoSi_2$.

$TiSi_2$ and $CoSi_2$ have the lowest resistivity among all silicides (10 to 15 $\mu\Omega$-cm); however, they are not good candidates for polycides. Both are vulnerable to processing chemicals, especially to HF solutions. The most difficult obstacle, however, is in polycide patterning. $TiSi_2$ is attacked by Cl-based chemistry during polysilicon etch, and this makes linewidth control difficult at the critical gate level. $CoSi_2$, on the other hand, lacks a volatile reaction by-product and is difficult to pattern by dry etching. $TiSi_2$ and $CoSi_2$, however, are ideal for self-aligned silicide or salicide application.

Micron Ex. 1029, p. 425
Micron v. YMTC
IPR2024-00911

8.3.2 Self-Aligned Silicide (Salicide)

A self-aligned silicide process is illustrated in Fig. 11*b*. In this process, the polysilicon gate is patterned without any silicide, and a sidewall spacer is formed to prevent shorting the gate to the source and drain during the silicidation process. A metal layer, either Ti or Co, is blanket-sputtered on the entire structure, followed by silicide sintering. Silicide is formed, in principle, only where the metal is in contact with Si. A wet chemical wash then rinses off the unreacted metal, leaving only the silicide. This technique eliminates the need to pattern the composite polycide gate structure and adds silicide to the source/drain area to reduce the contact resistance.

In practice, both $CoSi_2$ and $TiSi_2$ are formed by multistep processes:

$$Co + Si \rightarrow CoSi \qquad (450°C) \tag{8.32}$$

$$CoSi + Si \rightarrow CoSi_2 \qquad (700°C) \tag{8.33}$$

The selective etch by wet chemical is done after the formation of CoSi. If rapid thermal processing (RTP) is used for silicide formation, step 1 [Eq. (8.32)] can be skipped and $CoSi_2$ is formed directly at 700 to 750°C. The wet etch then is applied directly to $CoSi_2$.

To form $TiSi_2$, two RTPs in nitrogen are required:

$$Ti + 2Si \rightarrow TiSi_2 \quad (C\text{-}49) \tag{8.34}$$

$$TiSi_2(C\text{-}49) \rightarrow TiSi_2 \quad (C\text{-}54) \tag{8.35}$$

The first RTP is done at 620 to 680°C, and the second RTP is done at a temperature higher than 750°C. The second step [Eq. (8.35)] transforms the high-resistivity, metastable C-49 phase $TiSi_2$ to the low-energy low-resistivity orthogonal $TiSi_2$ (C-54 phase). The resistivity of C-49 $TiSi_2$ is about three to four times higher than that for C-54 $TiSi_2$. This phase transition occurs at 700 to 750°C. Direct heating of Ti in contact with Si to $\geq$ 750°C forms the stable C-54 $TiSi_2$ directly. However, severe shorting between the source/drain and gate occurs (see the discussion of processing issues later in this section).

Since $TiSi_2$ and $CoSi_2$ have the lowest resistivity, $\sim$ 10 to 15 $\mu\Omega$-cm, even a substantially thinner layer of $TiSi_2$ or $CoSi_2$ can provide a lower sheet resistance than polycide structures. For example, 50 nm of $TiSi_2$ has a sheet resistance of 2 to 3 $\Omega/\square$ whereas even 150 nm of WSi_2 can only give a sheet resistance of $\geq$ 6 $\Omega/\square$. Therefore, by using a self-aligned silicide, not only is the gate patterning simplified but also one can more fully use the benefit of a low-resistance poly-Si in circuit design.

Equally important is lowering the source/drain contact resistance. In most designs using 0.5 μm design rules, the source/drain contact resistance imposes no threat as a critical parasitic resistance. For wide transistors, multiple contacts are required to ensure that the entire width of the transistor is used. A self-aligned silicide can act as a shunt for the entire source/drain area, allowing a single contact and additional design freedom.

Micron Ex. 1029, p. 426
Micron v. YMTC
IPR2024-00911

Effects on transistor performance

A self-aligned silicide on the source/drain area lowers the contact resistance because the metal-Si contact area is increased from the area of the contact window(s) to the entire source/drain area. Consider the contact resistance of a nMOS transistor with a source/drain area of 1 μm × 1 μm and a contact window size of 0.4 μm. For the specific contact resistance, let us use 1×10^{-7} Ω-cm^2 for the metal/n$^+$ contact, and 0.2×10^{-8} Ω-cm^2 for the metal/metal contact. The contact resistance for the silicide/Si contact is

$$R_{c1} = 10^{-7}\ \Omega\text{-cm}^2/(10^{-8}\ \text{cm}^2) = 10\ \Omega$$

The contact resistance for the metal/silicide contact is

$$R_{c2} = 0.2 \times 10^{-8}\ \Omega\text{-cm}^2/(0.12 \times 10^{-8}\text{cm}^2) = 1.6\ \Omega$$

The total contact resistance is then $R_{c1} + R_{c2} = 11.6\ \Omega$.

If no silicide is formed in the source/drain, then only R_{c1} exists:

$$R_{c1} = 10^{-7}\ \Omega\text{-cm}^2/(0.12 \times 10^{-8}\ \text{cm}^2) = 80\ \Omega$$

Therefore, even for a nominal-size transistor, the self-aligned silicide significantly decreases the contact resistance.

Even 80 Ω of contact resistance is only a small fraction of the channel resistance. To reduce this quantity further does not improve the device performance. Let us consider, however, a wider transistor used to supply more current to subsequent stages. Assume that the transistor is 10 μm wide and that this nMOSFET is capable of delivering 4 mA of current at $V_g = 3.3$ V. Let us consider the effect of the R_c on the drain side and on the source side, individually:

1. The effect of R_c on the drain side. The effect of R_c on the drain side is to reduce the drain voltage V_d by the amount $R_c I_d$. If the transistor is operating in the saturated region (as in most digital circuits), then the reduction in V_d has little effect. If the transistor is operating in the linear region (as in analog circuits), then the output current I_d is reduced from

$$I_d = V_d/R_{\text{channel}} \tag{8.36}$$

to

$$I_d = (V_d - R_c I_d)/R_{\text{channel}} \tag{8.37}$$

or

$$I_d = V_d/(R_c + R_{\text{channel}}) \tag{8.38}$$

Therefore, the effective resistance in the linear region has been increased by the amount R_c. For this transistor, the channel resistance in the linear region, R_{channel}, can be calculated as V_{dsat}/I_d, or approximately 250 Ω. If only a single contact is made, the contribution of R_c, 80 Ω, is about 30%.

2. The effect of R_c on the source side. The effect of R_c on the source side is more severe. A resistance on the source side of the MOSFET not only decreases the

Micron Ex. 1029, p. 427
Micron v. YMTC
IPR2024-00911

drain voltage by the same amount as on the drain side but also raises the potential at the source by $R_c I_d$. The effective gate voltage V_g is then reduced by the same amount:

$$V_g' = V_g - R_c I_d \tag{8.39}$$

A reduction in effective gate voltage is equivalent to a reduction in transconductance:

$$g_m = \Delta I_d / \Delta V_g \tag{8.40}$$

$$g_m' = g_m / (1 + R_c g_m) \tag{8.41}$$

An nMOSFET with a $g_m = 300\ \mu S/\mu m$ and $R_c = 80\ \Omega$ has a $g_m' \approx 0.98 g_m$, or only a 2% loss in current drive for a 1-μm wide transistor; but for a 10-μm transistor, a 20% loss for the current drive results. Therefore, for wide transistors, either multiple contacts or a self-aligned silicide is needed to reduce the parasitic resistance.

The contact resistance will degrade the device performance more severely on further scaling of design rules. At a 0.25-μm design rule, the contact windows will be about 0.3 μm, giving a $R_c \approx 140\ \Omega$ compared with the 80 Ω in the example for 0.35-μm design rules. At a 0.18-μm design rule (corresponding to 1 Gbit DRAM), R_c will be ~250 Ω without self-aligned silicide. Even for a 1-μm wide transistor, the g_m' will be reduced to only about $0.9 g_m$. If self-aligned silicide is used, R_c drops to about 18 Ω in the 0.25-μm design and to about 25 Ω in the 0.18-μm design.

Effects on layout area

Figure 12 illustrates the benefit of self-aligned silicide in reducing layout area. Consider the current distribution along a wide transistor with no silicide in the source/drain area. Because of the need for shallow junctions, the diffusion resistance increases with each generation of scaling. As a consequence, at design rules of 0.35 μm or below, a diffusion R_{sh} of $< 100\ \Omega/\square$ is not achievable for n^+ diffusion. Consider the configuration in Fig. 12*a*, where two contact windows at one end of a wide transistor are used. The current distribution within the transistor is not uniform, and most of the current is crowded into the lower half of the transistor, as shown in Fig. 12*b*. This observation results in a current reduction of approximately 20 to 30%. The higher the diffusion R_{sh}, the more severe the current crowding becomes. If a self-aligned silicide is used, then the current crowding is avoided since the diffusion R_{sh} is largely replaced by the low silicide R_s.

The current crowding problem can be avoided by using multiple contacts. As shown in Fig. 12*d*, strategically placed contacts can reduce the current crowding. The use of more contacts, however, requires metallization to connect these contacts. The result is that valuable routing area is used for source/drain contacts, requiring a larger die size. This penalty can be reduced when more levels of metallization are used, since more degrees of design freedom are available to minimize the area penalty.

Processing issues

There are three important processing issues in self-aligned silicides: (1) the forming of a bridge between gate and source/drain; (2) the difficulty in forming

Micron Ex. 1029, p. 428
Micron v. YMTC
IPR2024-00911

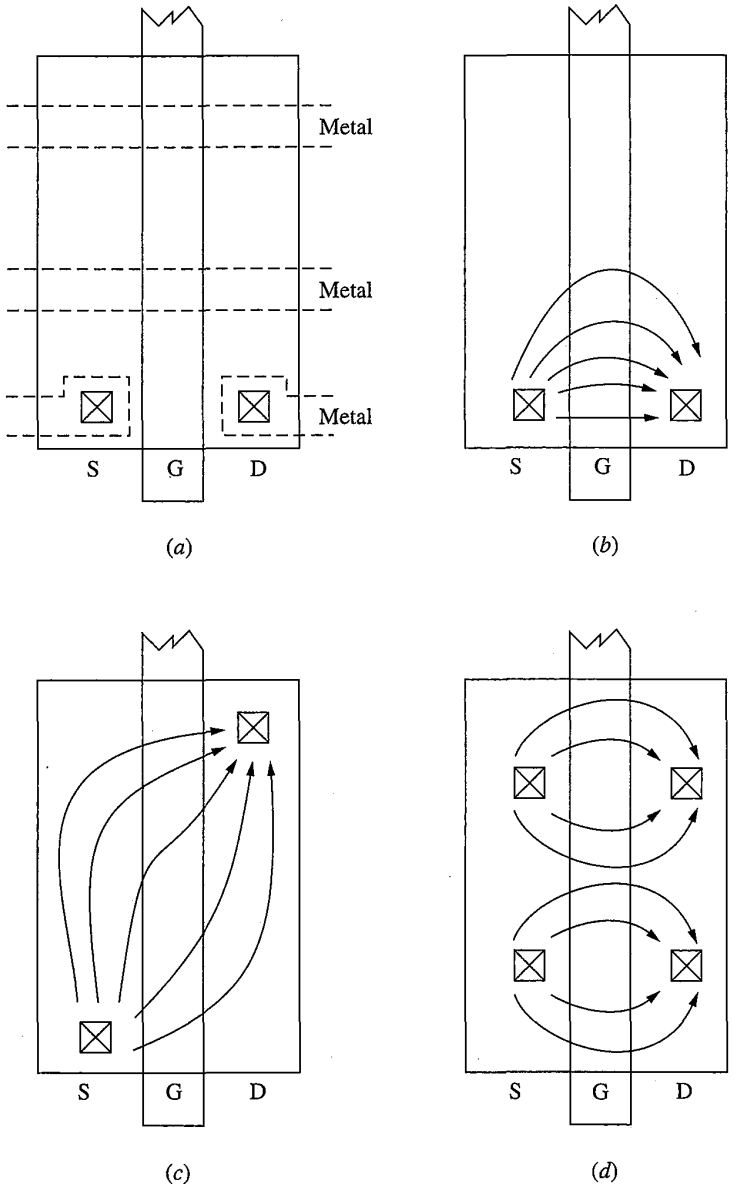


FIGURE 12
Effects of salicide on layout area: (*a*) If only one contact is made in the S/D, then the rest of the area can be used for routing; (*b*) if the S/D is not silicided, then current crowding occurs on the lower half of the transistor; (*c*) a different contact arrangement reduces the current crowding but takes away some routing area; (*d*) multiple contacts allow the use of full transistor width but leave no room for other metal runners over the transistor. If the S/D is silicided, then no current crowding will occur and only one contact, as in (*a*), is needed.

Micron Ex. 1029, p. 429
Micron v. YMTC
IPR2024-00911

silicide on thin strips of polysilicon and the diffusion region; and (3) the thermal stability of thin silicides.

The forming of a bridging short between the gate and source/drain is most apparent for $TiSi_2$. Figure 13 illustrates the formation of $TiSi_2$ along the sidewall of the gate, resulting in a short between the gate and the source/drain. The mechanism for this is "Si pumping" during the formation of $TiSi_2$. During silicide formation of

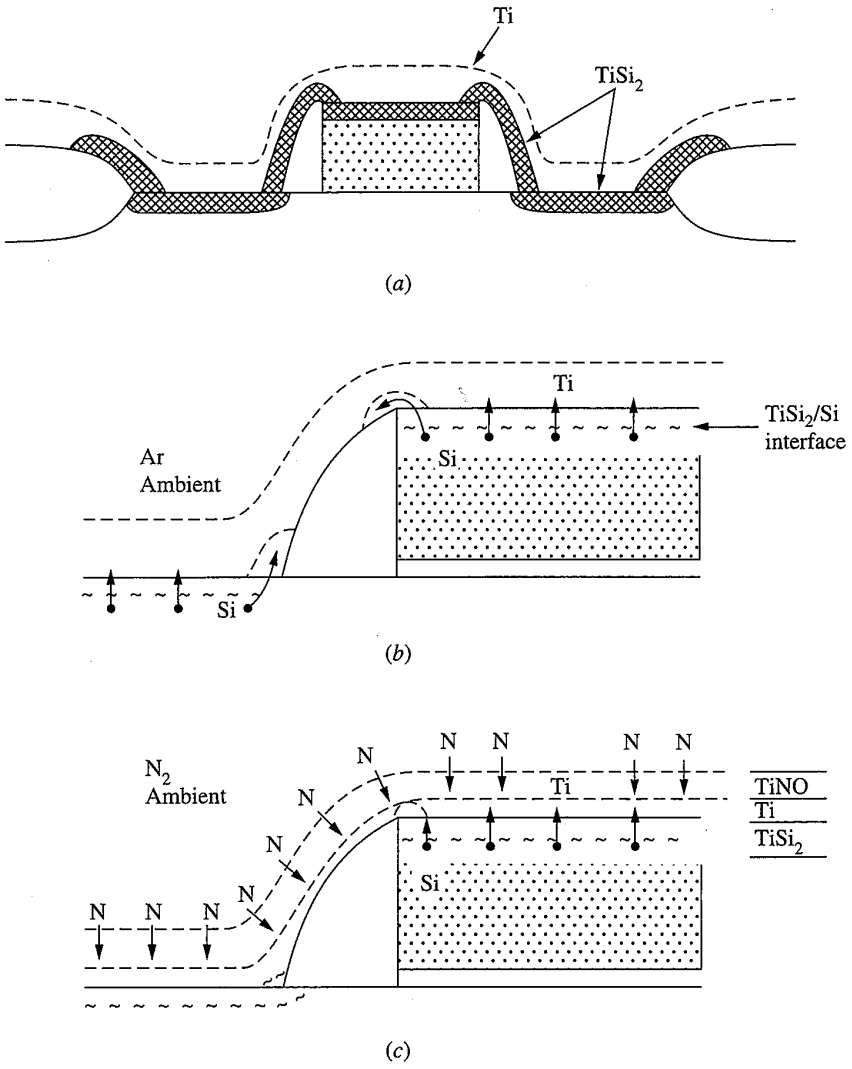


FIGURE 13
"Bridging" during the formation of Ti salicide: (*a*) $TiSi_2$ bridges over sidewall spacer and a short circuit is formed between S/G and D/G; (*b*) during $TiSi_2$ formation, Si diffuses into Ti to form $TiSi_2$; under favorable conditions Si can diffuse a long distance and form $TiSi_2$ on the sidewall spacer; (*c*) in a nitrogen atmosphere, nitrogen diffuses into Ti and blocks the diffusion of Si, reducing or eliminating bridging; TiN and/or Ti–O–N are formed on top of $TiSi_2$.

Micron Ex. 1029, p. 430
Micron v. YMTC
IPR2024-00911

$TiSi_2$, Si is the moving species, which diffuses into the Ti metal, as shown in Fig. 13*b*. As a result, $TiSi_2$ formation is no longer confined to where the Ti metal contacts Si. Naturally, the extent of bridge formation is determined by the temperature and time of the sintering process. Unfortunately, at least 600°C is needed to cause Ti to react with Si for $TiSi_2$ sintering and at this temperature extensive diffusion of Si in Ti already occurs. This problem can be reduced by processing in a nitrogen ambient and using RTP.[94–96] Within a finite temperature range between 620 and 680°C, nitrogen can diffuse into the Ti metal before extensive diffusion of Si occurs and stop the "Si pumping," as depicted in Fig. 13*c*. At temperatures higher than 680°C, Si diffuses rapidly in Ti and bridging occurs.

For $CoSi_2$, the diffusing species during silicide formation is alternately Co and Si. As a result, the problem of bridge formation is much less severe. $CoSi_2$ can be formed by using either a furnace in a reducing atmosphere[97] or by RTP.[98]

At a polysilicon line width below about 0.5 μm, the formation of $TiSi_2$ becomes difficult,[99,100] resulting in high poly resistance. The transformation of the high-resistance C-49 phase $TiSi_2$ to the low-resistance C-54 phase $TiSi_2$ was found to be the limiting factor.[99] After the first RTP, small-grained (0.1 to 0.2 μm) C-49 $TiSi_2$ was formed on both wide (> 10 μm) and narrow (0.4 μm) polysilicon lines. After a 725°C anneal, the $TiSi_2$ on the wide polysilicon lines transformed to large-grained (1 to 10 μm) C-54 phase. However, on narrow lines only smaller-grained (< 1 μm) C-54 $TiSi_2$ was observed, and only on a few lines. At a higher-temperature (775°C) anneal, C-54 phase was formed on both wide and narrow lines. At 800°C, $TiSi_2$ started to agglomerate, and the line resistance rapidly degraded.

An explanation[99] for the difficulty of the C-49 to C-54 phase transformation can be constructed based on the limitation of grain size on narrow lines. Generally, thin films do not grow into large grains several times the width of the line. Large-grained structures (such as Al) can be obtained by growing a large-grained thin film and then patterning it into lines. In the case of $TiSi_2$ the polysilicon lines are already defined. The nucleation of C-54 phase at 700 to 750°C is slow, and the density of nuclei is low. On wide lines, the nuclei grow into large grains, and only a few nuclei are needed to cover the entire line with large-grain C-54 $TiSi_2$. On narrow lines, however, the grain size is limited by the linewidth, and only a fraction of the line is covered by the C-54 $TiSi_2$ grown from the few nuclei available. This problem can be solved by heating the $TiSi_2$ to > 750°C to accelerate the nucleation of the C-54 phase. Unfortunately, thin Ti silicide agglomerates at > 750°C, rendering this solution unusable.

Several efforts have been made to reduce the narrow-line effects.[100–102] Preamorphization of the polysilicon[100,101] was used to improve the process margin. Selective W was used as a strapping[102,103] over $TiSi_2$ to improve the resistivity. $CoSi_2$ was proposed to replace $TiSi_2$ for very fine lines,[104,105] since $CoSi_2$ transformation occurs at a lower temperature (600 to 700°C) and thus full formation of low-resistance $CoSi_2$ is achieved before the silicide agglomerates.

Because of these processing issues, the use of salicide for 0.25-μm and smaller devices requires considerable process development and in some cases new processes have to be used. The salicide process, however, delivers unique advantages for logic and MPU (microprocessor unit) circuits at low cost. Continued scaling of devices will demand even more development work for this difficult process.

Micron Ex. 1029, p. 431
Micron v. YMTC
IPR2024-00911

8.3.3 Local Interconnect

Local interconnects are short polysilicon, polycide, or metal wires that connect adjacent devices to make simple functional groups such as logic gates, inverters, flip-flops, etc. Often, these short wires only connect part of the circuit to help handle cross-over of wiring. Figure 14 illustrates the use of a polysilicon line to connect the drains of the driving transistors to the gates of the opposite transistors in a SRAM cell. In this particular case, low resistance is not required and high resistance polysilicon can be used. For memory circuits such as the SRAM, more than one layer of polysilicon is often used as a resistive load and for local interconnect. This type of application is called *multi-poly technology.* Sometimes both multi-poly and multilevel metal are used with the poly for local connections and with metal (Al) for longer (global) interconnects.

For logic and microprocessors using a $TiSi_2$ salicide structure, one can take advantage of the TiN formed during the silicidation process and pattern it for use as a local interconnect.[106] As shown in Fig. 15, instead of the wet selective etch, the local interconnect area is masked and dry or dry/wet etched to remove only the TiN and unreacted Ti from unmasked areas. This process forms the salicide and the local interconnect in a single deposition/etching step. Good selectivity of TiN relative to $TiSi_2$ is the necessary condition for this approach.

TiN local interconnect can also be used with $CoSi_2$ salicide, although through a more complex process. After the salicide process is completed, a TiN layer is deposited and masked. A dry etching of the TiN completes the patterning of the local interconnect. This is a higher-cost process, since the TiN has to be deposited separately. The advantage is that good selectivity between TiN and $CoSi_2$ is guaranteed. Other silicide, polycide, or metal (such as W) can also be used in the same way as long as selective etching can be done.[107,108]

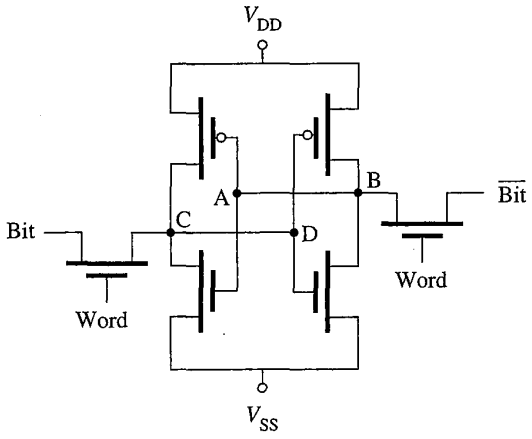
Local interconnect has the advantage of adding a (limited) degree of design freedom at relatively low cost. The process described in Fig. 15 requires only one mask to acquire this design freedom; no window mask is needed!

8.3.4 Silicide and Shallow Source/Drain Junctions

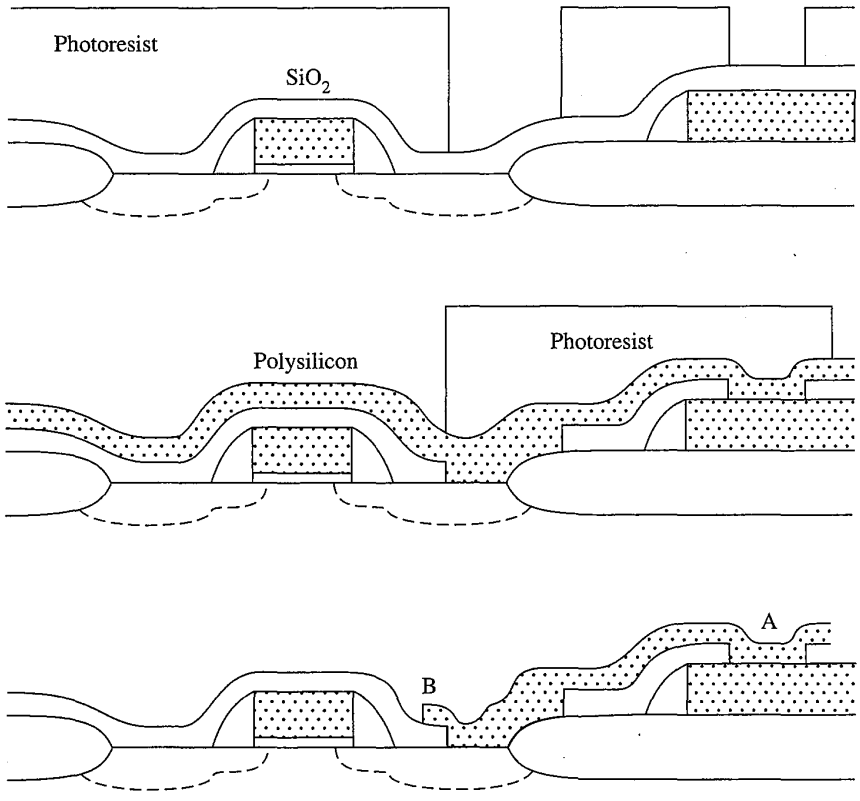
A severe disadvantage of a salicide in the source/drain region is the junction leakage caused by silicide. To form a silicide, a substantial portion of the Si from the junction is consumed. For example, to form 50 nm of $TiSi_2$ requires the consumption of approximately 50 nm of Si from the junction. For a shallow junction about 0.15-μm deep this represents one-third of the junction depth if the silicide is uniformly formed. Silicide tends to be not totally uniform, and even a minor nonuniformity (say, 20 nm) means that in some places 70 nm out of the 150-nm junction are consumed, leaving a very shallow effective junction.

In addition to the consumption of Si, dopants in the junction redistribute during silicide formation. This phenomenon[109] is especially damaging in $TiSi_2$. During the formation of silicide, because of the diffusion of Si into the Ti metal, a large number of vacancies are generated in the Si matrix, which enhances the diffusion of dopants

Micron Ex. 1029, p. 432
Micron v. YMTC
IPR2024-00911



(*a*)



(*b*)

FIGURE 14
Polysilicon local interconnect for SRAM: (*a*) A circuit diagram for a six-transistor CMOS SRAM cell; A, B and C, D are connected with polysilicon; (*b*) a two-mask process for implementing the polysilicon local interconnect: The first mask is used to open contact windows through a thin dielectric layer (interpoly dielectric or IPD) and the second mask is used to define the local interconnect.

Micron Ex. 1029, p. 433
Micron v. YMTC
IPR2024-00911

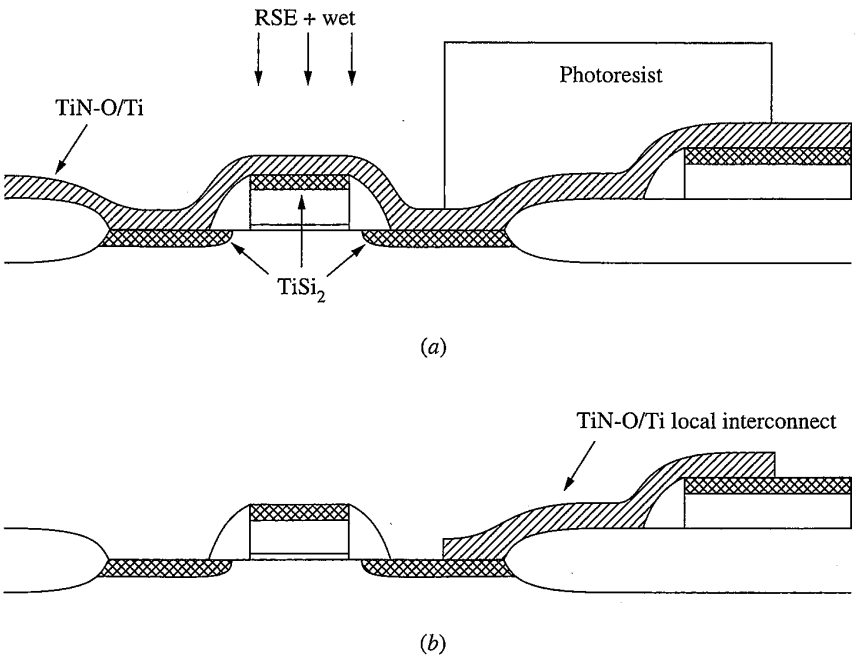


FIGURE 15
One-mask process for TiN local interconnect: (*a*) Save the Ti–O–N/Ti after $TiSi_2$ formation in N_2; (*b*) after photolithography and etching (dry + wet), a local interconnect is formed between the drain and the polysilicon of another device.

in the junction. In the first place, the Si containing the highest dopant concentration is consumed by the silicide. In addition, the enhanced diffusion absorbs more dopant from the remaining Si junction into the silicide, further depleting the junction of dopants, as shown in Fig. 16. The depleted junctions are more vulnerable to leakage,[110] and the low dopant concentration at the silicide/Si interface causes high contact resistance (see Fig. 4).

A technique to solve the junction leakage and dopant depletion problem is to implant the junction after the silicide is formed. To do this, the dopant has to be implanted either into the silicide or through the silicide. Good junctions formed by implanting into $CoSi_2$ and driving the dopants out by annealing have been demonstrated.[111] By using this technique, junctions as shallow as 0.12 μm can be made. For $TiSi_2$ the same process is difficult because of the very high affinity of As and B for $TiSi_2$; only very limited success has been reported.[112] It is difficult to form very shallow junctions ($\leq$ 0.15 μm) by implanting dopant through the silicide, since the peak of the implant has to be in the silicide and only the tail portion can be used. The junction depth then becomes a sharp function of the silicide thickness, which in the case of $TiSi_2$ is difficult to control precisely.[113]

Once the junction depth becomes less than $\sim$ 0.1 μm, further scaling (with salicide) becomes extremely difficult. Even when the junctions can be formed, the

Micron Ex. 1029, p. 434
Micron v. YMTC
IPR2024-00911

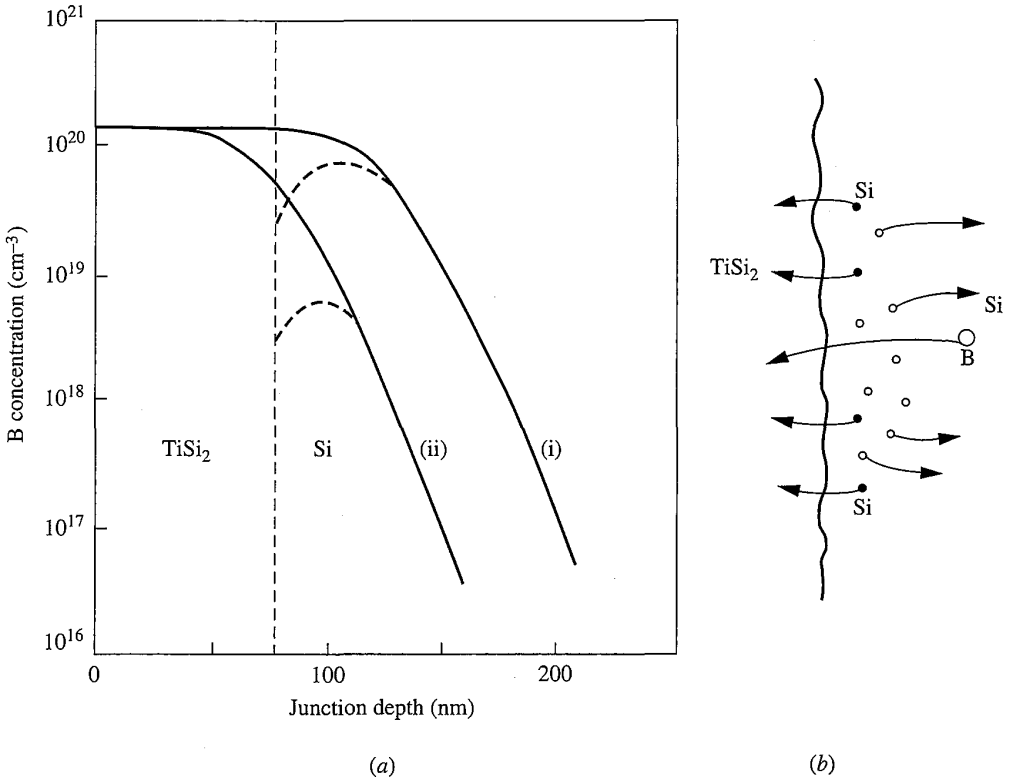


FIGURE 16
Redistribution of B during $TiSi_2$ formation: (*a*) B redistribution for 200-nm junction (*i*), and for 150-nm junction (*ii*). Solid curves are B concentration profiles before $TiSi_2$ and dashed curves are after $TiSi_2$. (*b*) Mechanism of enhanced diffusion during $TiSi_2$ formation: Si moves into Ti to form $TiSi_2$, leaving a vacancy behind; the vacancy diffuses rapidly into Si, enhancing the diffusivity of B. There is no back-diffusion flux from $TiSi_2$ even as the B concentration in $TiSi_2$ is higher because B binds tightly with Ti.

very shallow effective junction (< 0.05 μm) is vulnerable to other processing and environmental factors such as mechanical stress.[114] Eventually innovations, such as raised source and drain junctions, will be necessary to replace the conventional structures (see Section 8.7).

8.4
CVD W PLUG AND OTHER PLUG PROCESSES

One of the most difficult problems in metallization is to ensure enough metal continuity at contact windows and vias. The step coverage of sputtered Al decreases rapidly with increasing contact window aspect ratio, and at small design rules the step coverage at contacts and vias drops below 20%. No metal reliability can be guaranteed at this level of step coverage, and consequently various forms of metal plugs have been developed.

Micron Ex. 1029, p. 435
Micron v. YMTC
IPR2024-00911

8.4.1 Selective CVD W Plug

On a Si contact, the selective process starts from a Si reduction process as shown in Eq. (8.2). This process provides a nucleation layer of W grown on Si (but also consumes about 20 nm of Si) but not on SiO_2. The real W plug is grown by a hydrogen reduction process as shown in Eq. (8.1). The hydrogen reduction process deposits W rapidly on the nucleation layer, forming the plug. This process, however, does not have perfect selectivity, and as a result, spurious nucleation and W growth can occur on the SiO_2 (Fig. 17*a*). The mechanism of the loss of selectivity has been studied and reported numerous times,[115,116] but so far no foolproof technique has been identified to guarantee total selectivity.

Another factor that is unfavorable to a selective W plug is the difficulty in filling contact windows of different heights, as shown in Fig. 17*a*. Since the contact to the gate is always shallower (by an amount that equals the gate height + field oxide) than to the source/drain, selective W cannot fill both contact windows simultaneously. Therefore, selective W is suitable only for via contacts.

Various innovative techniques have been proposed to solve the loss of selectivity. The most promising deposits a nucleation layer such as TiN and then selectively removes this material from the SiO_2 area, leaving it only in the contacts,[117] as shown in Fig. 17*b*. This has several advantages: (1) It solves the different window height problem, because W grows from the sidewall as well as from the bottom of the contacts; (2) the selectivity loss is less severe, since now the plug grows from the sidewall and much thinner W is needed; and (3) the adhesion of the W plug to the contact is better because of growth from the sidewall. This process, however, shifts the difficulty of selective W deposition to the selective etching of TiN. Until a simple, low-cost selective W process is found, the majority of manufacturers will use the more expensive but more mature blanket W-plug process.

8.4.2 Blanket CVD W Plug

By depositing a metal nucleation layer on the entire wafer, CVD W can be blanket-deposited on the wafer and in the contact windows. The W on the SiO_2 is then etched off by RIE, leaving only the thicker W in the contact, as shown in Figs. 17*c* and 17*d*. Since this process relies on the removal of all CVD W except in the contacts, the uniformity of the W deposition and RIE etchback is critical for the control of this process. If this control is not exercised, then the W plug will be substantially recessed after the etchback, resulting in poor Al step coverage again.

The most commonly used nucleation layer is TiN. Although TiW also serves as a good nucleation layer, the RIE selectivity between TiW and W is poor, resulting in both localized etching along the sidewall and difficult end point detection. As a result, the process control of the plug process for TiW is inferior to that for TiN.

Usually, a silane reduction process [Eq. (8.2)] is used as the first step in blanket W deposition, to serve as a nucleation layer and also to reduce junction damage. After the silane reduction, hydrogen reduction as shown in Eq. (8.1) is used to grow the blanket W layer rapidly.

Micron Ex. 1029, p. 436
Micron v. YMTC
IPR2024-00911

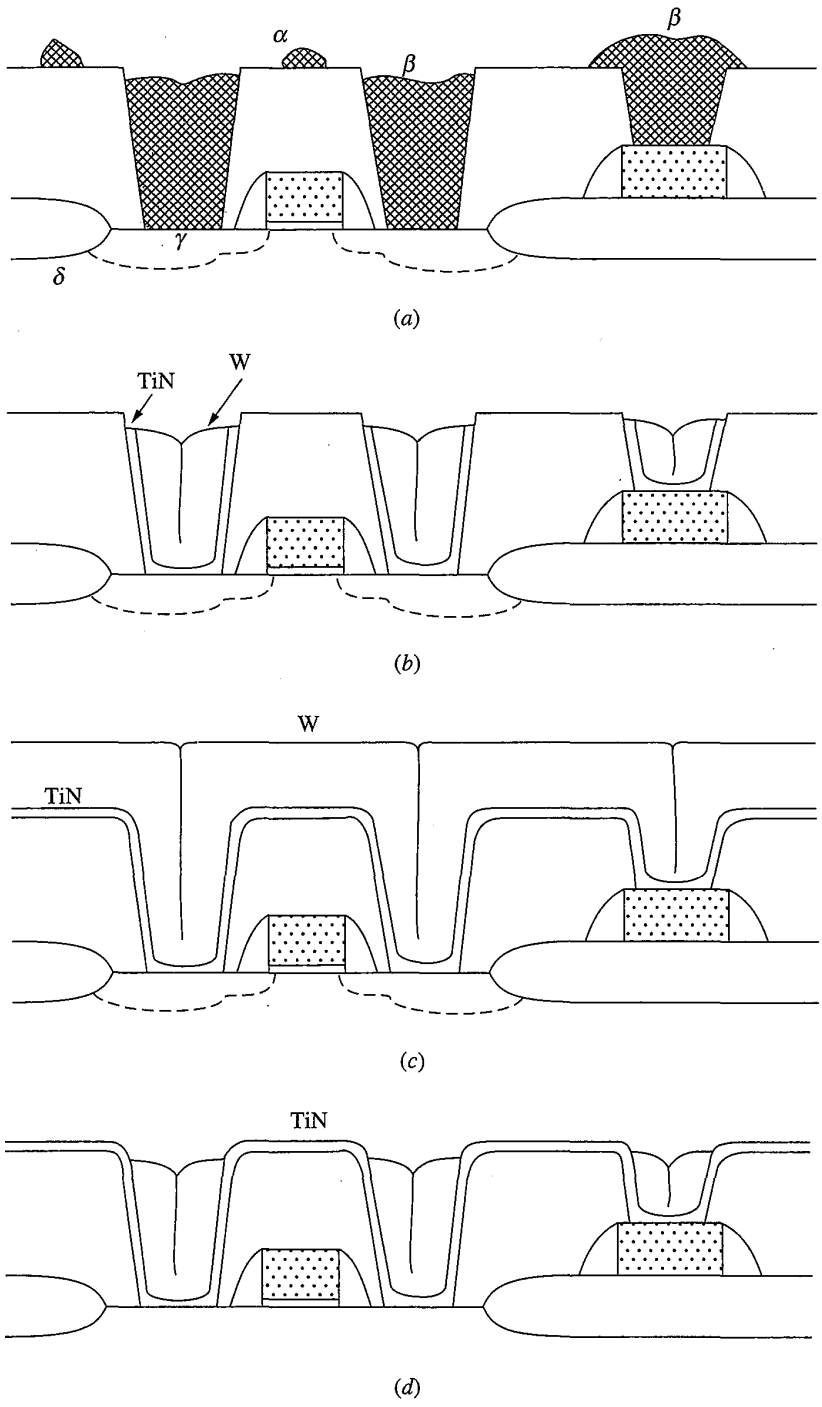


FIGURE 17
CVD W plug processes: (*a*) Selective W with no barrier layer: α: loss of selectivity; β: uneven fill; γ: junction leakage; δ: wormholes. (*b*) Selective W with barrier liner; does not have the problems shown in (*a*), but liners are hard to form. (*c*) Blanket deposition of W on TiN barrier. (*d*) After W etchback to form W plugs.

Micron Ex. 1029, p. 437
Micron v. YMTC
IPR2024-00911

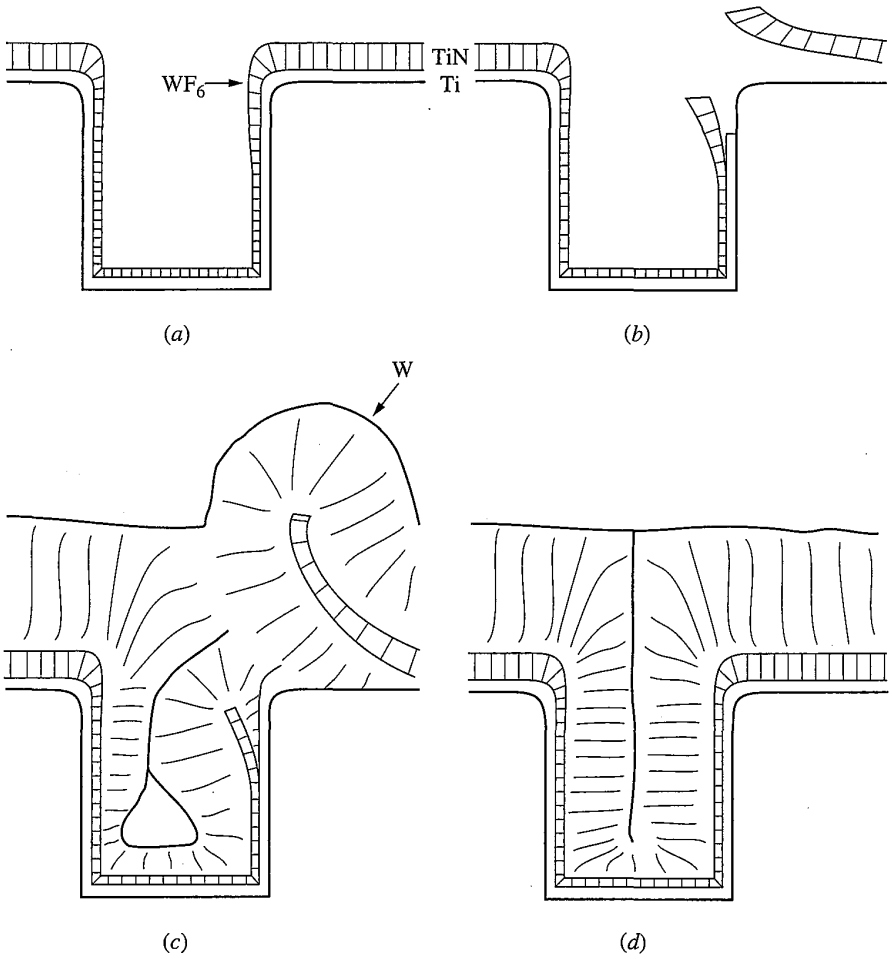


FIGURE 18
Tungsten volcano from barrier failure: (*a*) High stress and weak TiN at sharp corner attacked by WF_6; (*b*) Ti reacts with WF_6 and TiN peels back; (*c*) W deposits on both sides of peeled TiN, forming the volcano; (*d*) window with rounded corner stays intact.

TiN does not produce good contact resistance to n^+ and p^+ Si, and a thin layer of Ti ($\sim$ 30 to 50 nm) under the TiN is used. Ti, however, reacts readily with WF_6 (Table 1) and is more effective than H_2, Si, or SiH_4 in reducing WF_6. If there is any defect in the TiN film, or if the TiN is not properly "stuffed" with oxygen, then WF_6 can penetrate through the weak spots of the TiN film and react with the underlying Ti. Figure 18 shows the formation of "volcanoes" at the edge of contact windows. Sputter-deposited TiN has high stress, and the highest stress point is at the edge of a step (Fig. 18*a*). Because of the columnar growth of TiN, the material at the corner also is more porous and thus more vulnerable to WF_6 penetration. Once a reaction starts, the Ti is exposed to WF_6 and reaction proceeds rapidly. Losing support, the TiN peels off (Fig. 18*b*). Meanwhile, W nucleates on both sides of the peeling TiN (Fig. 18*c*) and grows into thick W films, forming a hump usually > 1 μm in size.

Micron Ex. 1029, p. 438
Micron v. YMTC
IPR2024-00911

Because these humps or "volcanoes" are large, the etchback RIE process cannot remove them completely, resulting in intra- and interlevel metal shorts.

The W volcano problem is more severe for smaller contact windows because higher stress develops at surfaces with larger curvatures. The barrier properties of the TiN are critical for the prevention of W volcanoes. The deposition process also plays a significant role, especially at the nucleation stage. Rounding the edge of the contact window and relieving the local stress (Fig. 18*d*) reduce the occurrence of W volcanoes.

The equipment for blanket CVD W deposition and etchback has been developed in the last decade, and this process is widely applied to both memory and logic circuits.

8.4.3 Aluminum Plug

Recently, reflow Al plugs formed at high temperatures have become popular. The concept of depositing Al and reflow by a high-temperature anneal by RTP[118] or by excimer laser[119] has been investigated before. Reflow by RTP causes junction damage because Al stays in the molten state for about one second and reacts with the barrier material. Heating the Al by an excimer laser, on the other hand, melts the Al in a microsecond and no visible damage to the barrier layer is done. However, Al is highly reflective even in the UV range; as a result, high fluence of laser light is needed to melt Al. Consequently, the laser dose required to fill contact holes is close to that which causes runaway laser ablation.[120] Antireflective coatings reduce the laser fluence needed for melting and increase the process margin.[121] However, the alloying of metals such as Ti into Al causes high resistivity, whereas Cu causes etching problems. Therefore, despite some reported success,[119] laser reflow of Al has not been widely used.

In situ reflow of Al by heating the wafer to a high ($> 500°C$) temperature for a few minutes after cold Al deposition has been reported.[122] Because of the high temperature, the barrier metal has to be robust; usually TiN is used. Figure 19*a* illustrates the formation of an Al plug using this reflow method. The driving force for the Al flow process is the reduction of surface area to minimize the surface energy. To completely fill contact windows with aspect ratio ≥ 1, the wafer needs to be heated to about 550°C for about 3 minutes. At this high temperature significant bulk diffusion of Al occurs. Since surface energy provides the driving force for Al flow, it is imperative that the Al surface remain oxide-free throughout the reflow process. High vacuum, on the order of 1×10^{-8} torr, with low residual oxygen and water vapor is necessary. Also, proper wetting of Al on the barrier layer is important to prevent the agglomeration of Al. Although TiN is a good barrier layer material, the wetting of Al on TiN is poor. Therefore, composite wetting/barrier layers such as Ti/TiN, $TiSi_2$/TiN, or Si/TiN are used to enhance the Al reflow.

Al plugs can also be formed by high-temperature deposition or by multistep deposition.[123–125] Figure 19*b* shows a multistep Al plug process. A thin Al layer is sputtered at low temperature to serve as a continuous nucleation layer. The majority

Micron Ex. 1029, p. 439
Micron v. YMTC
IPR2024-00911

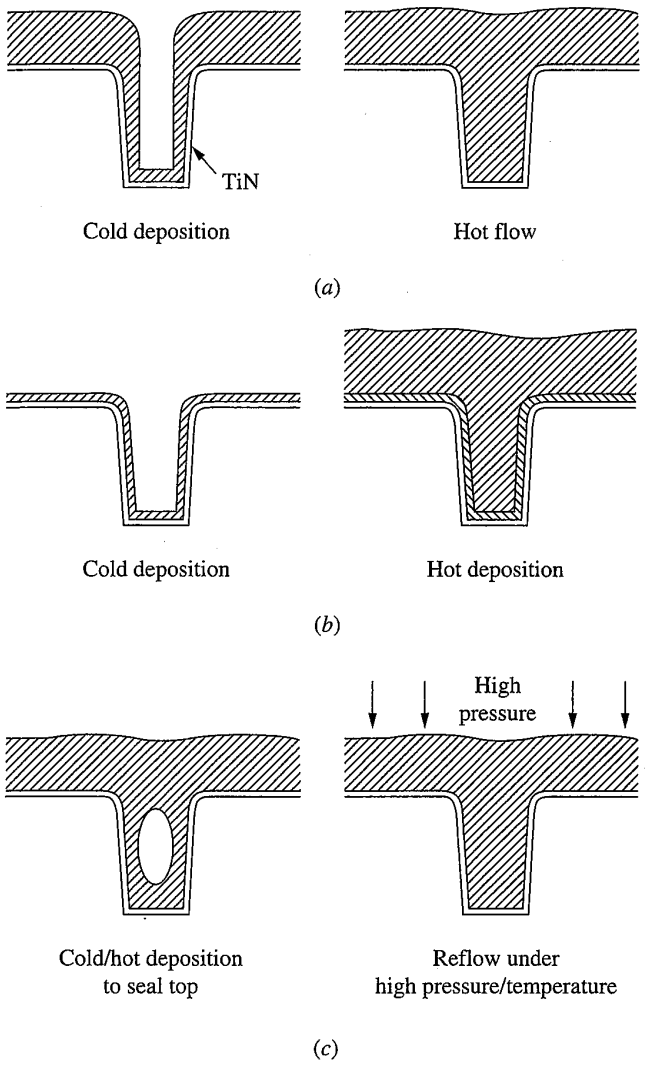


FIGURE 19
Reflow Al plug processes: (*a*) Cold deposition + hot reflow; (*b*) cold deposition + hot deposition; (*c*) high pressure reflow.

of the Al is then deposited at high temperature, 400 to 500°C. In this process, most of the mass transport occurs through surface diffusion. Therefore, the process can be done at a lower temperature (by about 50 to 100°C). The lower processing temperature puts a less stringent requirement on the barrier metal. However, because of the dependence on surface diffusion, the vacuum requirement is more demanding than for the Al reflow process.

For via applications the flow process needs to occur at < 450°C. This can be achieved by using special alloys. For example, for an Al-Ge alloy, the flow temperature can be reduced to below 400°C.[126] Ge, like Si, tends to precipitate out of Al during cooling and affects the contact resistance in small vias. Low-temperature Al flow is important for the opportunity it provides for low-cost fabrication of multilevel wiring. Recently, a low-temperature, high-pressure Al flow process[127] has

Micron Ex. 1029, p. 440
Micron v. YMTC
IPR2024-00911

been reported. In this process, a thick layer of Al is deposited first to seal the top of the contact vias. Ar at high pressure (600 MPa) is then introduced into the specially designed chamber where the wafer is held at 350 to 450°C. The combined effects of temperature (which makes Al soft) and pressure force a massive migration of Al, which flows and fills the vias, as shown in Fig. 19*c*.

Because of the dependence on surface and bulk diffusion, the flowed Al plug processes are sensitive to the shape and the aspect ratio of the contacts. The dependence on the wetting layer also causes processing problems, for example, the formation of $TiAl_3$ when Ti is used as the wetting layer and the formation of Si nodules when Si is used for wetting. The processing temperature range is small because Al does not flow if the wafer temperature is low, yet the barrier fails if the temperature is too high. Well-controlled and very uniform heating of the wafer is necessary. Therefore, although flowed Al plugs are used for the fabrication of some products at low cost, the technique has not reached maturity yet. This is still an area of intense R&D interest.

8.4.4 Other Metal Plug Processes

Contact plugs can be formed by other CVD processes such as CVD Al[20] and CVD Cu.[23,24] Both of these are still in their infancy. Polysilicon-based plug processes have been investigated. In this approach, polysilicon is blanket-deposited and then removed by RIE etchback to form a Si plug in the contact windows. A metal is deposited next and then annealed at low temperature to form a metal silicide with the Si plug. The excess metal is then stripped off by a selective chemical etch, and a silicide plug is left in the contact windows. By using this approach, silicide plugs of $CoSi_2$[128] and Ni_3Si[129] have been reported. Electroless plating of Co, CoW, and Ni to form selective plugs have been reported.[86] These processes, even though promising, have not been tested in large circuits to verify their compatibility with ULSI processing.

8.5 MULTILEVEL METALLIZATION

8.5.1 Introduction

The most difficult problems in multilevel metallization are (1) to deposit a void-free interlevel dielectric (ILD) that can fill the small gaps between metal lines without particle or reliability problems and (2) to smooth out the surface morphology and to reduce the accumulated topography to allow sufficient margin for the depth of focus for lithography. In this section we will focus in these two areas—ILD deposition and ILD planarization. We will also discuss the merit of more advanced architecture—stacked contacts, borderless contacts, and packaging limitations.

Micron Ex. 1029, p. 441
Micron v. YMTC
IPR2024-00911

8.5.2 Interlevel Dielectric (ILD) Deposition

Plasma-enhanced SiO_2

Even though high-quality CVD oxides can be deposited without plasma assistance at high temperatures (650 to 850°C), they are not suitable for ILD because high temperature is not compatible with Al wiring. Oxides deposited at lower temperatures (300 to 400°C) are porous and water-absorbing and, in general, unsuitable for ILD use. With plasma enhancement (PECVD), oxides can be deposited at $\leq$ 400°C with acceptable quality. The details of CVD oxide and PECVD deposition have been described in Chapter 5. here we will examine the merit of PECVD oxides for ILD.

A particularly important precursor for plasma oxide deposition is TEOS, or tetraethylorthosilicate (also called tetraethoxysilane) $Si(OC_2H_5)_4$, which without plasma dissociates pyrolytically at about 650 to 750°C. TEOS dissociates cleanly into SiO_2 and complex by-products[130] with little C residue in the oxide. The chemistry of TEOS dissociation follows roughly the reaction

$$Si(OC_2H_5)_4 \rightarrow SiO_2 + \text{by-products} \tag{8.42}$$

In the presence of plasma, TEOS can react with oxygen according to

$$Si(OC_2H_5)_4 + O_2 \rightarrow SiO_2 + \text{by-products} \tag{8.43}$$

In Eq. (8.43), oxygen is added to help the oxidation of TEOS. With plasma enhancement, good-quality oxide can be achieved at $<$ 400°C. What makes TEOS an important precursor for CVD oxide is its capacity to cover the steps conformally. The molecular structure of TEOS consists of a silicate (SiO_4) core with an ethyl group attached to the oxygen atoms at each corner of the tetrahedral silicate structure. This structure is particularly difficult to chemisorb on the surface because Si cannot reach the surface directly. Therefore, it forms weak bonds with the surface and is mobile there before chemical reaction occurs. (Alternative theories propose that fragments of TEOS are mobile on the surface.) Consequently, oxides deposited with a TEOS precursor have good step coverage.

Silane-based chemistry can also be used for plasma oxide deposition. Commonly used chemistries are reactions with N_2O or with O_2:

$$SiH_4 + 2N_2O \rightarrow SiO_2 + 2H_2 + 2N_2 \tag{8.44}$$

$$SiH_4 + O_2 \rightarrow SiO_2 + 2H_2 \tag{8.45}$$

Silane-based oxides have poor step coverage, contain a large amount of hydrogen, and are seldom used. However, with advanced high-density plasma sources excellent step coverage and film properties can be achieved with silane-based chemistry. (See Section 8.5.2 on high-density plasma deposition).

Although the step coverage of PETEOS is conformal, it has a tendency to build up at the upper corners of metal lines, as shown in Fig. 20*a*. This overhang is caused by the shadow effect of LPCVD. At low pressure, the mean free path of the precursor molecule is larger than the feature size. Consequently, the bombardment of the surface by the precursor molecules follows a line-of-sight path, which produces

Micron Ex. 1029, p. 442
Micron v. YMTC
IPR2024-00911

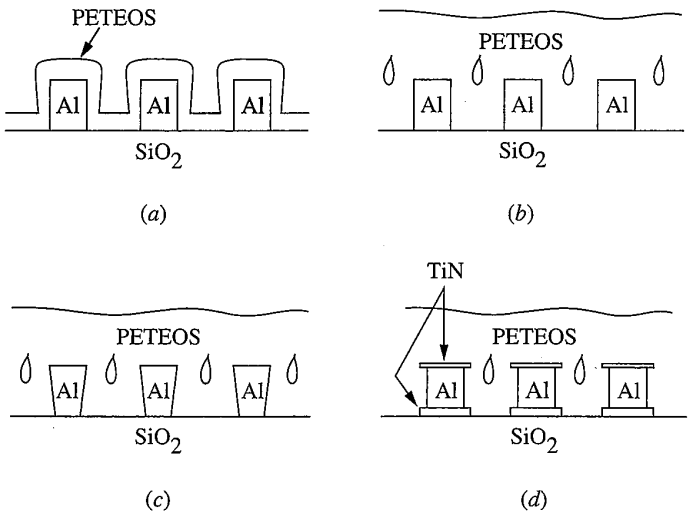


FIGURE 20
Keyholes in plasma-enhanced CVD oxide: (*a*) PECVD TEOS tends to build up at corner of metal lines; (*b*) keyhole voids; (*c*) re-entrant metal etching profiles cause larger keyholes even when the oxide is perfectly conformal; (*d*) metal etching profile control is more difficult for multilayered metal structures, resulting in keyholes.

the shadow effect. As a result, when PECVD TEOS is used to fill a small gap between two metal lines a void can form in the center, as shown in Fig. 20*b*. Note that even without the shadow effect and the overhang, metal profiles may be imperfect, as shown in Fig. 20*c*, which also results in voids. These are especially serious when multilayered metal structures for which uniform etching profiles are difficult to maintain are used (Fig. 20*d*). Voids in ILD not only pose a reliability concern, for fear of trapped chemicals in the void, but also may cause breaks in metal lines if a void is etched open during the planarization process, as shown in Fig. 21.

The task of depositing ILD without voids therefore cannot be solved by a process that is 100% conformal. Consequently, multistep processes using ion bombardment to round off corners to enhance the ILD filling capability have been developed.[131] As shown in Fig. 22, the multistep process (also called dep/etch/dep process) alternately deposits and etches the ILD to create a desired profile. The Ar etch step is used to cut the corners of the ILD and form a less severe geometry for the next PECVD process. Since a sidewall on the metal is created, this process is not sensitive to the etching profile of the metal lines. The multistep process, however, is more costly to fabricate because it requires multiple deposition and etching steps.

The repeated dep/etch/dep process can fill metal gaps with a height/space ratio (the aspect ratio, or AR) of about 1.0 without any voids. In principle the multistep process of deposition and etching can be tailored to fill gaps with higher aspect ratios, as shown in Fig. 23 By increasing the number of PECVD deposition and Ar etch steps, higher aspect ratio gaps can be filled. However, the cost for the ILD process also increases, with diminishing returns for each additional layer. Less expensive

Micron Ex. 1029, p. 443
Micron v. YMTC
IPR2024-00911

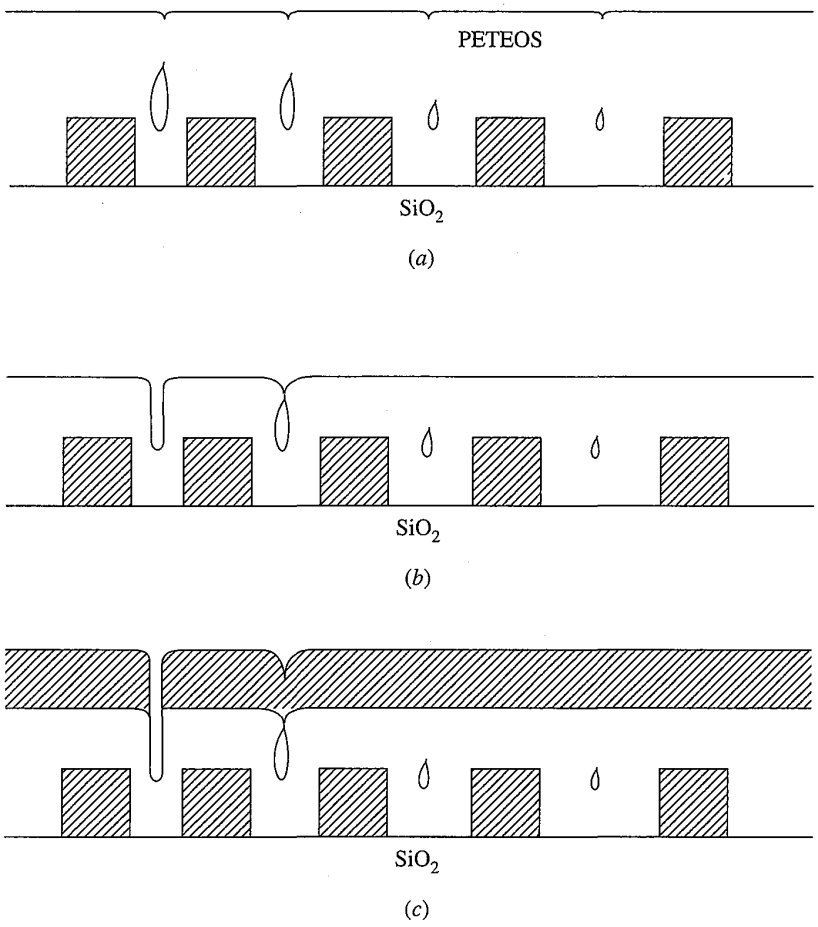


FIGURE 21
Effects of keyholes in ILD on metal step coverage: (*a*) ILD before planarization; (*b*) after ILD planarization, some keyholes open into trenches; (*c*) after metal deposition.

processes that can fill a wider range of gaps are more desirable. These will be discussed in the next four sections.

Atmospheric CVD TEOS/O_3 oxide

Silicon dioxide can be deposited by TEOS at low temperatures (150 to 400°C) with ozone[132] without plasma. Because the decomposition of TEOS is induced by O_3 and a plasma is not required, this process can be done at atmospheric or nearly atmospheric pressure to get better step coverage. At atmospheric pressure, the mean free path is $\leq 0.1\ \mu m$ and the deposition is conformal. The oxide can be deposited in a batch-type atmospheric pressure (APCVD) or in a single wafer or minibatch system at subatmospheric pressure (SACVD) for better process control. The chemistry of the two processes is the same.

The mechanism of TEOS/O_3 deposition is still poorly understood. Figure 24 shows a typical deposition rate vs. deposition temperature curve, and Fig. 25 shows

Micron Ex. 1029, p. 444
Micron v. YMTC
IPR2024-00911

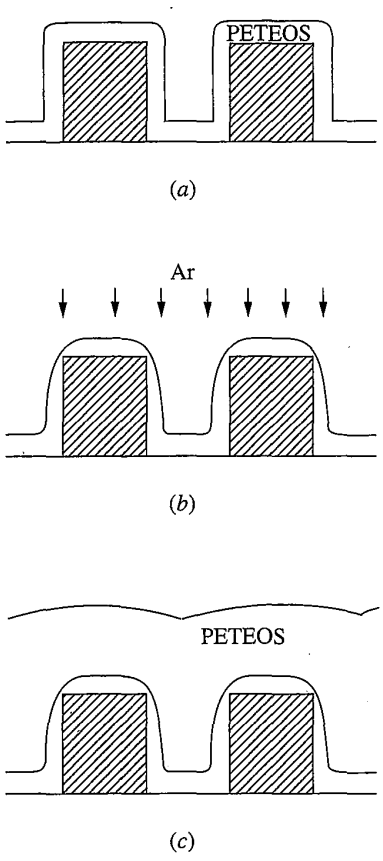


FIGURE 22
A multistep ILD process: (*a*) PECVD TEOS deposition; (*b*) Ar sputter etch to remove overhang; (*c*) new layer of PECVD TEOS to complete void-free gap fill.

a typical deposition rate vs. O_3 concentration curve. The deposition rate follows roughly an Arrhenius relation with $1/T$ at low temperature up to about 250°C, where it peaks and then declines at higher temperatures. The concentration of O_3 has a similar effect on the deposition rate. The deposition rate and the oxide properties are strong functions of the underlying Si surface.[133] For example, on thermal oxide

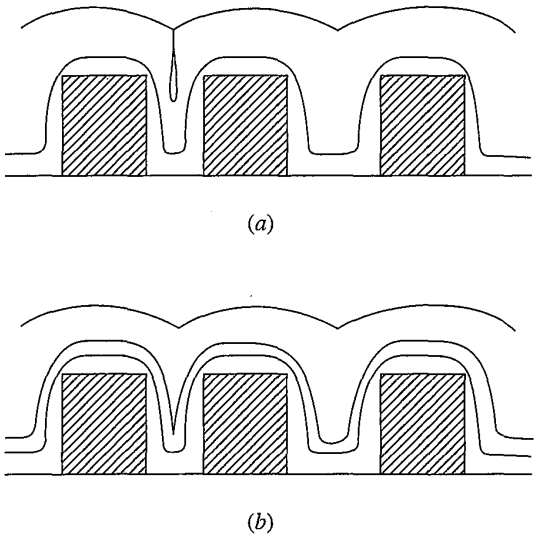


FIGURE 23
ILD gap-filling processes by repeated deposition and Ar etch: (*a*) dep/etch/dep; (*b*) dep/etch/dep/etch/dep.

Micron Ex. 1029, p. 445
Micron v. YMTC
IPR2024-00911

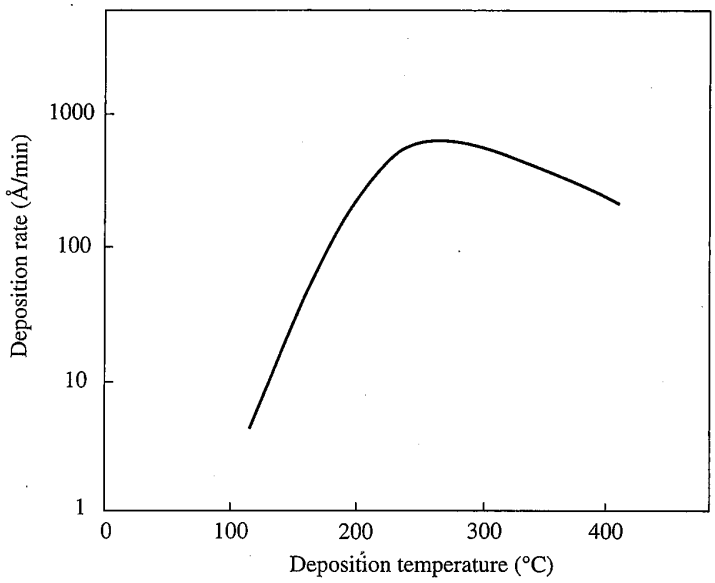


FIGURE 24
Deposition rate of O_3-TEOS vs. wafer temperature.

the growth rate is only about 70% that on Si. This sensitivity to surface condition suggests that a heterogeneous reaction on the surface is responsible for the decomposition of TEOS. The behavior in Fig. 24 also suggests that there is more than one reaction at higher deposition temperatures.

A peculiar characteristic of TEOS/O_3 deposition is the exhibition of flow-like step coverage,[134] as shown in Fig. 26. This led to a model[134] for the formation of liquid-like polymers of TEOS on the surface:

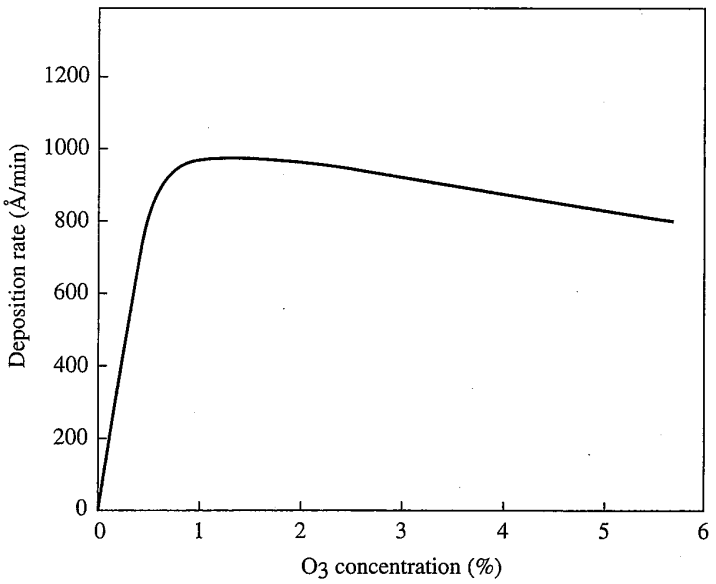


FIGURE 25
Deposition rate of O_3-TEOS vs. O_3 concentration.

Micron Ex. 1029, p. 446
Micron v. YMTC
IPR2024-00911

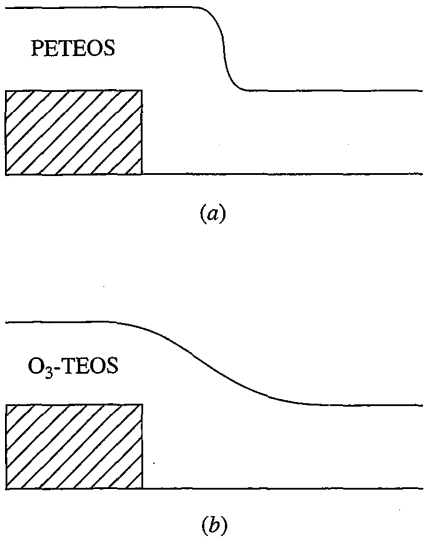


FIGURE 26
Flow-like characteristics of O_3-TEOS deposition: (*a*) PECVD TEOS; The step coverage is conformal; (*b*) O_3-TEOS APCVD; The step is flow-like.

$$n\text{Si}(\text{OC}_2\text{H}_5)_4 \rightarrow \text{Si}_n\text{O}_{(n-1)}(\text{OC}_2\text{H}_5)_{2(n+1)} + 2n\text{C}_2\text{H}_4 + n\text{H}_2\text{O} \qquad (8.46)$$

In this model, the TEOS oligomer eventually decomposes into SiO_2:

$$\text{Si}_n\text{O}_{(n-1)}(\text{OC}_2\text{H}_5)_{2(n+1)} \rightarrow n\text{SiO}_2 + 2n\text{C}_2\text{H}_4 + n\text{H}_2\text{O} \qquad (8.47)$$

Note that, although the mechanism for TEOS/O_3 deposition is not known in detail, the foregoing liquid model is not widely accepted. The flow-like behavior has also been explained by the more recent surface reaction model. These models[135,136] propose that the dissociation of TEOS is through the oxidation by atomic O from O_3. Although the details are beyond the scope of this book, Fig. 24 can be understood with a simplified model. If we assume that both the dissociation of O_3 (into $O + O_2$) and the oxidation of TEOS by atomic O occur on the surface, the Arrhenius behavior of Fig. 24 can be explained. The activation energy matches that for O_3 dissociation, and one may conclude that this is the rate-limiting factor.[135,137] At higher temperatures, desorption of TEOS from the surface becomes significant, which lowers the deposition rate. The flow-like structure can be explained[137] using the well-known fact that the vapor pressure of a droplet of radius r follows the Kelvin equation:[138]

$$\log_e p/p_0 = 2\gamma V/rRT \qquad (8.48)$$

where γ = surface tension
V = molar volume
p_0 = vapor pressure of a planar surface

The shoulder portion of the deposition (Fig. 26) has a higher vapor pressure for TEOS, which results in a lower deposition rate because of a higher desorption rate. For the knee portion the situation is the opposite: A convex geometry causes a low vapor pressure (for TEOS) and a higher deposition rate. Further experimental verification of these models is necessary to establish a detailed mechanism for TEOS/O_3 deposition.

Micron Ex. 1029, p. 447
Micron v. YMTC
IPR2024-00911

The characteristics of O_3-TEOS can be summarized: (1) It can be deposited by APCVD or SACVD at 250 to 400°C at rates depending on both temperature and O_3 concentration; (2) the oxide has good step coverage and can fill higher aspect ratio gaps than PECVD TEOS; (3) the film properties are sensitive to surface conditions; and (4) the film properties are functions of deposition temperature and O_3 concentration. Higher temperature and higher O_3 concentration lead to higher-quality oxides but at the cost of lower deposition rates.

O_3-TEOS oxide is somewhat porous, with a shrinkage of 1 to 3% after a 450°C anneal. When left in air, the film absorbs moisture gradually. The water absorption can cause gate oxide damage and via poisoning and has to be avoided. Figure 27 shows the process sequence of a structure where the O_3-TEOS oxide is etched back to provide a gap filler. In this approach, a thin layer of O_3-TEOS oxide is first deposited to fill the gap (Fig. 27*a*). If it is left on top of the metal, then moisture absorption and re-release during metal deposition will lead to oxidation and poisoning of the metal contact (Fig. 27*b*). Instead, an RIE process is used to remove the O_3-TEOS oxide (Fig. 27*c*). Finally, a thick PECVD TEOS is deposited to complete the ILD process (Fig. 27*d*). This process can fill narrow gaps with aspect ratios of about 1.3 to 1.5. For higher aspect ratio gaps new deposition processes are needed.

Spin-on-glass (SOG)

Spin-on-glasses are widely used for low-cost fabrication of IC circuits. There are two basic types of SOG: organic and inorganic. The inorganic SOGs are silicate-based, and the molecular structure is shown in Fig. 28*a*. The structure of an organic

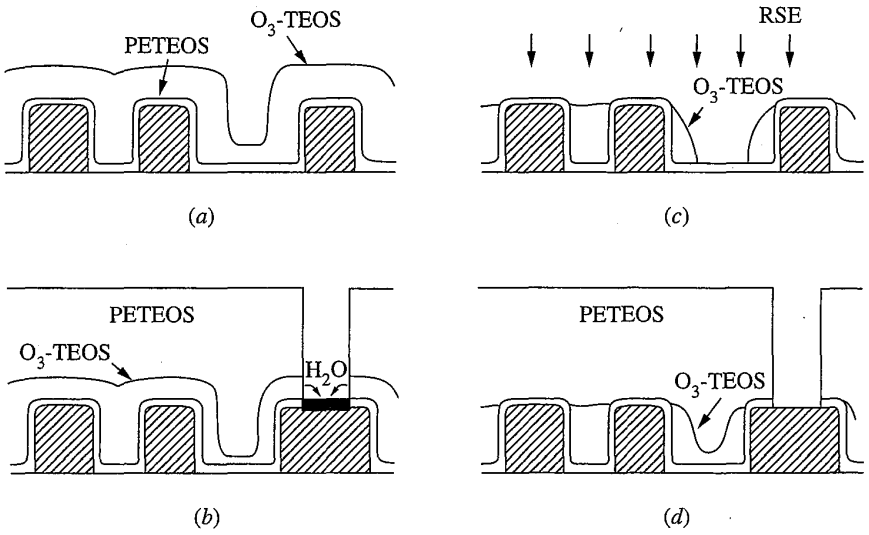


FIGURE 27
O_3-TEOS/PETEOS ILD processes: (*a*) PETEOS liner before O_3-TEOS deposition; (*b*) if the O_3-TEOS is left over metal, the exposed portion may absorb moisture and outgas, oxidizing the metal contact; (*c*) to avoid contact problems, remove O_3-TEOS from contact area by RSE; (*d*) a process that leaves no exposed O_3-TEOS.

Micron Ex. 1029, p. 448
Micron v. YMTC
IPR2024-00911

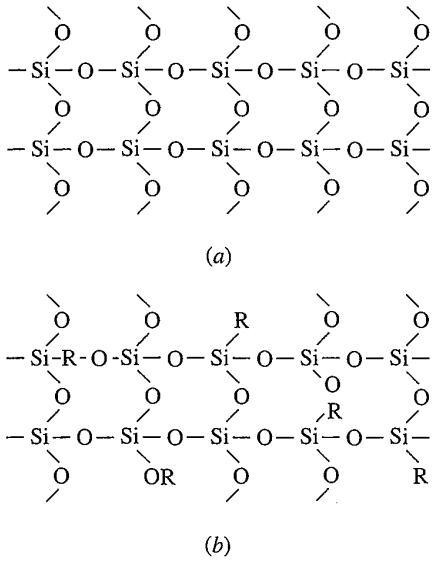


FIGURE 28
Schematic of the molecular structures of SOGs: (*a*) silicate SOG; (*b*) siloxane SOG.

SOG is shown in Fig. 28*b*. The organic, or siloxane-based, SOGs are featured with radical groups replacing or attaching to oxygen atoms. From these two basic structures, numerous commercial modifications have been made to adjust the molecular weight, the viscosity, and the final film properties for specific applications.

After full curing, the silicate SOG behaves like SiO_2. It does not absorb water in significant quantity and is thermally stable. However, the volume shrinkage during curing is large, and as a result the silicate SOG develops high stress ($\sim$ 400 MPa) and cracks easily during curing. Applying only a thin layer (100 to 200 nm) avoids the cracking problem. To build up thicker layers to fill gaps, multiple application and curing are needed. The ability to fill narrow gaps is also limited since low-viscosity SOGs (which have better filling properties) usually have higher shrinkage.

Silicate SOGs can be modified by adding P (up to 4 at%) to produce a softer SOG. The P is incorporated into the silicate by replacing Si atoms and breaking Si–O bonds. Therefore, adding P to a silicate SOG creates a structure similar to the siloxane SOG (Fig. 28*b*). The modified SOG (called P-silicate) is softer and has less shrinkage. The stress after curing is reduced to $\sim$ 200 MPa, and the SOG is more crack-resistant. It also has a better gap-filling capability and can be applied in thicker layers. Adding P, however, increases the water absorption and the modified SOG is both mechanically and electrically less stable than the silicate SOG.

Siloxane SOG, on the other hand, retains carbon even after curing. It has lower stress ($\sim$ 150 MPa) and can be applied in thicker layers. The organic SOG, however, is sensitive to water absorption, cannot tolerate even moderate temperature (> 400°C) cycles, and has poor mechanical strength. Organic SOGs may also become unstable under plasma exposure. The amount of organic component in the siloxane is an adjustable parameter, and the SOG can be tailored to specific applications.

Because silicate SOGs are more SiO_2-like, they are used mainly for gap filling even when multiple applications and curing are necessary. The organic SOGs are more suitable for sacrificial SOG, e.g., the planarizing layer for etchback.[139] (See

Micron Ex. 1029, p. 449
Micron v. YMTC
IPR2024-00911

Section 8.5.3.) Because of water absorption and subsequent via poisoning by SOGs, they are used only for gap fill and the excess is removed. The processing sequence shown in Fig. 27*a*, *c*, *d* applies to SOG with the simple replacement of SOG for O_3-TEOS oxide.

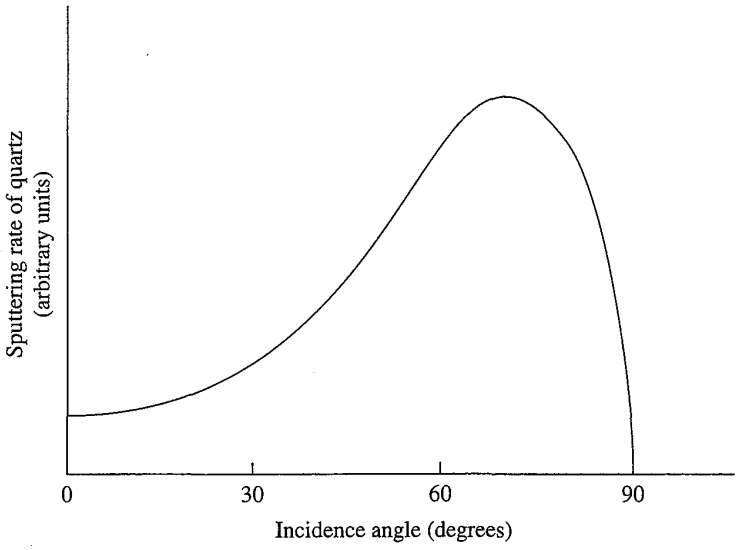
Because both the material and the method of application are inexpensive, SOG has become a popular low-cost ILD for gap fill. The etchback step (Fig. 27*c*) is expensive and a potential source of particles. Therefore, a nonetchback SOG process is important. Recent work has reported spin-on materials that can be left on metal without jeopardizing via contacts.[140,141] Fluorine treatment of a siloxane SOG to use as a nonetchback ILD is also reported.[142] No large-scale testing has been completed for these new materials and therefore the acceptance is still experimental. Although some fully cured SOGs have lower dielectric constants (3.0 to 3.6 compared with 3.9 to 4.5 for SiO_2) than oxides, in general SOGs absorb water and both the dielectric and mechanical properties are inferior to those of PECVD oxides. Lower cost, good gap-filling capability, and better ILD planarity still make SOG an attractive alternative to PECVD oxide. Currently, SOGs are used for manufacturing devices from > 1 μm to 0.35 μm in size.

Bias-sputtered quartz

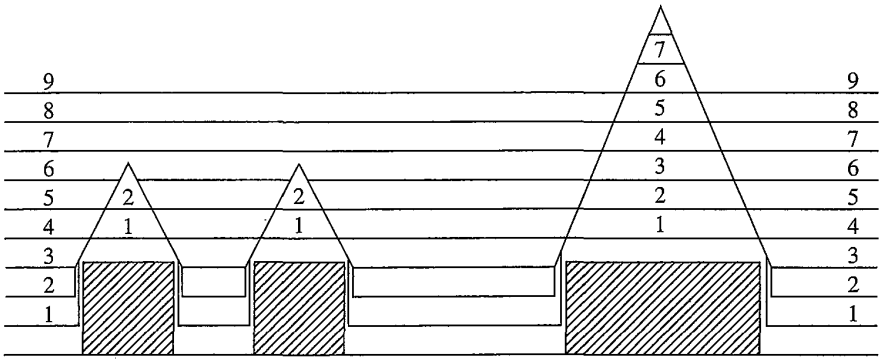
SiO_2 can be deposited by rf sputtering from a quartz target at room temperature. The sputter-deposited oxide is porous, has poor step coverage, and is not suitable for ILD use. Adding a dc bias, however, changes both the film quality and the step coverage drastically. The ion bombardment during sputter deposition provides sufficient energy to densify the oxide to a level similar to thermal oxide. The energetic ions sputter off the overhang from feature edges and create shadow-free deposition. Consequently, a complete filling of gaps between metal lines is achieved. Figure 29*a* shows the resputtering rate of SiO_2 as a function of ion incident angle.[143] The sputtering rate is low at normal incident angle and stays low up to about 30°. It has a broad peak at about 70° and then falls off sharply to zero at 90°. In bias quartz sputtering the deposition rate and bias are adjusted so that the deposition rate is between the minimum and the peak of the resputtering rates. If the deposition rate falls below the minimum resputtering rate, then the substrate is eroded and there is no deposition. If the deposition rate is higher than the peak of the resputtering rate, then bias sputtering is not effective, resulting in poor step coverage. Figure 29*b* shows the layer profiles over different features for biased sputter deposition. Note the building of sharp ridges over metal features. As the SiO_2 grows in thickness the ridges on smaller features disappear first. Bias-sputtered deposition (Fig. 29*b*) can fill narrow, high-aspect-ratio gaps solidly when parameters are properly adjusted.

Bias-sputtered quartz has not been used widely for ILD applications for the following reasons: (1) the low deposition rate; (2) particulates from chamber walls and shielding; (3) metal contamination from chamber parts; (4) radiation damage from energetic electrons (which produce x-rays); and (5) high compressive stress ($\sim$ 1000 MPa) of the film, which cracks easily. However, the benefits of bias sputtering during deposition have been recognized. The combination of bias sputtering with plasma-enhanced CVD of SiO_2 led to the most promising ILD process (see next section on HDP oxide).

Micron Ex. 1029, p. 450
Micron v. YMTC
IPR2024-00911



(*a*)



(*b*)

FIGURE 29
Bias-sputter deposition: (*a*) The sputtering rate of quartz vs. the incidence angle of Ar^+; (*b*) profiles of bias-sputter-deposited layers.

High-density plasma (HDP) enhanced SiO_2

Many of the shortcomings of bias-sputtered quartz can be overcome by combining bias sputtering with plasma-enhanced deposition. CVD oxide has a much higher deposition rate and is free from particulates from chamber walls. In addition, CVD oxide is free from metallic contamination and can produce films with lower stress. However, to achieve high deposition rate and effective bias sputtering without radiation damage one needs a high density of low-energy ions. Recent advances in high-density plasma sources (see Chapter 5) provide the necessary

Micron Ex. 1029, p. 451
Micron v. YMTC
IPR2024-00911

ingredient for combined CVD SiO_2 and bias sputtering for high-quality ILD at low temperature.[144,145]

Modern high-density plasma sources are capable of supplying low-energy ions with density $\geq 10^{12}$ cm^{-2}. The high density of plasma provides oxides of exceptional quality even at low deposition temperatures. In addition, by applying a bias, the re-sputtering effect, similar to bias-sputtered quartz, allows the filling of narrow gaps. As a result, even hard-to-fill gaps caused by the undercutting of Al wiring during the RIE of multilayered structures can be filled with a good quality oxide.

$TEOS/O_2$ chemistry similar to PECVD TEOS oxide has been used to deposit SiO_2 using electron cyclotron resonance (ECR) plasma at low temperature.[146] More commonly, silane and oxygen are used following the chemistry in Eq. (8.45). Silane is preferred over TEOS because (1) good step coverage is ensured by bias sputtering and therefore the conformal coverage for TEOS is not a significant advantage, (2) silane chemistry leaves no C residue, (3) silane is a gaseous precursor and is easier to handle than the liquid TEOS precursor, and (4) the poor film properties of silane-based oxide in PECVD are remedied by the higher-density plasma.

One drawback of high-density plasma is that, because it is so efficient, the plasma is sustained at a pressure one order of magnitude lower than for PECVD. Lower pressure (mtorr) means a lower deposition rate, about 200 nm/min. Note that this is not one order of magnitude lower than the rate of PECVD oxide, because higher plasma density helps to accelerate the oxidation of silane. To increase the deposition rate one must increase the pressure, the flow rate, or the microwave power. It is desirable to maintain a low pressure to avoid gas phase reactions and subsequent particulate generation. Usually, the maximum microwave power has already been applied, and it is hard to increase the plasma density further. By increasing the silane flow rate, a deposition rate as high as 650 nm/min has been achieved.[147] At this rate low-cost single wafer processing is possible.

In HDP oxide deposition, bias sputtering is achieved by an rf bias separately applied to the substrate to avoid charge accumulation. During deposition, the substrate is self-biased to a dc voltage that serves to accelerate the Ar ions for bias sputtering. This independent rf supply also provides an additional adjustable parameter for controlling the film properties, deposition rates, etc. Because of the high density of plasma, heating of the wafer during deposition is significant. Consequently, the thermal chuck that is used to maintain the wafer temperature needs to be carefully designed to provide effective cooling.

The high plasma density also raises the question of device damage. Although the energy of the ionized species in the plasma is low (< 40 eV), the substrate rf self-bias can be several hundred volts. Consequently, the oxide is deposited under constant ion bombardment with significant energy. Studies have been conducted to investigate the plasma damage.[148,149] These results indicate that there is no direct gate oxide damage (10 to 15 nm) during ECR oxide deposition. Larger-scale and long-term effects on gate oxide have not been reported yet.

High-density plasma oxide deposition equipment has just become commercially available. Because of the ideal properties of this process for ILD and for gap filling, the HDP oxide will become the dominating ILD process for devices 0.35 μm and below.

Micron Ex. 1029, p. 452
Micron v. YMTC
IPR2024-00911

Device and contact degradation by ILD

Device and contact degradation can be caused by the ILD process through three different paths: (1) hydrogen or hydroxyl ions in low-quality oxides, (2) plasma damage, and (3) moisture absorption and the subsequent oxidation of metal contacts. The first two can cause V_t shift and hot carrier degradation, whereas the oxidation of metal contacts can cause high via contact resistance and early via electromigration failure.

PECVD oxides have a large number of unsaturated bonds. In addition, the oxides are more defective and have a density only $\sim$ 90% of that of thermal oxide. If they are not to be used as an ILD, a high-temperature anneal (800 to 950°C) in O_2 or steam can densify these oxides to nearly 100% of thermal oxide. However, if they are to be used as an ILD, the maximum temperature for annealing is restricted to $\leq$ 450°C and the oxide property remains inferior. An exception to this is the oxide deposited by high-density plasma, for which good oxide quality is achieved at low temperature.

Hydrogen and hydroxyl groups ($-$OH) are known to cause dangling bonds in the gate oxide and to accelerate hot carrier aging.[150] These species diffuse rapidly through PECVD oxides and are not effectively gettered by phosphosilicate glass (PSG). Consequently, some ions can reach the gate oxide where traps are created. The detailed mechanisms of hot carrier aging are discussed in Chapter 12.

Both PECVD oxide and HDP oxide employ a plasma and can potentially cause plasma damage to the gate oxide. Since the ILD is deposited after the first level of metal, some transistors are subjected to a high "antenna ratio." An example of the antenna effect is when a long Al wire is attached to the gate of a MOSFET. During a plasma process (RIE or deposition) the entire length of this wire can collect charges and charge up the gate oxide of the MOSFET. Since the ratio of the wire length to the width of the MOSFET can be large (100 to 1,000) significant charge accumulation can occur. Plasma damage effects are discussed in detail in Chapter 12.

O_3-TEOS and SOG are porous and water-absorbing. When exposed to aqueous solutions, or even cleanroom air, these oxides absorb water quickly from the atmosphere. The absorbed water can outgas in the vacuum chamber before contact metal deposition. Because the vias are opened directly to the metal, and the pumping in the via is always poor, the water vapor is chemisorbed on the metal and oxidizes the metal just before the contact metal deposition. This oxidation causes high contact resistance, known as via poisoning. To avoid this problem, these materials are used only as a gap filler, as shown in Fig. 27, to prevent air exposure.

8.5.3 Planarization

The demand for more levels of metal wiring for the ever-increasing complexity of circuits can only be fulfilled if ILD planarity can be maintained. The planarity of ILD serves two purposes—to provide a smooth surface to ensure good metal step coverage and to provide a flat-enough field, within the lithography depth of focus, that contact vias and metal wires can be patterned.

Micron Ex. 1029, p. 453
Micron v. YMTC
IPR2024-00911

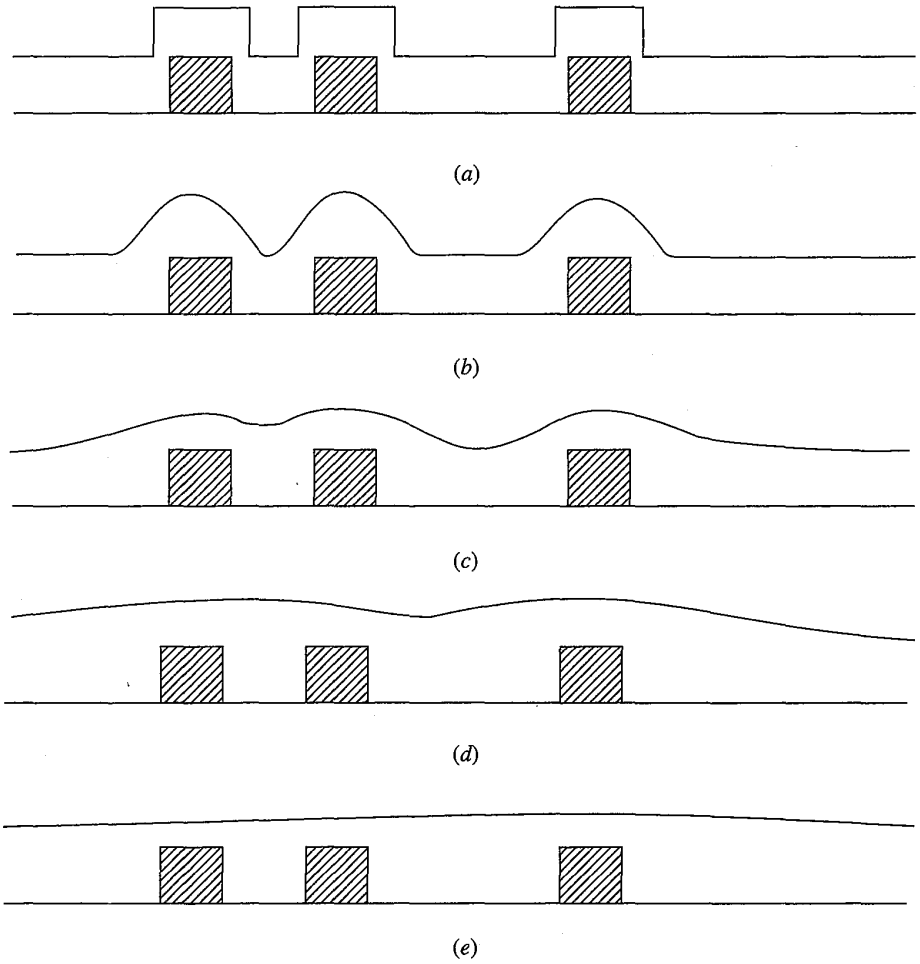


FIGURE 30
Qualitative definition of planarization: (*a*) No planarization; (*b*) smoothing; (*c*) partial planarization; (*d*) local planarization; (*e*) global planarization.

It is useful to define planarization terms unambiguously to avoid confusion. There is currently no universally recognized standard for planarization. Here we will define planarization both qualitatively and quantitatively. We follow the terminology defined by Wolf,[151] as shown in Fig. 30. Planarization is divided qualitatively into four categories:

1. *Smoothing:* Corners rounded and sidewalls sloped, but the step height not significantly reduced (Fig. 30*b*).
2. *Partial Planarization:* Smoothing plus a reduction in step height locally, as in Fig. 30*c*.
3. *Local Planarization:* Complete filling up of smaller gaps (1 to 10 μm); total step height to flat areas not significantly reduced (Fig. 30*d*).
4. *Global Planarization:* Local planarization plus a significant reduction in the total step height, as shown in Fig. 30*e*.

Micron Ex. 1029, p. 454
Micron v. YMTC
IPR2024-00911

There are two ways to define planarization quantitatively. The first is to measure the reduction in step height. The metric used, the planarization factor β, is defined by $\beta = 1 - h_f/h_i$, where h_i is the initial step height before planarization and h_f the final step height. This definition does not consider the lateral extent of the planarization and is used for characterizing local planarization. For global planarization, the planarity is defined by the relaxation distance,[152] or its equivalent, the planarization angle θ. As shown in Fig. 31, at a single step edge the dielectric becomes thicker but eventually falls back to the original thickness. This behavior can be approximated by three straight line segments. The horizontal span of the tapered region is defined as the planarization relaxation distance R. The angle between the taper and the x-axis is defined as the planarization angle θ. R and θ are related simply by $R(\tan\theta) = h$, where h is the step height.

Both R and θ are used in the literature as metrics for planarity. Figure 32 shows the ranges of R and θ for various ILD deposition and planarization processes. PECVD, O_3-TEOS, and HDP oxides are mainly for ILD and gap fill and have relaxation distances in the range of < 5 μm and $\theta \approx 30$ to 45°. Local planarization is achieved with SOG or resist etchback with R between 1 and 100 μm and θ between 1 and 45°. Chemical/mechanical polishing (CMP), on the other hand, provides true global planarization with R in millimeters and $\theta < 0.01°$.

Lithography depth of focus

The limiting factors of resolution vs. depth of focus for optical lithography have been discussed in Chapter 6. In simple forms the resolution and the depth of focus can be written as

$$\text{Resolution} = K_1\lambda/\text{NA}^2 \tag{8.49}$$

$$\text{DOF} = \pm K_2\lambda/(\text{NA})^2 \tag{8.50}$$

where K_1 and K_2 are factors controlled by both the optical system and the photoresist and NA is the numerical aperture of the lens. To achieve ever higher resolution, the wavelength has been decreasing and the NA increasing for each generation of new technology. Consequently, the depth of focus for lithography systems is shrinking faster than improvements in the resolution are appearing. This decrease, plus the

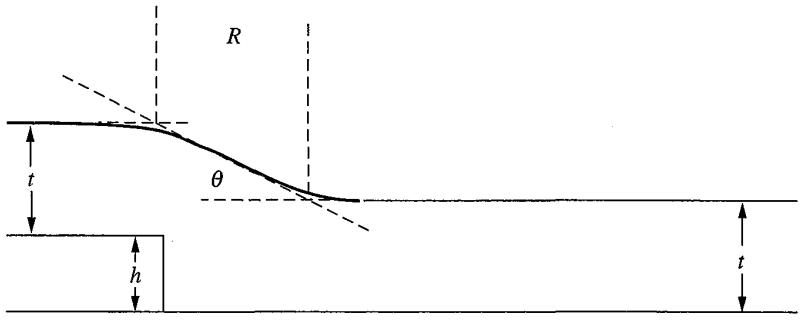


FIGURE 31
Quantitative definition of planarization: R = the relaxation distance, θ = planarization angle.

Micron Ex. 1029, p. 455
Micron v. YMTC
IPR2024-00911

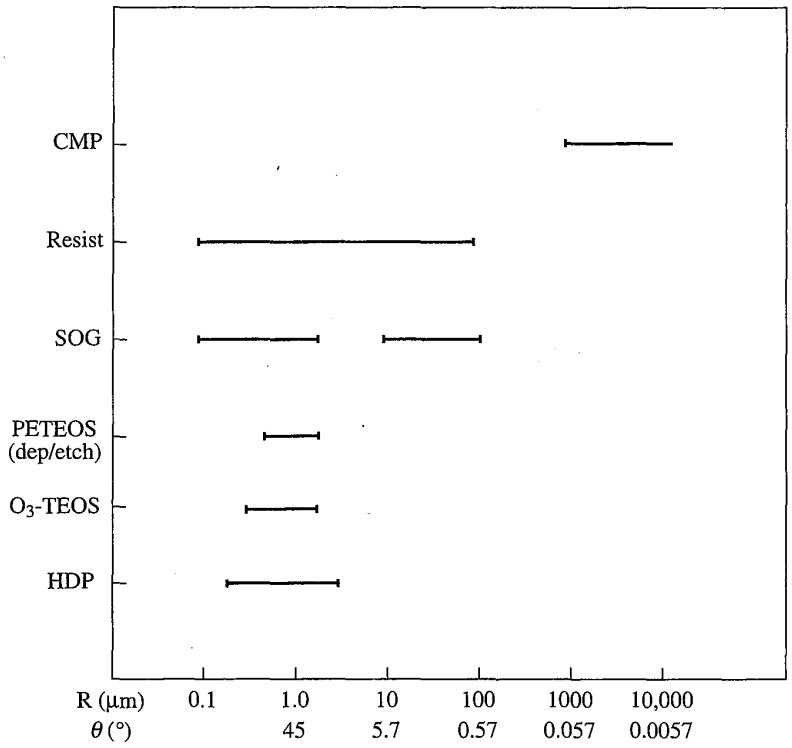


FIGURE 32
The range of planarization for various processes.

demand for more levels of metal wiring, is rapidly pushing the surface topography beyond the depth of focus capability. Figure 33 shows how the depth of focus is exceeded quickly by the buildup of topography when multiple levels of metals are used. The film thickness in Fig. 33 is typical for a 0.5-μm design. At window-2 and metal-2, the total topography has already exceeded the depth of focus limit if 0.5-μm resolution is desired from I-line lithography. Without using global planarization to reduce the topography, window-2 and metal-2 design rules have to be relaxed to allow the use of I-line lithography. At window-3 and metal-3, the design rules have to be grossly relaxed to stay within the depth of focus limits. The increase in die size causes a substantial penalty in yield and cost.

Etchback planarization

The surface topography caused by previous polysilicon or Al wiring can be smoothed by applying a thick layer of dielectric and a sacrificial layer of planarizing material such as photoresist or SOG and etching back, as shown in Fig. 34. The photoresist or SOG can smooth the surface features from a few microns to a few tens of microns. The entire structure is then etched back to the final ILD thickness by RIE, as shown in Fig. 34*b*. The smoothing of the ILD surface helps to achieve better Al step coverage as well as better metal linewidth control. The latter occurs through the removal of sharp edges at steps, which helps to reduce reflection from metal surfaces during lithography. Since this planarization is only local, it does not reduce the

Micron Ex. 1029, p. 456
Micron v. YMTC
IPR2024-00911

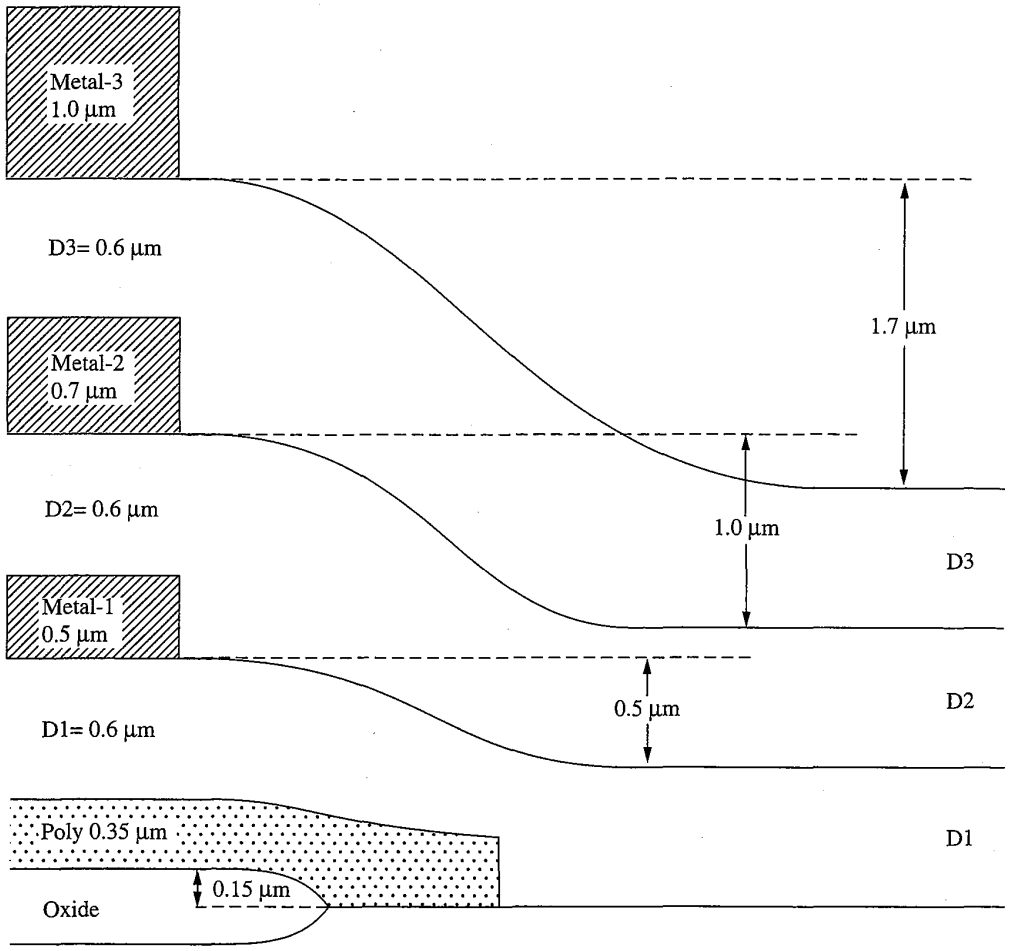


FIGURE 33
The buildup of topography by multilevel metallization.

total topography within a lithography field. The buildup of total surface topography increases just as rapidly as without planarization (Fig. 34*d*). Consequently, this process can be applied only to the first and at most the second dielectric and is difficult to extend to ILD planarization for three or more levels of metallization.

An etchback process can be used for global planarization in combination with a dummy resist, metal, or oxide mask.[153] (The dummy mask is sometimes called a planarization block mask, or PBM.) For example, if the metal pattern is designed so that there is no open area greater than a few micrometers in size, then the etchback process is adequate to provide uniform topography for the entire lithography stepper field. In practice, such a restriction cannot be part of the design rules, since it would compromise design freedom excessively. A method that does not pose any restriction to the design rules is to add electrically inactive "dummy" features whenever an open area greater than a set number (e.g., 4 μm) is detected. This approach of adding dummy metal features is shown in Fig. 35. The dummy metal features are incorporated into the metal mask and the processing sequence is not changed.

Micron Ex. 1029, p. 457
Micron v. YMTC
IPR2024-00911

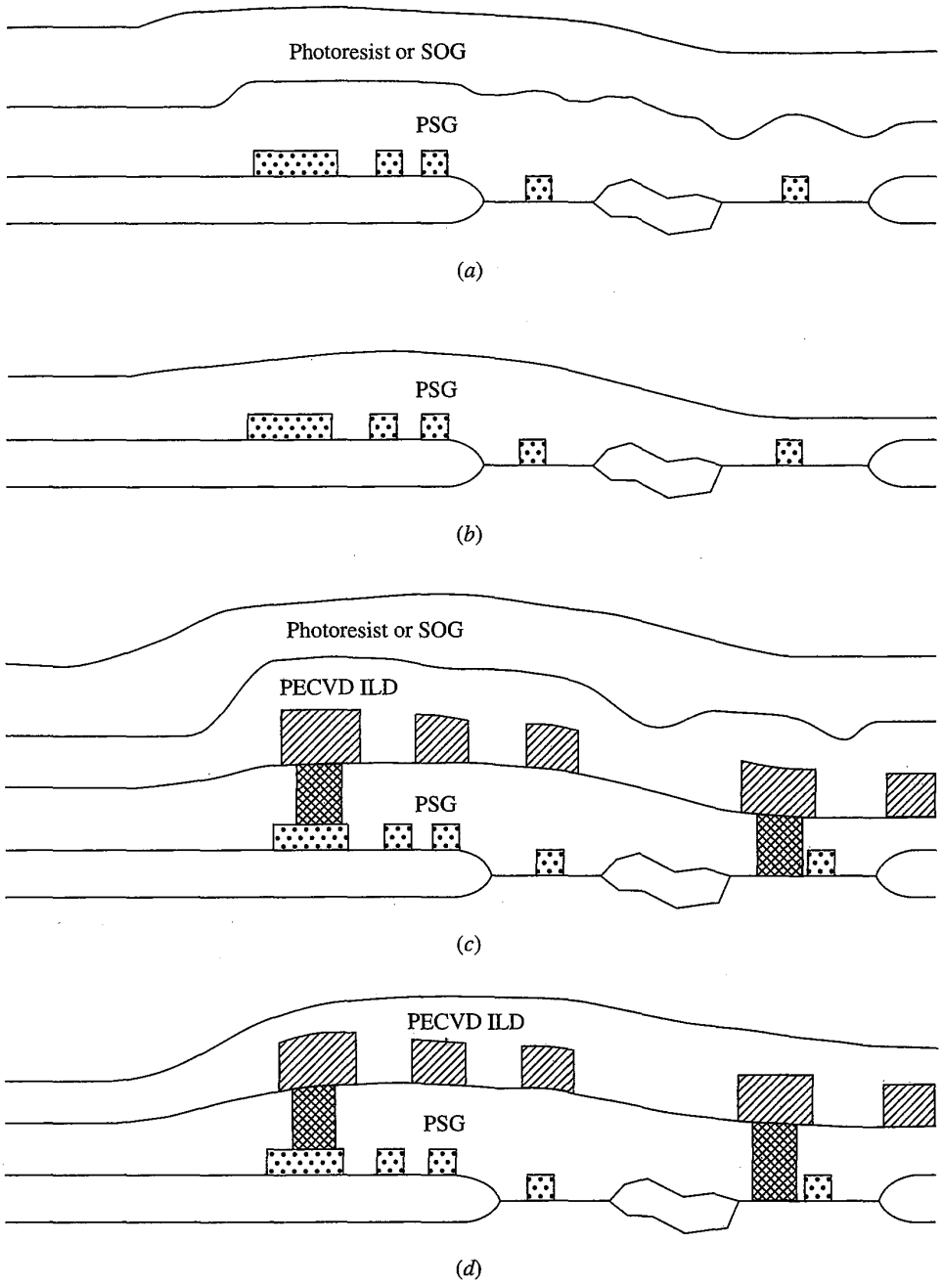


FIGURE 34
Etchback planarization: (*a*) Sacrificial layer over PSG; (*b*) topography smoothed by etchback; (*c*) after W plug, metal-1, ILD, and the deposition of a new sacrificial layer; (*d*) etchback provides a smooth morphology again, but does not remove built-up topography.

Micron Ex. 1029, p. 458
Micron v. YMTC
IPR2024-00911

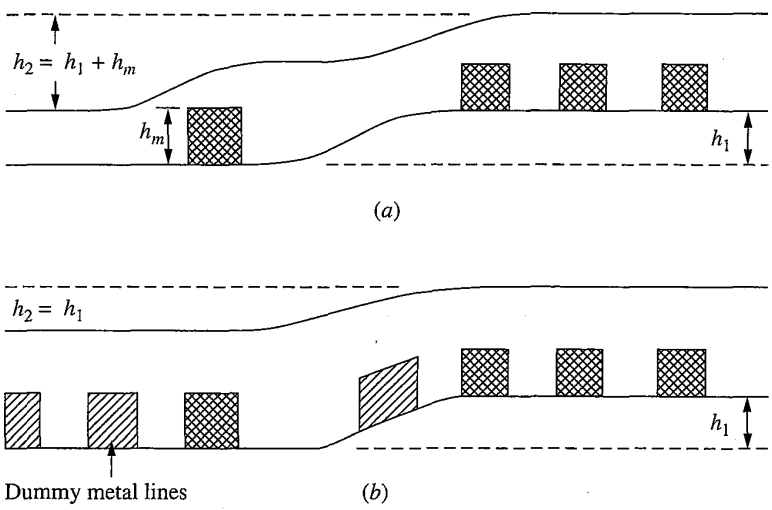


FIGURE 35
Global planarization using etchback and dummy metal features: (*a*) etchback with no dummy metal features results in local planarization only; (*b*) dummy metal + etchback provides global planarization.

Generating dummy features of any kind, however, requires substantial CAD development and CAD processing time. As a result, the design cost and the design cycle time are increased even when the processing cost is not always increased. Metal dummy features are also undesirable from the performance point of view, since wiring capacitance is increased. Even when the dummy features are electrically inactive, they must be held either at ground or at certain voltages to avoid interference with the circuit. ULSI circuits are designed through CAD simulation that does not account for nearby dummy features that can float to any voltage. To ground and connect the dummy metal features not only requires more CAD work but also complicates the design work. Therefore, even though dummy metal lines are easy to process, from a design point of view they are unacceptable.

An alternative is to use dummy oxide features instead of metal features. This is a more complex process because the oxide features need to be deposited, patterned, and etched independent of the metal features. The benefit is that it does not impose capacitance and grounding problems and is a more acceptable solution than metal dummy features.

One dummy oxide feature approach is shown in Fig. 36. After the lower-level metal patterning and etching are completed, a regular ILD is applied. A dummy oxide mask, which has to be generated by a CAD tool and contains information about densely patterned and open areas, is then used to pattern the ILD. As shown in Fig. 36*b*, the area containing densely packed metal lines is then etched by roughly the amount corresponding to the height of the underlying metal. This approach leaves a final topography between dense and open areas of roughly the same height, as shown in Fig. 36*c*. After this, a new layer of ILD and a sacrificial layer of photoresist or SOG are applied, and the entire structure etched back to the final ILD thickness. This approach is relatively simple, the oxide mask is noncritical and does not cost a great deal of CAD time, and the processing is straightforward. A drawback is its

Micron Ex. 1029, p. 459
Micron v. YMTC
IPR2024-00911

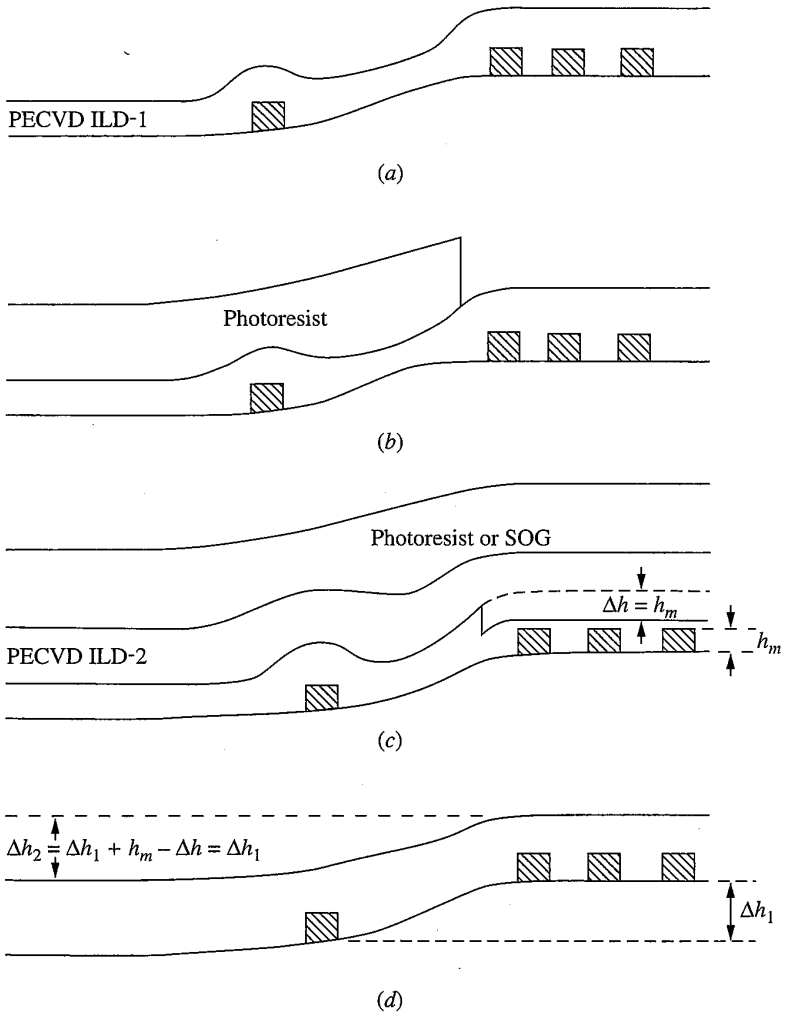


FIGURE 36
Global planarization by oxide mask and etchback: (*a*) Deposit ILD-1 for dummy features, the thickness ≈ metal height; (*b*) lithography for dummy features, (*c*) RIE ILD-1, apply a thicker ILD-2 and a sacrificial etchback material, (*d*) etchback planarization. Note that this process provides both local smoothing and global planarization, but does not remove topography from previous layers.

inability to handle gray areas where metal lines are separated by 5 to 20 μm and are somewhat irregular. To handle all possible designs, a refinement of this process is necessary, which will make the process more complex and the CAD work more elaborate.

Global planarization can also be achieved with an oxide etching process. As shown in Fig. 37, the topography introduced by metal lines can be removed by etching trenches into the next layer ILD and then smoothing out the structure by SOG or resist etchback. First, a thick layer of ILD is applied over the lower-level metal. Then a reverse-tone metal mask is used to pattern the ILD. RIE is used to etch trenches into the ILD, followed by wet etching to smooth the features somewhat, as in Fig. 37*c*.

Micron Ex. 1029, p. 460
Micron v. YMTC
IPR2024-00911

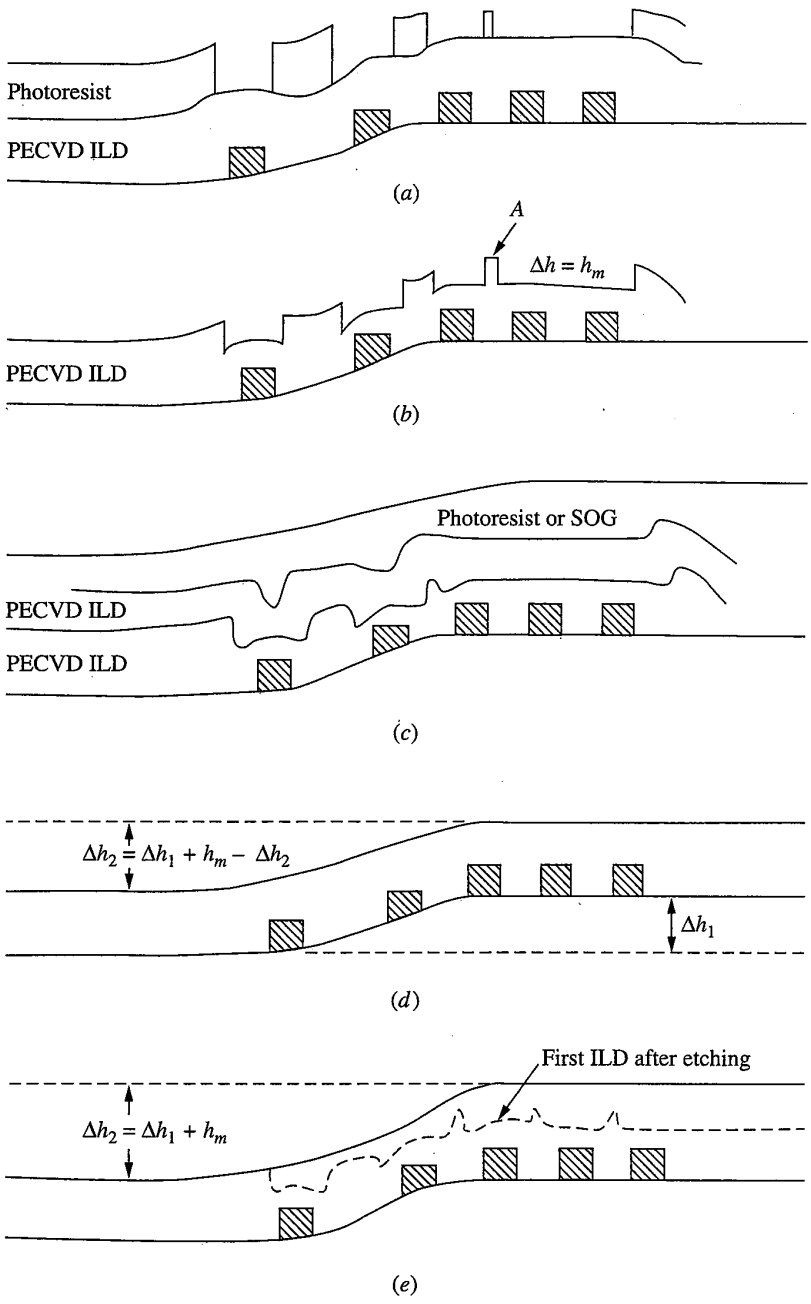


FIGURE 37
Global planarization by reverse-tone oxide mask and etchback: (*a*) Reverse-toned mask has larger features to relax alignment difficulties; (*b*) after oxide etch, the features are still abrupt; (*c*) wet etch removes sharp corners and thin "walls" such as A; then deposit ILD-2 and a sacrificial layer; (*d*) etchback planarization; (*e*) topography in areas with certain metal spacings is not reduced.

Micron Ex. 1029, p. 461
Micron v. YMTC
IPR2024-00911

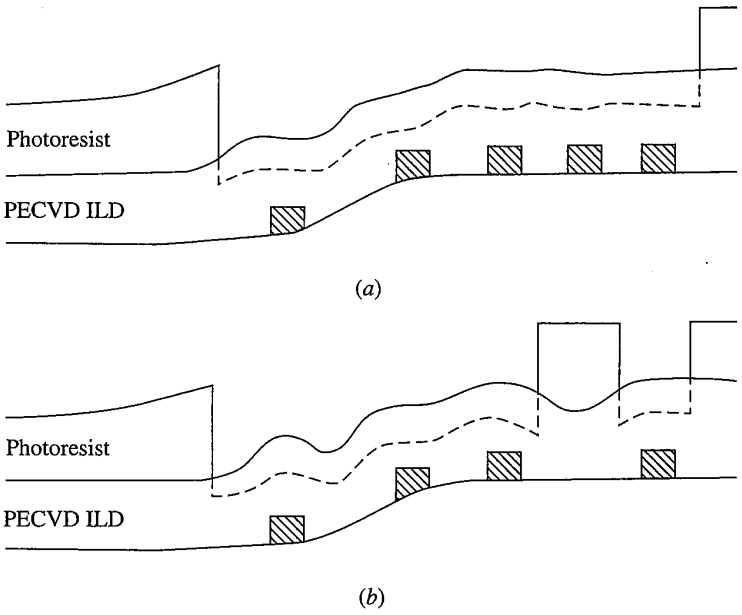


FIGURE 38
Using CAD to refine reverse tone mask: (*a*) When metal spacings are less than a certain value, no pattern is put between neighboring lines; (*b*) when two lines are separated by a certain distance, an extra line is put in. The dashed lines show the trenches after RSE of ILD-1. With the CAD refinement, the situation as in Fig. 37*e* will not occur.

Finally, a sacrificial photoresist or SOG is applied and the entire structure blanket-etched back to the final ILD thickness. An immediate problem is the misalignment between the trenches and the metal lines. In addition, as can be seen from Fig. 37*e*, for clustered metal lines separated by a certain distance the topography reduction is poor. At this separation an array of "walls" is left after RIE and wet etching, rendering planarization ineffective.

The reverse tone oxide patterning technique can provide a global planarization but is imperfect. This problem can be reduced by applying some CAD surgery to the reverse tone oxide mask. A thorough design with a CAD tool can decide where to put trenches and the proper size of the trenches. By applying these techniques, a reverse tone oxide mask can produce good global planarization, as shown in Fig. 38.

Even though the etchback approach can provide global planarization when combined with dummy feature masks or reverse tone masks, it does not provide totally flat topography and requires considerable process development. The additional mask required, the CAD work, and the additional oxide etch also increase the processing cost. These processes are rapidly being replaced by a low-cost, truly global planarization process, discussed next.

Chemical/mechanical polishing (CMP)

A true global planarization can be achieved by chemical/mechanical polishing (CMP) of the ILD.[154] This process has been used to prepare Si wafers for more than

Micron Ex. 1029, p. 462
Micron v. YMTC
IPR2024-00911

30 years, as well as for glass polishing and for bonded silicon-on-insulator (SOI) wafers. Although its application to ILD planarization is more recent, both the equipment and the technique are well known to the semiconductor industry.

The requirements and the process for polishing a dielectric are quite different from those for polishing Si. For dielectric polishing, the goal is to remove topography and yet maintain good uniformity across the entire wafer. The amount of material removed is about 0.5 to 1 μm, compared to Si polishing where several tens of micrometers of material are removed. The uniformity requirement for ILD polishing is much more stringent than for Si polishing, since nonuniform ILD films lead to window etching and plug formation difficulties. Consequently, even with more than 30 years of experience in Si wafer polishing, the industry found that new processes and new equipment had to be developed for ILD polishing. Recently, CMP has been applied to polishing metal for W plug formation and embedded metal structures. Metal polishing involves chemistries significantly different from those for oxide polishing and requires further research and development.

The schematic of the CMP equipment is shown in Fig. 39. The essence of a CMP process is an automated rotating polishing platen and a wafer holder, which can both exert a force on the wafer and rotate the wafer independent of the rotation of the platen. The polishing is accomplished by a polishing slurry consisting of colloidal silica suspended in a KOH solution. An automatic slurry feeding system is used to ensure the uniform wetting of the polishing pad and the proper delivery and recovery of polishing slurry. For a unit designed for industrial use, automatic wafer loading and a cassette-to-cassette handler are also incorporated.

The basic polishing mechanism for an SiO_2 dielectric is the same as for glass polishing and has been summarized recently.[155,156] The mechanical removal rate of the glass is given by the Preston equation,[157]

$$R = K_p p v \tag{8.51}$$

where R = rate of removal
p = applied pressure
v = relative velocity between the wafer and the polishing pad
K_p = proportionality constant

K_p has units of (pressure)$^{-1}$ and is known as the Preston coefficient. K_p is a function

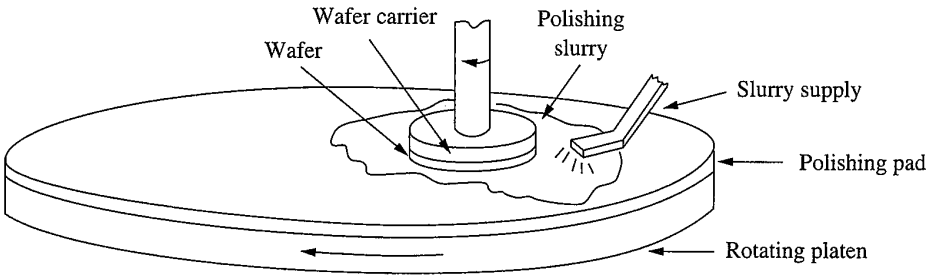


FIGURE 39
A schematic of a CMP polisher.

Micron Ex. 1029, p. 463
Micron v. YMTC
IPR2024-00911

of the mechanical properties of the glass (hardness, Young's modulus), the polishing slurry, and the composition and the structure of the polishing pads.

Although Eq. (8.51) is essentially mechanical in nature, the microscopic action of polishing is both chemical and mechanical. The exact mechanism for polishing is still not fully understood. Figure 40 shows the current understanding of the chemical/mechanical events during polishing. The chemical reactions can be divided into four stages: (1) the formation of hydrogen bonds with the oxide surfaces of both the wafer and the slurry particles (hydroxylation), as in Fig. 40*a*; (2) the formation of hydrogen bonds between the wafer and the slurry (Fig. 40*b*); (3) the formation of molecular bonds between the wafer and the slurry (Fig. 40*c*); and (4) the breaking of the oxide bonds with the wafer (or the slurry) surface when the slurry particle moves away (Fig. 40*d*).

From the chemical/mechanical events shown in Fig. 40, there are three important implications: (1) Polishing is not a mechanical abrasion of slurry against the wafer surface; (2) both the presence of water and the pH of the solution affect the formation of hydrogen bonds; and (3) the size and the composition of the slurry particles are important. Relying on chemical/mechanical action rather than mechanical abrasion avoids a mechanically damaged surface layer on the ILD. The microscopic nature of CMP distinguishes it from mechanical abrasion, which is unacceptable to ULSI applications. Because of the chemical nature of bond breaking, the polishing slurry has a large effect on the polishing rate. The most commonly used slurry for SiO_2 polishing is silica with a particle size about 10 to 90 nm. In general, slurries made from oxides with higher oxygen bond strengths give higher polishing rates. The highest polishing rate achieved, with cerium oxide, is several times higher than that of silica.

Figure 41 illustrates the arrangement of a patterned wafer on a polishing pad. Instead of the wafer being held directly on the chuck, a backing film is sandwiched between the wafer and the chuck. The film provides elasticity between the chuck and the wafer. A composite set of two pads is used to achieve the desired rigidity/elasticity for polishing. Figure 42*a* illustrates the need for elasticity between the chuck and the wafer. Without a backing film any defect or particle on the chuck or on the back of the wafer will cause a thin spot. In addition, wafer breakage is more likely. If the polishing pad is too soft, as in Fig. 42*b*, then polishing is conformal to the wafer surface and no planarization is achieved. Planarization is caused by the capability of the polishing pad to bridge over low spots on the wafer and thus preferentially removing material from the high spots, as shown in Fig. 42*c*.

Therefore, microscopically, the removal of material in contact with the polishing pad is mainly chemical in nature. Planarization occurs because only high spots on the wafer touch the polishing pad. Because the macroscopic removal of material depends on the mechanical contact between the wafer and the polishing pad, the Preston equation [Eq. (8.51)] still applies well to CMP.

The detailed mechanism for planarization has not been fully elucidated and is an active field of research. A phenomenological model[158] using three parameters, corresponding to high features, low features, and horizontal erosion, is successful in simulating the erosion profiles of isolated and clustered features. Recently, a fluid model, which treats the slurry as a hydrodynamic layer, has been developed.[159] This

Micron Ex. 1029, p. 464
Micron v. YMTC
IPR2024-00911

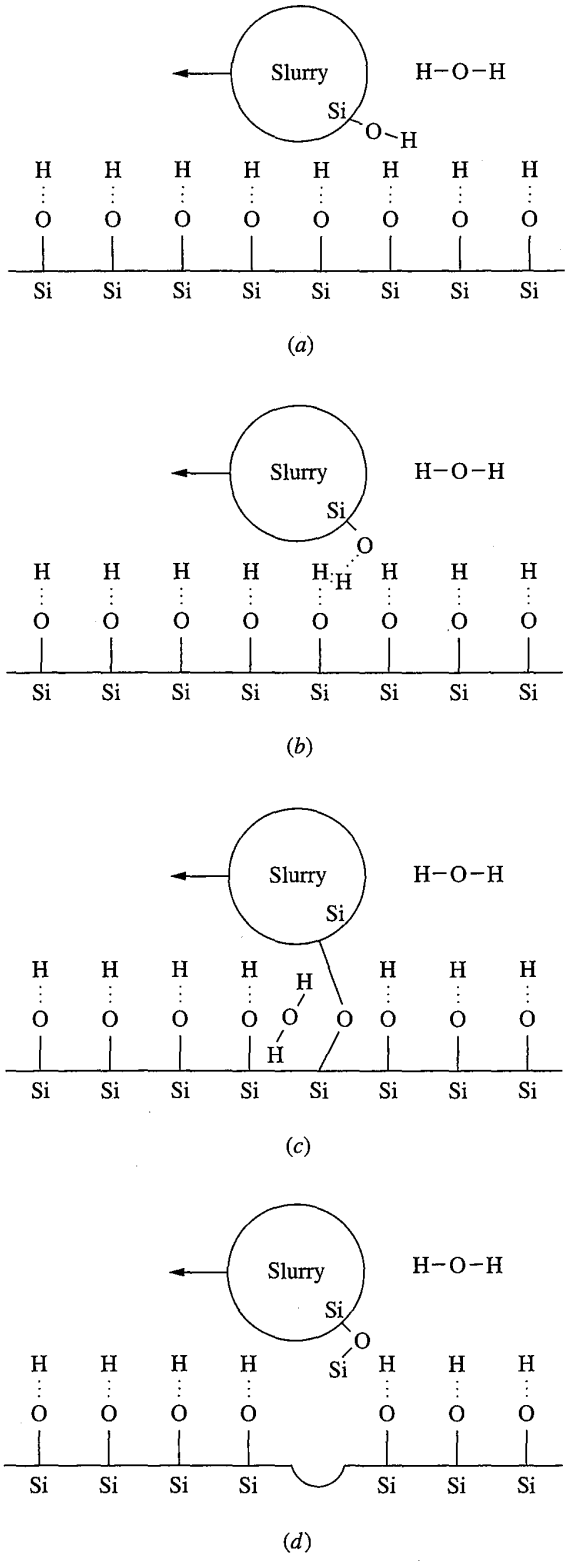


FIGURE 40
The mechanism of chemical/mechanical polishing: (*a*) In aqueous solution oxide forms hydroxyls; (*b*) hydrogen bond is formed between the slurry particle and the wafer; (*c*) Si–O bonds are formed by releasing a water molecule; (*d*) the Si–Si bond breaks when the slurry particle moves away.

Micron Ex. 1029, p. 465
Micron v. YMTC
IPR2024-00911

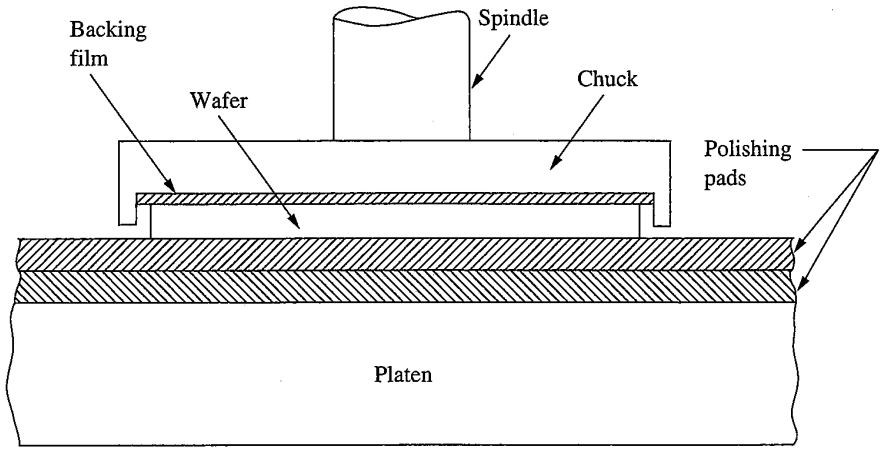


FIGURE 41
Details of the CMP wafer carrier and polishing pads.

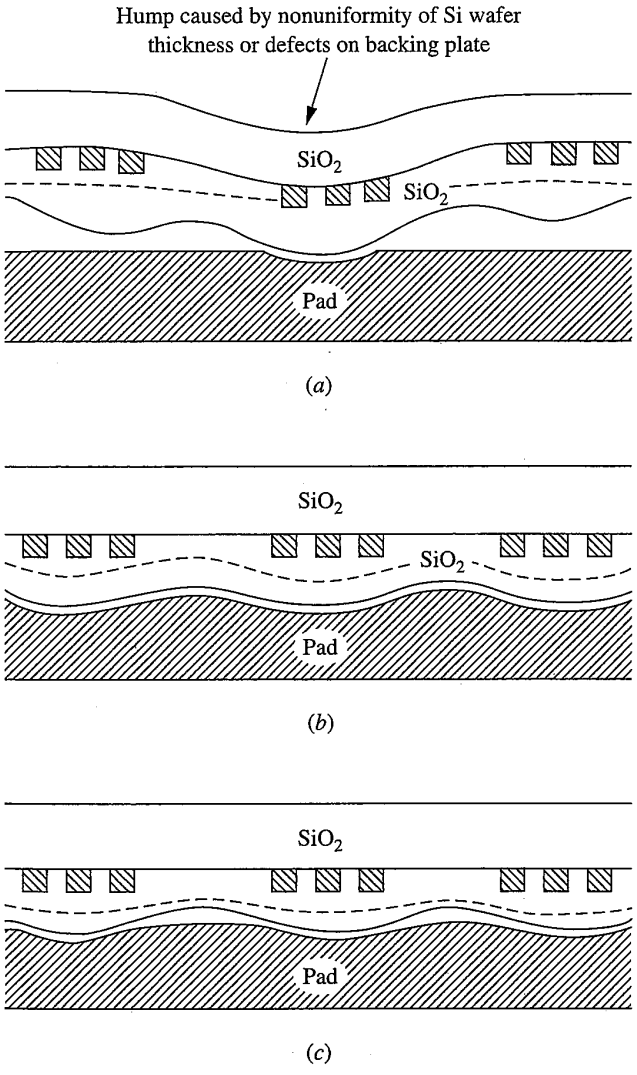


FIGURE 42
The effect of rigidity of the backing chuck (film) and the polishing pad: (*a*) A rigid chuck may bend the wafer and cause a thin spot; (*b*) a soft pad conforms to the surface and does not planarize; (*c*) a moderately rigid pad polishes the high spot on the wafer. The dashed lines show the final polished profile.

Micron Ex. 1029, p. 466
Micron v. YMTC
IPR2024-00911

model showed limited success in simulating the erosion profiles. Therefore, qualitatively the planarization by CMP is understood, but quantitative understanding is still lacking.

The mechanism for CMP of metal is less studied and less understood and is more complex than oxide polishing. Figure 43 illustrates one proposed mechanism for metal polishing.[160] This model employs both a chemical etching and a passivation mechanism. Metals form an oxide layer on the surface, as shown in Fig. 43*a*. During polishing, this oxide layer is removed by a mechanism similar to that for oxide polishing (i.e., hydroxylation + bond formation with slurry + bond breaking from wafer). Once the oxide is removed, the metal is etched by the chemicals in the slurry solution, as shown in Fig. 43*b*. Simultaneously, the exposed metal forms a new passivation layer through oxidation by the solution, as in Fig. 43*c*. In reality, the three processes (the removal of metal oxide, metal etching, and repassivation) occur simultaneously

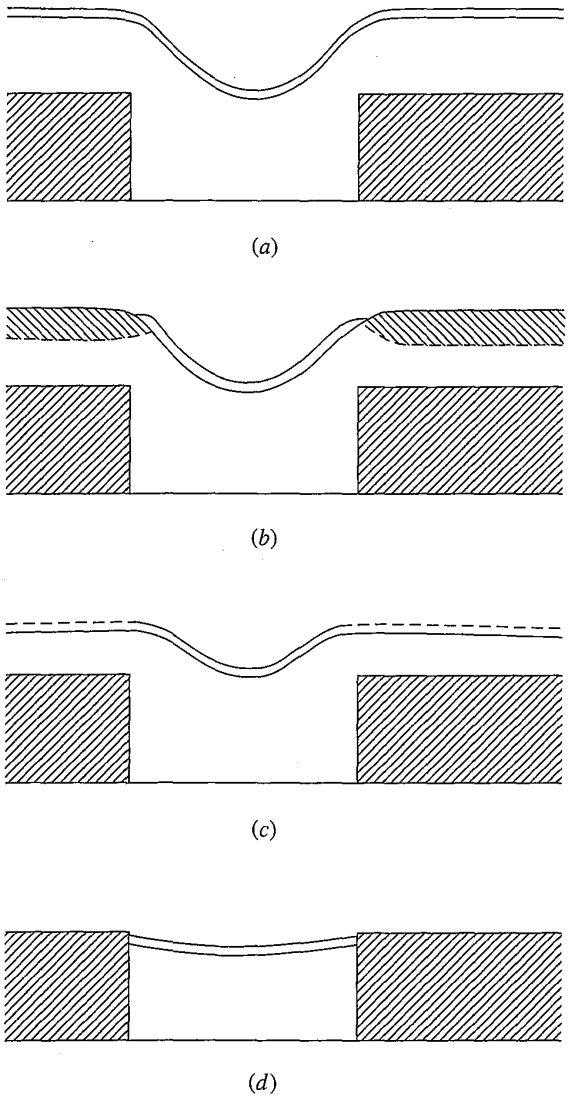


FIGURE 43
Mechanism for metal polishing: (*a*) Free metal surface is covered by a passivation oxide; (*b*) the oxide on the high spots is removed by CMP and the metal is etched by the solution (shaded area); (*c*) regrowth of oxide; (*d*) processes in (*b*) and (*c*) repeat until no high spot is left.

Micron Ex. 1029, p. 467
Micron v. YMTC
IPR2024-00911

and the polishing rate is determined by the balance of these processes. Therefore, for metal CMP the polishing slurry must contain all three components: the fine slurry particles, a corrosion (etching) agent, and an oxidant. Figure 43*d* shows the eventual end point of metal CMP. Planarization is achieved by the mechanical rigidity of the polishing pad similar to oxide polishing.

For metal CMP, the most extensive experience is with W CMP. Fine alumina powder is the most commonly used slurry. The slurry solution consists of potassium ferricyanide, $K_3Fe(CN)_6$, and potassium dihydrogen phosphate, KH_2PO_4, with the pH adjusted to about 5.0 to 6.5.[160] In this solution, the ferricyanide serves both as the etchant and the oxidant. The KH_2PO_4 provides the desired pH for etching. The etching reaction is as follows:

$$W + 6[Fe(CN)_6]^{3-} + 4H_2O \rightarrow (WO_4)^{2-} + 6[Fe(CN)_6]^{4-} + 8H^+ \quad (8.52)$$

Competing with the etching reaction is the passivation reaction:

$$W + 6[Fe(CN)_6]^{3-} + 3H_2O \rightarrow WO_3 + 6[Fe(CN)_6]^{4-} + 6H^+ \quad (8.53)$$

Because of the delicate balance between etching, passivation, and oxide removal, the control of the slurry is important. If etching [Eq. (8.52)] becomes dominant, then W is etched isotropically and no planarization is achieved. To improve planarity, a weak base, ethylenediamine, is added to the slurry to adjust the pH closer to neutral. Potassium ferricyanide, however, is highly toxic, and the disposal of used slurry causes severe environmental impacts. This compound is now replaced by ferric nitrate $Fe(NO_3)_3$.

A major difficulty in the application of CMP is to guarantee uniform polishing over the entire wafer (uniformity). To achieve uniform polishing, three conditions have to be met: (1) Every point on the wafer has to travel at the same relative velocity to the polishing pad, (2) the polishing slurry is uniformly distributed under the entire wafer, and (3) the wafer itself should be symmetrical. Condition 3 is difficult to achieve for 150-mm wafers with a primary or a secondary flat and is easier to achieve for 200-mm wafers with a notch instead of flats. Condition 2 is hard to guarantee because new slurry always gets to the edge of the wafer first. Condition 1 is a controllable parameter and is examined in the following.

Consider the configuration in Fig. 44. Let the rotation speed of the wafer, which is at a distance R from the center of the platen, be ω_2, and let the rotation speed of the platen itself be ω_1. Consider three points A, B, and C on the wafer. A and B are at the opposite edge of the wafer whereas C is at the center of the wafer. The relative velocities of points C, A, and B to the platen are given by

$$v_C = \omega_1 R \quad (8.54)$$

$$v_A = \omega_1(R - r) + \omega_2 r \quad (8.55)$$

$$v_B = \omega_1(R + r) - \omega_2 r \quad (8.56)$$

It is easy to see that these three points will move at the same relative velocity to the platen if $\omega_2 = \omega_1$. Figure 45*a* shows the trajectories of the points A, B, and C on the platen after a complete revolution. Under the condition of $\omega_2 = \omega_1$, the loci of points A and B are circles and both A and B travel the same distance as C.

Micron Ex. 1029, p. 468
Micron v. YMTC
IPR2024-00911

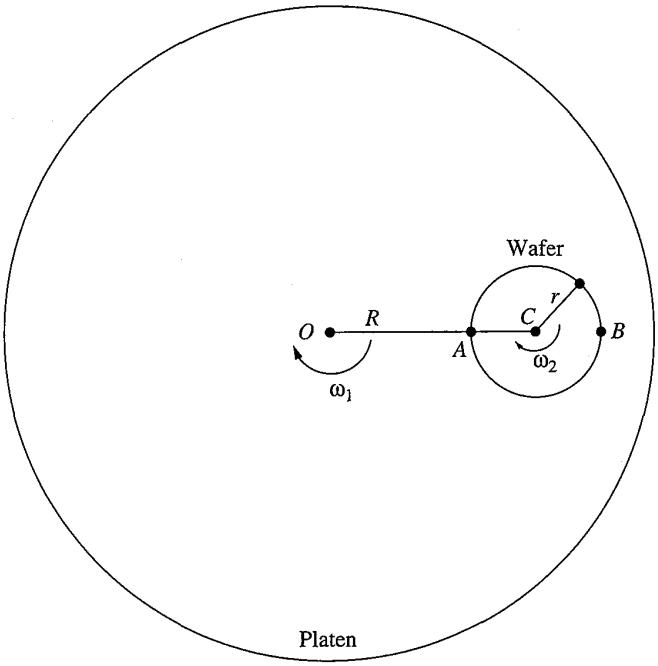


FIGURE 44
The relative motion of a wafer to the platen. Point *C*, the center of the wafer, is fixed in space. Both the platen and the wafer rotate.

This is equivalent to $v_A = v_B = v_C$. (It has also been proven mathematically that when $\omega_2 = \omega_1$ every point on the wafer moves at the same relative velocity.[161]) If $\omega_2 = -\omega_1$, however, the loci of points *A* and *B* travel a longer distance than *C*, and the edge of the wafer will polish at a higher rate than the center, as in Fig. 45*b*. If the wafer rotates faster than the platen, i.e., when $\omega_1 < \omega_2$, then a point near the edge of the wafer travels a longer distance than a point near the center, as shown in Fig. 45*c*. Thus, we can conclude that when the wafer rotates at a higher speed than the platen, the edge of the wafer will polish at a higher rate, according to the Preston equation. Figure 45*d* shows the case when the wafer rotates at a speed of only $\frac{1}{2}$ of the platen. Under the condition of $\omega_1 = 2\omega_2$, a point at the edge of the wafer follows a cardiod curve whose averaged travel path over a complete turn is comparable to that at the center of the wafer.

From the foregoing illustration the best uniformity is achieved with $\omega_1 \geq \omega_2$. The edge of the wafer polishes at a higher rate than the center when $\omega_1 < \omega_2$. In practice, the polishing surfaces are not rigid and other factors contribute to the polishing uniformity. Depending on the values of *R* and *r*, a condition between $\omega_1 = 2\omega_2$ and $\omega_2 = 2\omega_1$ is used.

The other condition that must be met for uniform polishing is the uniform distribution of the polishing slurry under the wafer. Currently, the details of how slurry is transported from the edge to the center of the wafer are not clear. Since slurry transport is mostly through the surface of the pad, the edge of the wafer gets a more abundant supply of slurry than the center of the wafer. Therefore, no matter whether

Micron Ex. 1029, p. 469
Micron v. YMTC
IPR2024-00911

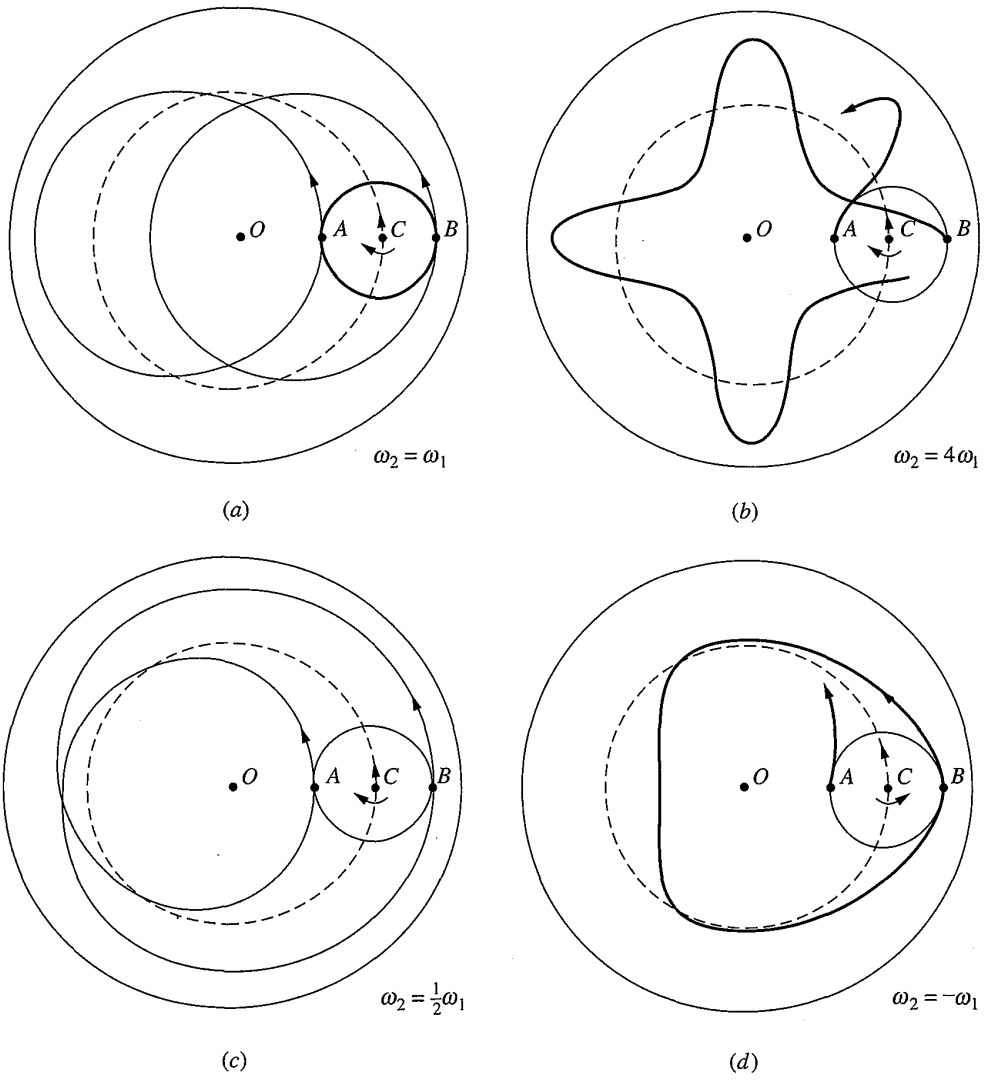


FIGURE 45
The loci of different points on a wafer: (*a*) When the wafer rotates at the same angular speed as the platen; (*b*) when the wafer rotates at a higher speed than the platen; (*c*) when the wafer rotates at $\frac{1}{2}$ the speed of the platen; (*d*) when the wafer rotates at the same speed as the platen, but in the opposite direction.

from the relative velocity aspect or from the slurry supply aspect, the edge of the wafer has a polishing rate equal to or higher than the center. To compensate for this, some equipment uses a slightly convex curvature on the carrier to exert a higher pressure toward the center of the wafer.

Consistent polishing rate and uniformity (reproducibility) are not routinely achieved for CMP yet. The lack of reproducibility is attributable to multiple sources and is still poorly understood. The Preston equation [Eq. (8.51)] connects the polishing rate to K_p (the Preston coefficient), the pressure, and the relative velocity. From our earlier analysis uniform relative velocity is achievable. The pressure p and the Preston coefficient are both functions of the *consumables* and the pattern

Micron Ex. 1029, p. 470
Micron v. YMTC
IPR2024-00911

on the wafer. The polishing pads (and to a lesser degree the backing film) and the polishing slurry are commonly referred to as the consumables. The real pressure for CMP depends both on the wafer pattern and on the contact with the polishing pad. A soft pad gives a larger contact area than a harder pad and generates a lower pressure under the same load force on the wafer carrier. Therefore, a soft pad gives a lower polishing rate. Similarly, wafers with different patterns generate different contact areas under the same load force and are polished at different rates.

The polishing pad also changes with use. Polishing pads are made of polyurethane with 1-mm perforations punched into the pads to help transport the slurry and facilitate polishing. The surface of the polishing pad is intentionally kept rough with many asperities about 1 to 10 μm in size to help hold/transport the slurry. After polishing a number of wafers, the surface asperities are flattened,[162] resulting in a state called *glazing*. When a pad is glazed, it loses some of its capacity to hold polishing slurry. In addition, glazing increases the contact area and thus causes a drop in pressure. Both changes reduce the polishing rate. A glazed pad can be rejuvenated by a process called *conditioning*. A diamond-embedded wheel is applied to the pad while the platen continues to rotate. The detailed mechanism for conditioning is not known. There is no consensus on whether the pad material is actually removed by the diamond wheel lapping. The net effect of the conditioning, however, is the regeneration of rough surfaces, which can hold/transport the slurry and provide a smaller contact area. Pad conditioning provides a key for consistent polishing rate. It is applied after the polishing of each wafer, and in some equipment concurrent pad conditioning and polishing are used.

The polishing slurry is the other important consumable, and consistency in the slurry is a key to reproducibility. For dielectric polishing, the most commonly used slurry is colloidal silica or fumed silica suspended in a KOH-based solution. The polishing rate depends on the pH of the solution, the concentration (in wt %) of the solid particles in the slurry, and the size and the distribution of the slurry particles. The polishing rate increases with increasing pH, with increasing slurry particle concentration, and with increasing particle size.[163]

To summarize: although the material removal rate in CMP can be described by the Preston equation, out of the three parameters in this equation only the relative velocity can be directly controlled. The pressure and the Preston coefficient are complex functions of processing, consumables, and wafer pattern variables. A manufacturable CMP process can only evolve from the understanding and control of these parameters.

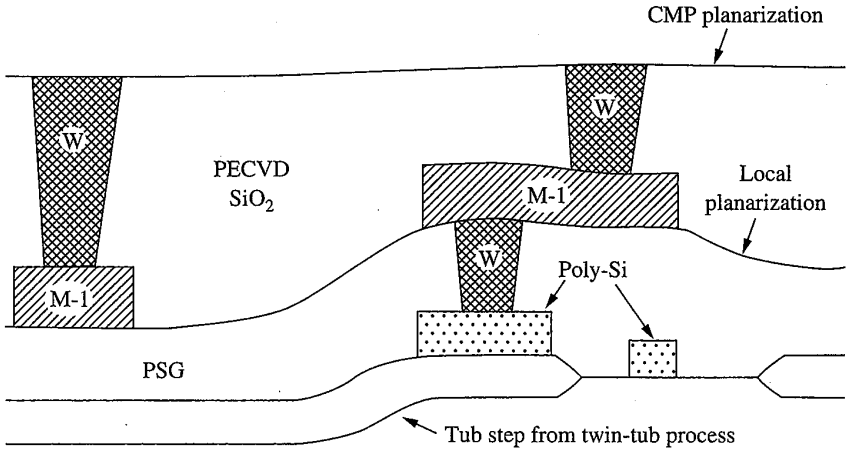
A serious drawback for CMP is the difficulty in end point detection. Without a clear signal about when the process is completed, CMP is done using empirical polishing rates and timed polish. Since the process control is immature, the empirical method fails from time to time, causing yield drops. Attempts to install an end point mechanism include capacitive measurement and optical measurement.[164] A commercial end point apparatus with limited capability has recently been introduced.[165]

The need to remove the polishing slurry from the wafer and to ensure a low level of particles on the wafer led to the specialty of "post-CMP clean." The slurry particles used in CMP range from 10 to 100 nm. These small particles attach to the wafer mainly by electrochemical forces. To establish a cleaning chemistry, one

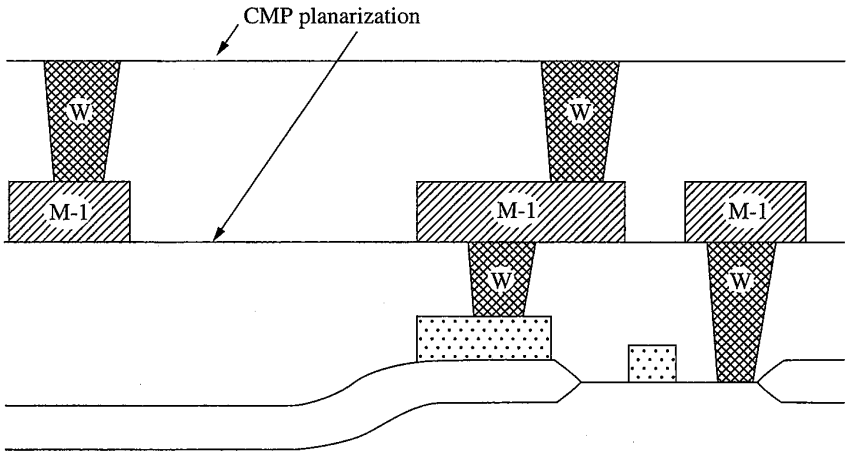
Micron Ex. 1029, p. 471
Micron v. YMTC
IPR2024-00911

must understand the details of particle attachment. So far, the post-CMP clean uses brute force—second platen buffing, soft scrubbing, ultrasonic cleaning, and chemical etching. Experience indicates that this approach is adequate for 0.5-μm device fabrication. For future, more demanding, applications, fundamental understanding of the fine particle issues will be necessary.[166]

One risk of the CMP process is the abrupt introduction of a globally planarized ILD, which creates uneven via heights over a previously unplanarized or locally planarized dielectric. Figure 46 shows the difference in via depth after CMP of an ILD over a lower level of metal on a locally planarized dielectric. Depending on the thickness of the metal wire and the ILD, the difference in via depth can be as much as 100%. Vastly different via depths increase the difficulty in the already difficult via



(*a*)



(*b*)

FIGURE 46
Via depth differences by different planarization approaches: (*a*) Large difference in via depths when CMP is used after local planarization; (*b*) vias of similar depth when CMP is used twice.

Micron Ex. 1029, p. 472
Micron v. YMTC
IPR2024-00911

lithography and etching processes. This problem can be solved by applying CMP or global planarization on earlier levels of the dielectric. The ideal approach is to apply a global planarization at the source/drain/gate contact level. Unless a totally flat structure is used, such as recessed isolation and recessed gate, the contact windows are of different depths. A global planarization at this level does not add more difference in window depth compared to a local planarization process, as shown in Fig. 46*b*. When CMP or global planarization is applied to this dielectric (the poly-metal dielectric or PMD), uniform via depth can be achieved without any penalty. The choice of how many global planarization processes should be used and where they should be applied is a complex process integration issue of cost/yield analysis.

Damascene and dual damascene

The damascene process derives its name from the ancient art of the Middle East involving inlaying metal in ceramic or wood for decoration. In ULSI application, a damascene process refers to a similar structure, as shown in Fig. 47. After the via plug process the ILD is deposited without planarization, since the surface is already flat. Trenches for metal lines are then defined and etched in the ILD and

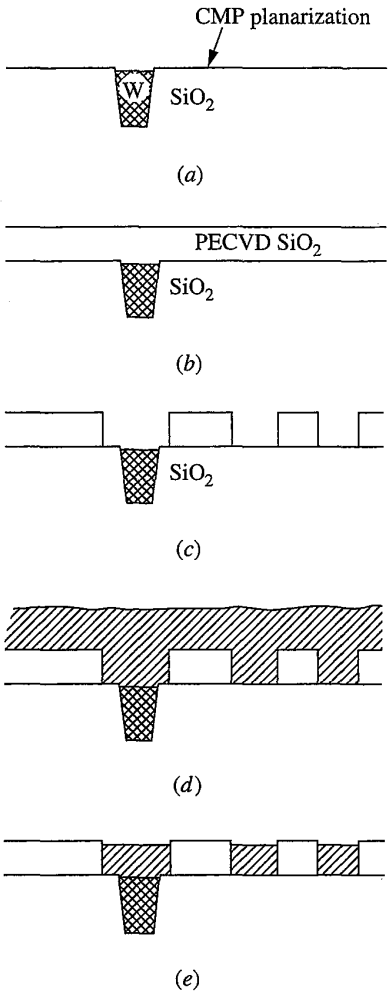


FIGURE 47
Damascene process: (*a*) Metal plugs formed after the SiO_2 is planarized by CMP; (*b*) deposited PECVD SiO_2 ILD; (*c*) trench patterning and RIE for metal lines; (*d*) metal deposition to fill trenches; (*e*) CMP metal to complete metal definition.

Micron Ex. 1029, p. 473
Micron v. YMTC
IPR2024-00911

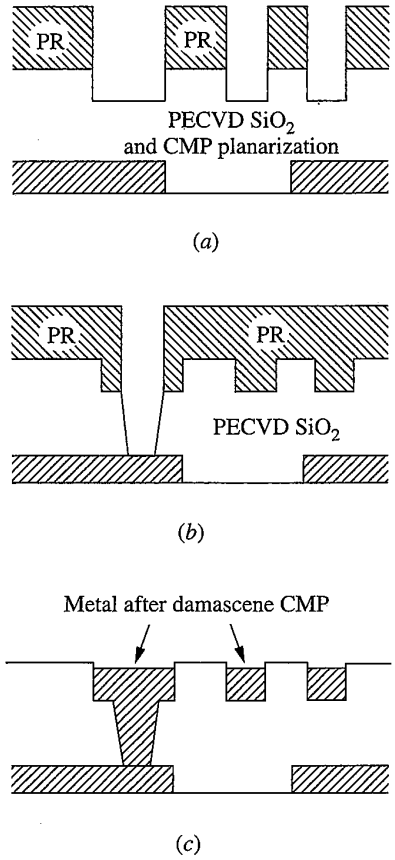


FIGURE 48
Dual damascene process: (*a*) After ILD planarization, define trench pattern; (*b*) strip photo resist, apply new photoresist, define via pattern, and RIE vias; (*c*) deposit metal in both vias and trenches, and CMP to remove excess metal. Note that the via plug is of the same material as the upper-level metal in this process.

filled with metal, either by a CVD process or by a metal flow process. The excess metal on the surface is removed and a planar structure with metal inlays in the dielectric is achieved. This process has several advantages over the traditional metal/ILD/planarization approach: (1) The surface at any time is totally flat; (2) the process eliminates the difficulty in filling small gaps between metal wires; and (3) it eliminates the difficulty in metal etching, especially for Cu and other hard-to-etch metals. On the other hand, the damascene process is more complex since (1) it still requires the via plug process and the CMP of both metal and dielectric, and (2) it requires a flat topography to start with and thus forces the use of recessed isolation, etc.

A dual damascene process is illustrated in Fig. 48. In this process vias and trenches are defined using two lithography and RIE steps, but the via plug is filled in the same step as the metal line, as shown in Fig. 48*c*. Dual damascene reduces the number of processing steps by reducing the barrier layer depositions from two to one and by eliminating the CVD W plug processes. One special benefit of dual damascene is that the via plug is now of the same material as the metal line and the risk of via electromigration failure is reduced (see Section 8.6.1).

Damascene has been demonstrated on a number of applications. The most commonly applied process is the first metal or local interconnect.[167–170] Some early damascene structures were achieved with RIE,[167] but CMP is exclusively used today. Metal interconnects using damascene of Cu[171–174] and damascene of Al[175–177] have been reported.

Micron Ex. 1029, p. 474
Micron v. YMTC
IPR2024-00911

Stacked contact and borderless contacts

As the technology scales to smaller dimensions, one of the most difficult obstacles to continued scaling is the level-to-level alignment in lithography, especially in complex structures of metals and contacts. Consequently, the metal packing density becomes limited by design rules governing the separation of one level of contacts from another, and by design rules for nesting tolerance or for borders used around contacts.

Figure 49 illustrates the area penalty imposed by contact-to-contact separations and by borders. From a fully bordered, no-stacking structure, Fig. 49*a*, to a borderless, fully stacked structure, Fig. 49*e*, a total area reduction of 62% is achieved. Intermediate area savings of significant portions can be realized by various combinations of borders and stacking, Fig. 49*b–d*. Unless the circuit is bonding-pad limited, the use of stacked contacts and borderless contacts can save more area than going to a new generation of design rules.

However, the practice of stacking contacts and borderless contacts is still in its infancy, mainly owing to processing difficulties. Figure 50 illustrates the problem with stacked contacts. If the lower contact is made with a nonplug process, then the poor Al step coverage prohibits a good contact, as shown in Fig. 50*a*. If a metal plug is used in the lower-level contact, as in Fig. 50*b*, then a good contact can be made when the second contact is stacked on top of the first one. If the metal-plug control

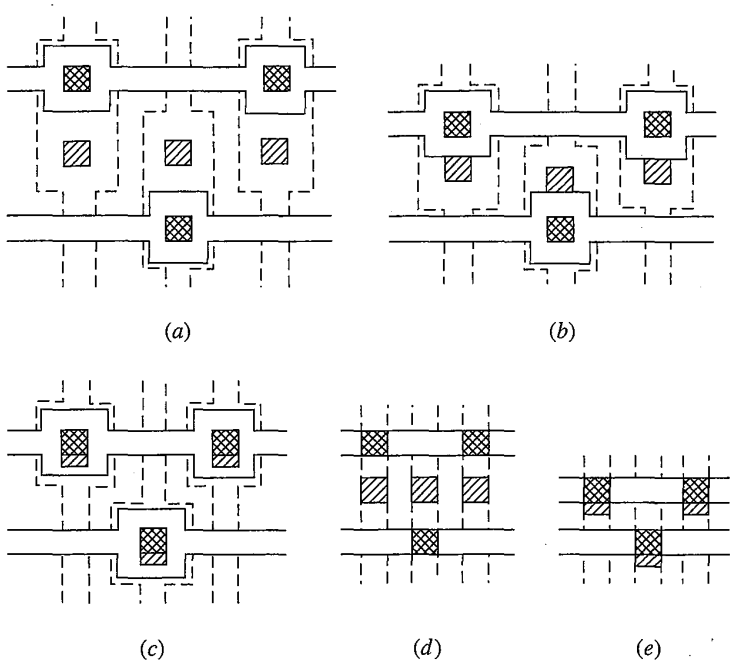


FIGURE 49
Stacked contacts and borderless contacts: (*a*) No stacking, design rules on all lines; (*b*) no stacking, no design rule separating metals on different levels; (*c*) stacked, with border; (*d*) no stacking, borderless; (*e*) stacked and borderless.

Micron Ex. 1029, p. 475
Micron v. YMTC
IPR2024-00911

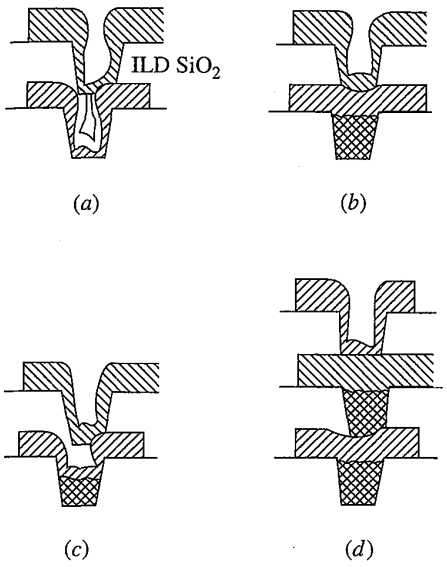


FIGURE 50
Stacked contacts: (*a*) No metal plugs in the lower contact via results in poor contact from the upper via; (*b*) a good metal plug ensures a good via contact; (*c*) a poor metal plug can still cause a poor via contact; (*d*) good via plugs allow multiple stacking of vias.

is poor, as shown in Fig. 50*c*, then poor Al step coverage can occur even with a metal plug, and a good stacked contact is difficult to make. If the metal-plug control can be maintained, as in a damascene process, then multiple stacking of contacts is possible, as shown in Fig. 50*d*.

Figure 51 illustrates the problems with borderless contacts. For a nonplug process, Fig. 51*a*, borderless contacts provide very little contact area between the upper-level via and the lower-level metal and the contact is unreliable. With any misalignment, even making contact is not ensured. If the lithography misalignment is significant, the contact area is still small, and current crowding can occur at the via contact, as in Fig. 51*b*. To make reliable borderless contacts, one can tighten the lithography alignment tolerance or relax the via design rule, as in Fig. 51*c*. True borderless contacts remain difficult to make reliably because of these problems. The use of dual damascene will help attain borderless contacts, because the filling of via contact holes and the metal wiring are accomplished in a single operation.[178]

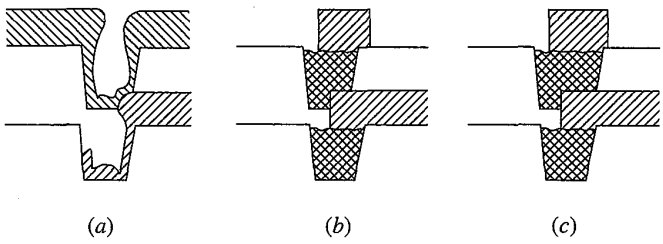


FIGURE 51
Borderless stacked vias: (*a*) Nonplug process gives poor contact; (*b*) misalignment can also cause poor contact; (*c*) a larger upper via provides better contacts to both levels of metal.

Micron Ex. 1029, p. 476
Micron v. YMTC
IPR2024-00911

8.5.4 Bonding Pad Limitations

For many ASIC applications, the power, ground, and I/O signals require hundreds of bonding pads. With the increasing capability to pack more components into a chip, the circuit sophistication increases with each generation of design rules, resulting in further demand for additional bonding pads. For many designs, the number of bonding pads becomes the limitation on chip size and, until bonding technology advances further, improvement in packing density cannot be realized by simply shrinking design rules or adding more levels of metal wiring.

Currently, bonding pitches of 80 to 120 μm are used for wire bonding in Si chip packaging. Smaller pitches are still in the exploratory stage. Consider a 400-pin device with all bonding pads distributed along the perimeter of the chip. If we assume a 100-μm bonding pad pitch, then to accommodate 100 pins per side the bonding pads will occupy 100×100 μm or 1 cm of length on each side. In this case, no further increase in circuit packing density can be achieved until the chip area becomes greater than 1 cm^2. For current technologies in volume production (0.5 to 1.0 μm minimum design rule), 1 cm^2 is the limit of optimum cost at the current achievable defect level. As a result, the strongest limitation in cost reduction is the number of bonding pads. Until packaging technology advances to reduce the bonding pitch, sophisticated circuits cannot enjoy the benefit provided by further shrinking the design rule or adding more levels of metallization.

A technique to circumvent the bonding pad limit is to use an array of pads instead of using only perimeter pads. This has the benefit of providing not only more bonding pads per chip but also shorter metal wires and thus faster circuits. This point will be discussed further in Chapter 10. It is clear, however, that packaging technology must scale and advance together with the other ULSI processing technologies. Otherwise, it will become the limitation, quickly rendering all other efforts of scaling useless.

8.6 METALLIZATION RELIABILITY

The reliability of metal wires under high current density operation and under mechanical stress has been a focus of intensive studies. These two very important phenomena are discussed in detail in Chapter 12. In this section we will discuss the contact electromigration associated with multilevel metallization, and metal corrosion.

8.6.1 Contact Electromigration

With multiple levels of wiring, it is necessary to pass current from one level of metal to another through vias. When the metal design rules scale down, the size of the contact vias also shrinks accordingly, and the current density in the vias can be as high as, and sometimes even higher than, that in the metal conductors. In practice via electromigration has been observed for two reasons: (1) The metal step coverage

Micron Ex. 1029, p. 477
Micron v. YMTC
IPR2024-00911

in the vias is poor, resulting in very high current density; and (2) a different material is used in the via (such as a W plug), causing localized current crowding[179,180] and high current density. Either of these two can cause severe contact electromigration.

The first case is illustrated in Fig. 52*a*. Consider the case of a contact via that carries a current of 0.4 mA from a transistor. If the step coverage in the via is about 10%, which is normal for a sputtered film in a hole with an aspect ratio of 1 : 1, then the amount of Al on the via wall is only about 50 nm for a 500-nm layer of Al deposition. The current density in the via can be calculated as $4 \times 10^{-4}\text{A}/0.075 \times 10^{-8}\text{cm}^2 = 5 \times 10^5\text{A/cm}^2$, or about six times that of a fully filled via. This high current density is three to five times higher than the EM design limits and will cause early EM failure. Since the current is from a nominal-size transistor, the via electromigration problem cannot be solved without an increase in the step coverage. This is accomplished by using W or Al plugs.

Figure 52*b* illustrates the EM-induced voiding when a CVD W plug is used in a via and Al wiring is in direct contact with the W plug. Even with a plug in the via, current crowding still occurs when current goes through the plug and into the next

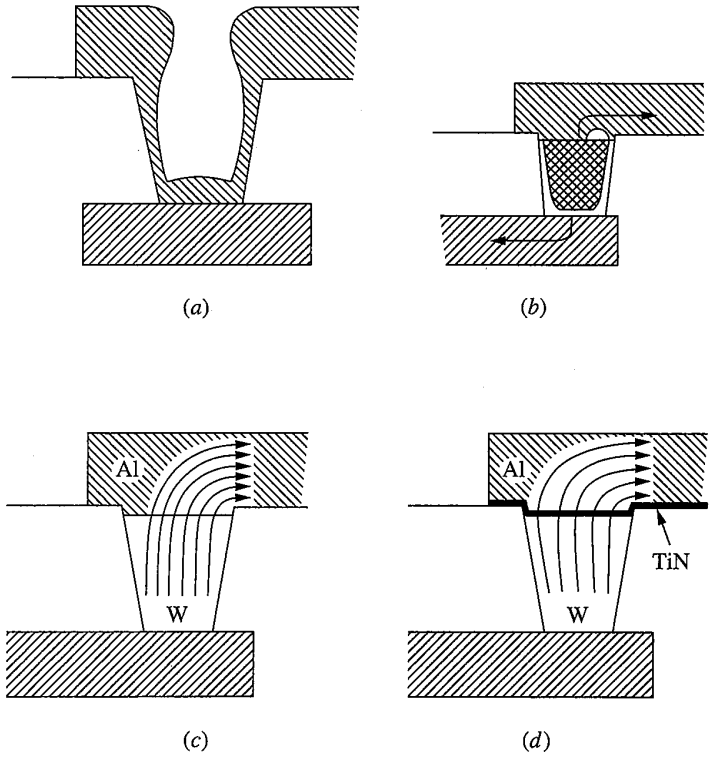


FIGURE 52
Contact electromigration: (*a*) Poor Al step coverage can cause high current density and Joule heating; (*b*) Al/W-plug contacts cause voiding at the corner of the Al/plug interface; (*c*) current crowding occurs at the inner corner of an Al/W-plug contact; (*d*) adding a TiN layer between the W plug and the Al wire reduces the current crowding.

Micron Ex. 1029, p. 478
Micron v. YMTC
IPR2024-00911

level of Al wire. This happens because the via intersects the Al wire at a right angle, and the electrical current has to turn 90° when entering the Al wire. As a result, various current paths offer different resistances, and current tends to go through the least resistive path—the inner corner, as shown in Fig. 52*b*. The current crowding drives more current through the inner corner of the plug and the adjacent Al wire. If the plug is made of Al, then no void will form since the Al flux leaving the high-current spot will be compensated by Al flux from inside the plug and from Al farther up the line. For a W plug, however, there is no Al flux that can come from the plug to replenish the Al flux leaving the hot spot. Consequently, voids will develop at the inner corner where current leaves the W plug into the Al wire, as shown in Fig. 52*c*.

There are two approaches to reduce the EM voiding at the via contact: (1) Replace the W plug with an Al plug to ensure the continuity of the Al flux, or (2) reduce the current crowding at the corner of the via contact. The second approach is illustrated in Fig. 52*d*. By applying a layer of TiN between the W plug and the Al wire, the current crowding at the inner corner can be reduced. The high resistance of the TiN helps to distribute the current from the via to a wider area in the Al wire, thus reducing the current crowding. To calculate the current distribution accurately three-dimensional analysis is needed, although the calculation is straightforward. A simple two-dimensional calculation is given in Problem 18 at the end of this chapter. Interestingly, the worst current crowding occurs when an Al plug is used. Voiding does not occur, however, because the Al flux through the via is continuous. Voids form from electromigration only when dissimilar materials are used and when current crowding occurs.

8.6.2 Metal Corrosion

Aluminum grows a passivating oxide in air and is naturally protected against corrosion. Aluminum wiring used in ULSI circuits, however, contains Cu, which has no passivating oxide, and the Al-Cu alloy used is more vulnerable to corrosion. The corrosion of Al wires comes from four sources: (1) Cl transported through the plastic packaging and passivation materials, (2) Cl from etching compound and etching by-products, (3) phosphorous acid formed from excess P in the phosphosilicate glass (PSG), and (4) electrochemical (galvanic) corrosion from dissimilar materials. Chlorine plays an important role in the corrosion of Al through these reactions:

$$Cl^- + H_2O \rightarrow HCl + OH^- \tag{8.57}$$

$$3HCl + Al \rightarrow AlCl_3 + 1.5H_2 \tag{8.58}$$

$$AlCl_3 + 3H_2O \rightarrow Al(OH)_3 + 3HCl \tag{8.59}$$

Note that the reactions in Eqs. (8.58) and (8.59) are cyclic. After the initial formation of HCl no additional Cl is needed. The presence of the Cl ion is only to facilitate the net reaction [from Eq. (8.58) and Eq. (8.59)]:

$$Al + 3H_2O \rightarrow Al(OH)_3 + 1.5H_2 \tag{8.60}$$

Therefore, only a small amount of Cl is needed to cause severe local corrosion of the Al lines, because the Cl^- ions recycle themselves during the corrosion process.

Micron Ex. 1029, p. 479
Micron v. YMTC
IPR2024-00911

Since most chemicals used for Al dry etching (RIE) contain Cl, $AlCl_3$ or a similar compound is formed on the Al surface after the plasma etching. Upon exposure to the moisture in air, the reaction in Eq. (8.59) and subsequently the cycling of Eq. (8.58) and Eq. (8.59) can proceed rapidly. Therefore, Al corrosion can occur very quickly after metal etching. To prevent etching-induced corrosion, chlorine compounds and elemental Cl must be removed from the metal surface immediately after plasma etching. Usually, a water rinse or a water vapor treatment is given right after etching. Even with this precaution corrosion can still occur from time to time. Modern metal-etching equipment now uses multiple chambers with an in-situ water rinse or water vapor treatment to ensure the removal of all Cl compounds.[181]

Phosphorus-doped glasses usually contain 4 to 5 wt % of P. P in the PSG can leach out if the concentration is higher than about 6 wt %. The P then interacts with the H_2O in the glass to form metaphosphoric acid (HPO_3), which subsequently etches the Al and causes corrosion. Note that the P content in PSG may vary locally in a wafer. Consequently, the P content can exceed 6 wt % locally even when the average P content is well below 4 to 5 wt %.

Modern metal structures use multilevels of dissimilar materials such as Ti/TiN/Al-Cu/TiN or Ti/Al-Cu/TiN, and the chance of electrochemical corrosion is increased. For example, the emf (electromotive force) between Al and W is about 1.5V, and in W-plug/Al-wiring structures the Al is in direct contact with W. In this kind of structure it is important that the dielectric material does not form an electrolyte to allow a galvanic reaction to occur. Therefore, water-absorbing materials should be avoided. Note that electrochemical corrosion can also be accelerated by the presence of Cl^- (from, e.g., plasma etching). In the presence of Cl^- and moisture, the galvanic action,

$$Al \rightarrow Al^{3+} + 3e^- \qquad \text{(Anode: Al)} \qquad (8.61)$$

$$3H^+ + 3e^- \rightarrow 1.5H_2 \qquad \text{(Cathode: W)} \qquad (8.62)$$

can proceed rapidly. Since Al is more electronegative than W it becomes the anode and is corroded away.

Copper is electropositive (relative to hydrogen) and is not vulnerable to electrochemical corrosion. However, in air copper oxide grows linearly with time, indicating the lack of a protective oxide. This lack of a passivating oxide makes Cu more vulnerable to chemical corrosion. Since there is practically no ULSI circuit made with Cu wiring today, experience in corrosion is nearly nonexistent. Most proposals for using Cu metallization involve some protective layers.[174]

8.7 SUMMARY AND FUTURE TRENDS

8.7.1 Summary

In summary, the demand for multilevel metallization has forced advances in all fronts, from film deposition to new architecture. Precise control of metal deposition and uniformity on large wafers is achieved by new designs of magnetron

Micron Ex. 1029, p. 480
Micron v. YMTC
IPR2024-00911

sputtering, and reactive sputtered and collimated deposition of Ti/TiN become standard processes. Polycides and self-aligned salicides provide more design freedom at low cost. CVD W has been widely adopted and will be followed by CVD TiN and CVD Cu for future applications. Plasma-enhanced CVD oxides are standard processes for 0.8-μm and 0.5-μm devices and will eventually be replaced by O_3-TEOS and HDP oxide for ILD gap fill. The advent of CMP and damascene builds the foundation for global planarization and totally planar structures and allows the building of four, five, and more levels of metals.

The rapid expansion of new applications in telecommunication, networked computation, multimedia processors, personal digital assistants, etc., in the next decade will no doubt further emphasize the need for high levels of integration. Higher levels of integration and high levels of performance will further complicate metallization while the conventional Al-alloy and SiO_2-dielectric approach their limits of *RC* delay and *IR* drop. Therefore, new materials and structures that can alleviate the parasitic limits and add higher sophistication to circuits are needed to ensure continued growth in the early 2000s.

8.7.2 New Conductor Materials

New metallization materials to reduce *RC* delay and *IR* drop or to improve electromigration resistance and thus allow the use of higher drive current are urgently needed. Since Al is already a good conductor the improvement in *R* seems limited. Ag, Cu, and Au are the only three metals that have higher electrical conductivity than Al. Fortunately, all three are heavier than Al and have better electromigration resistance, but even for Ag the improvement in *R* cannot exceed a factor of 2.

Cu is by far the most studied substitute for Al. CVD deposition of Cu has been discussed in Section 8.2.3. The use of Cu wiring for multilevels of interconnects has been demonstrated by dual damascene,[24,173,182,183] but the commercial application of this material to devices has not yet begun. Cu can diffuse through SiO_2, and a barrier layer, such as Si_3N_4, is needed. The most difficult obstacle, however, is its vulnerability to corrosion because Cu does not possess a tight, self-passivating oxide as does Al. To protect Cu from corrosion, passivation layers are needed that, unfortunately, take precious space and lower the conductance per unit volume. Consequently, Cu metallization can only reduce the effective *R* by 20 to 30%. Properly done, however, Cu can improve the electromigration resistance of the metal lines[184,185] and allow the use of higher current, which in turn improves the circuit speed. Investigations of Au[186] and Ag[187] for interconnects have been reported only recently. Much work is needed to make these new materials viable.

8.7.3 Low-Dielectric-Constant Materials

Low-permittivity insulators provide a larger margin for improving the *RC* delay than new conductors. In addition, a lower dielectric constant also reduces the line capacitance and thus cuts down the cross talk between conductors. An ideal low-permittivity dielectric, however, is still beyond reach today, because few materials

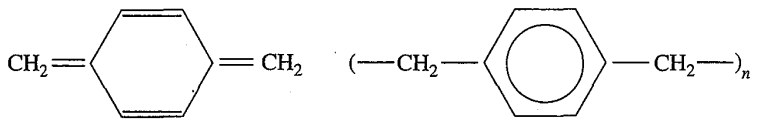
Micron Ex. 1029, p. 481
Micron v. YMTC
IPR2024-00911

can substitute for the properties of SiO_2. A dielectric material must meet at least the following criteria: (1) low permittivity, (2) high breakdown field, (3) low leakage, (4) no moisture absorption or excessive permeability to moisture, (5) thermal stability to $> 400°C$, (6) good adhesion to metal and to dielectric, (7) no metallic contaminants, (8) low defect density, (9) no significant outgassing, and (10) resistance to plasma. Unfortunately, few inorganic insulators have low permittivity, and few organic insulators show good thermal stability, good adhesion to metal, no outgassing or moisture absorption, and resistance to plasma.

Among the inorganic dielectric materials, BN has a low dielectric constant of 2.9. However, BN is hygroscopic[188] and is unsuitable as an ILD. SiBN with a composition of $Si_{0.10}B_{0.39}N_{0.51}$ has a dielectric constant of 3.6 and is stable in room air.[188] Other reliability factors for SiBN have not been fully studied yet. Recently, fluorinated silicon dioxide, or SiOF, has shown promise as a viable candidate.[189,190] SiOF has a dielectric constant of about 3.0 to 3.6 and can be deposited by PECVD[189] using TEOS and C_2F_6, or by a room temperature process.[190] The PECVD process is particularly attractive because it represents a modification of a process already widely used, with the benefit of a 15 to 30% lower dielectric constant.

Among the organic materials, most of the experience is with polyimides. These have been used extensively in packaging, multichip modules, and for some ULSI circuits as ILDs and as passivation. Reference 191 gives an overview of the experience and the issues on the applications of polyimides. With a dielectric constant of 3.0 to 3.6 and relatively good thermal and mechanical properties the application of polyimides to ULSI circuits is within reach. Some polyimides (e.g., PMDA-ODA from DuPont) can withstand temperatures to about 400°C without losing mechanical stability. Integration of polyimides with CVD W and Al-Cu wiring has been reported.[192] Multilevel metal structures using Cu/polyimide and dual damascene have been demonstrated.[24]

Organic materials with dielectric constants lower than 3.0 include parylene, BCB, and Teflon™. Parylene [poly-(*p*-xylylene)] has a dielectric constant of about 2.6 and can be vapor-deposited. Figure 53 shows the chemical structure of parylene, a polymer of *p*-xylylene. It consists of a benzene ring with 2 CH_2 radicals attached in the *para* positions. Parylene does not absorb moisture, but it is stable only to about 300°C. Fluorinated parylene (parylene-F) demonstrates a higher thermal stability to about 350 to 400°C. Investigation of this material for ULSI application has been reported recently.[193,194] Both the thermal stability and the resistance to plasma etching are promising. Much more characterization and probably synthesis of new materials will be needed to demonstrate full compatibility with ULSI processing. BCB, or



(*a*) (*b*)

FIGURE 53
Molecular structures for (*a*) para-xylylene, which is a precursor for parylene; (*b*) parylene (poly-*p*-xylylene) polymer.

Micron Ex. 1029, p. 482
Micron v. YMTC
IPR2024-00911

benzocyclobutane, has properties similar to parylene (no moisture absorption, dielectric constant of about 2.6, and stable to ~300°C). Chemicals related to the Teflon™ family have dielectric constants ≤ 2.0 but are thermally and mechanically less stable. In addition, Teflon has poor adhesion to metals. Therefore, this family of chemicals is hard to use for ILD. Recently, fluorinated polyimides with dielectric constants about 2.6 have been synthesized. Newly synthesized polymers, including some Si-containing polymers with dielectric constants below 3.0, have been listed in Ref. 195.

To summarize: the barrier to using dielectric materials with dielectric constants of 3.0 to 3.6 seems surpassable in a few years. Both organic polyimides and the inorganic SiOF may serve this purpose. To go below 3.0, the barrier is much higher. The existing polymers (commercially available) are probably all insufficient to fulfill the stringent requirements for ILD. A substantial number of new materials are being synthesized. The characterization and the trial application of these materials to ULSI circuits will be important.

8.7.4 New Contact Structures

The benefit of area reduction by stacked contacts and borderless contacts is substantial, rivaling that of design rule reduction. With the progress and acceptance of fully plugged contacts, structures with stacked contacts will become widely accepted. Completely borderless contacts are difficult to make, especially for small design rules. But partially relaxed border design rules that allow some contacts to be nearly borderless can still reduce the design area and increase circuit density. In the next decade, the stacked contact and variations of borderless contacts will be fully utilized to achieve higher packing density.

Packaging with arrays of pins instead of perimeters-only pins has already been designed. Because the pin count can now increase as the square of the die size, instead of linearly with the die size, this approach will alleviate the bonding pad limitation faced by many designs. In addition, the use of array bonding allows many signals to travel a shorter distance than in perimeter bonding and will improve circuit performance. Advances in packaging technology will further encourage the use of multilevel metallization.

8.7.5 Raised Source/Drain Structures

At 0.20 μm and below, silicides in the source/drain area are needed to reduce the contact resistance to an acceptable level (Section 8.3.4). However, it is difficult to make shallow junctions (0.10 μm or below) with silicide. Even if such junctions can be made, they will be so vulnerable[114] to stress, window RIE, etc. that it is impractical to make traditional contacts to them. Therefore, shallow junctions for 0.2-μm or below devices need to be made differently.

Raised source/drain junctions have been demonstrated.[196] By using selective growth of Si, 100 nm or more of epitaxially grown Si can be added on the

Micron Ex. 1029, p. 483
Micron v. YMTC
IPR2024-00911

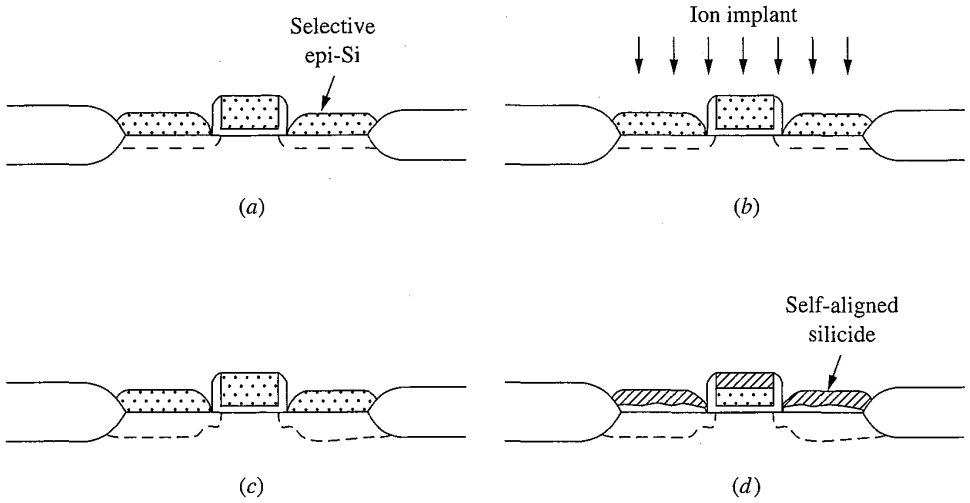


FIGURE 54
Raised source/drain junctions: (*a*) Selective growth of epitaxial Si on the source/drain; (*b*) implant of dopants into the epi-Si; (*c*) driving the dopants in to form the junction; (*d*) formation of salicide on the epi-Si and on poly-Si.

source/drain, as shown in Fig. 54*a*. Shallow junctions can then be formed by implanting into the epitaxial Si, as in Fig. 54*b*, and driving into the substrate, as in Fig. 54*c*. Finally, a salicide is formed on the epitaxial Si, Fig. 54*d*. Since the Si consumed for forming the salicide is from the epitaxial Si, the silicide does not erode the shallow junction. Because of the thicker Si, the sheet resistance of the source/drain diffusion is also lower. In addition, a thicker silicide can be used since the Si consumed does not come from the shallow junction. This allows a lower sheet resistance for the poly lines, for more design freedom. A drawback of raised source/drain junctions is the addition to the sidewall capacitance (Miller capacitance). To reduce this parasitic, a thicker sidewall spacer should be selected so that the Miller capacitance is only a small fraction of the gate capacitance.

8.7.6 Low-Cost Processing

The trend of using three, four, or more levels of metal wiring for ASICs and microprocessors started in the 1980s and will continue throughout the 1990s and into the next century. Multilayered Al structures are already widely used, and CMP of ILD and metal are beginning to be accepted as mainstream processes. The cost of processing the metallization structure is now comparable to the cost of processing CMOS devices. To keep the processing cost (per unit area) constant or to lower it, process complexity must be reduced.

The popular W plug and multilayer Al structure require repeated metal deposition and etching. For example, five metal depositions, two etchings, and one lithography are needed to complete one level of metal. Compare this to a simple Al wiring where only one metal deposition, one lithography, and one etching are used. Clearly,

Micron Ex. 1029, p. 484
Micron v. YMTC
IPR2024-00911

TABLE 2
Processing complexity of various multilevel metal structures

Structure	Process	Dielectric deposition	Metal deposition	RIE or planarization	Lithography	Total
Simple Al, non-plug	ILD deposition	2	0	1	0	3
	ILD planarization	0	0	1	0	1
	Via pattern	0	0	1	1	2
	Metal deposition	0	1	0	0	1
	Metal pattern	0	0	1	1	2
	Total =	2	1	4	2	9
Full-plug layered-metal	ILD deposition	2	0	1	0	3
	ILD planarization	0	0	1	0	1
	Via pattern	0	0	1	1	2
	Plug (barrier + CVD W + etch)	0	3	1	0	4
	Metal deposition	0	3	0	0	3
	Metal pattern	0	0	1	1	2
	Total =	2	6	5	2	15
Al-plug layered-metal	ILD deposition	2	0	1	0	3
	ILD planarization	0	0	1	0	1
	Via pattern	0	0	1	1	2
	Metal deposition	0	3	1	0	4
	ARC deposition	0	1	0	0	1
	Metal pattern	0	0	1	1	2
	Total =	2	4	5	2	13
Dual damascene-layered-metal	ILD deposition	1	0	0	0	1
	Trench pattern	0	0	1	1	2
	Via pattern	0	0	1	1	2
	Metal deposition	0	3	0	0	3
	Metal polish	0	0	1	0	1
	Total =	1	3	3	2	9

the current process is much too complex. If sputtered Al can also be used for the plug, then a much simpler process can be developed. However, since a barrier metal and an antireflective coating (ARC) are needed, the all-Al process only saves one metal deposition and one etching. Dual damascene, on the other hand, can reduce the cost by eliminating the metal lithography and the ARC layer. Table 2 compares the processing complexity of several different approaches. If all processes are mature (i.e., ILD and metal CMP can be done at low cost, and CVD metals are available), then dual damascene will provide the simplest process and the lowest cost.

Other cost reduction approaches are (1) high throughput equipment, (2) standardization of equipment, and (3) standardization of process modules. For example, using RTP and cluster tools, the cycle time can be reduced[197] substantially and fabrication cost is reduced. These measurements are more general than for just metallization. However, being a repetitive process, multilevel metallization will benefit the most from standardization.

Micron Ex. 1029, p. 485
Micron v. YMTC
IPR2024-00911

REFERENCES

1. K. K. Ng and W. T. Lynch, "The Impact of Intrinsic Series Resistance on MOSFET Scaling," *IEEE Trans. Electron Dev.,* **ED-34,** 503 (1987).
2. S. M. Sze, *Physics of Semiconductor Devices,* 2nd ed., Wiley, New York, 1981, p. 304.
3. R. Glang, "Vacuum Evaporation," in L. I. Maissel and R. Glang, eds., *Handbook of Thin Film Technology,* McGraw-Hill, New York, 1970, p. 1-3 to p. 1-130.
4. D. B. Fraser, "Metallization," in S. M. Sze, ed., *VLSI Technology,* McGraw-Hill, New York, 1983, pp. 350–357.
5. J. L. Vossen and J. J. Cuomo, "Glow Discharge Sputter Deposition," in J. L. Vossen and W. Kern, eds., *Thin Film Processes,* Academic, New York, 1978, pp. 12–75.
6. R. K. Watts, "Planar Magnetron Sputtering," in J. L. Vossen and W. Kern, eds., *Thin Film Processes,* Academic, New York, 1978, pp. 131–174.
7. J. A. Thornton, "Influence of Apparatus Geometry and Deposition Conditions on the Structure and Topography of Thick Sputtered Coatings," *J. Vac. Sci. Technol.,* **11,** 666 (1974).
8. L. I. Maissel, "Application of Sputtering to the Deposition of Films," in L. I. Maissel and R. Glang, eds., *Handbook of Thin Film Technology,* McGraw-Hill, New York, 1970, p. 4-1 to p. 4-44.
9. M. Hansen and K. Anderko, *Consititution of Binary Alloys,* McGraw-Hill, New York, 1958, pp. 132–134.
10. C. Y. Ting and M. Wittmer, "The Use of Titanium-Based Barrier Layers in Silicon Technology," *Thin Solid Films,* **96,** 327 (1982).
11. B. N. Chapman, *Glow Discharge Processes,* Wiley, New York, 1980, p. 369.
12. M. Wittmer, "Properties and Microelectronic Applications of Thin Films of Refractory Metal Nitrides," *J. Vac. Sci. Technol.,* **A-3,** 1797 (1985).
13. S. M. Rossnegal and D. Mikalsen, "Collimated Magnetron Sputter Deposition," *J. Vac. Sci. Technol.,* **A-9,** 261 (1991).
14. R. V. Joshi and S. Brodsky, "Collimated Sputtering of TiN/Ti Liners into Sub-Half-Micron High Aspect Ratio Contacts/Lines," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 253.
15. F. H. Baumann, R. Liu, C. B. Case, and W. Y.-C. Lai, "Application and Constraints of Collimated Ti/TiN Sputtering for 0.25 μm and sub-0.25 μm Contact Structures," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 412.
16. W. M. Bolton, R. E. Breining, and F. O. Deutscher, "Deposition of Aluminum," U.S. Patent 2,990,295, 1961.
17. M. L. Green, R. A. Levy, R. G. Nuzzo, and E. Coleman, "Al Films Prepared by Metal-Organic Low Pressure Chemical Vapor Deposition," *Thin Sol. Films,* **114,** 367 (1984).
18. H. W. Piekaar, L. F. Tz. Kwakman, and E. H. A. Granneman, "LPCVD of Aluminum in a Batch-Type Load-Locked Multi-Chamber Processing System," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 122.
19. W. Y.-C. Lai, K. P. Cheung, D. P. Favreau, C. J. Case, and R. Liu, "CVD Aluminum for Submicron VLSI Metallization," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnection Conference,* 1991, p. 89.
20. T. Ohta, N. Takeyasu, E. Kondoh, Y. Kawano, and H. Yamamoto, "New Structure of Selective-CVD-Al for Submicron Plug Applications and Process Using UHV Cluster Tool," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 329.

Micron Ex. 1029, p. 486
Micron v. YMTC
IPR2024-00911

21. R. W. Moshier, R. E. Sievers, and L. B. Spendlove, "Vapor-Plating Metals from Fluorocarbon Keto Metal Compounds," U.S. Patent 3,356,527, 1967.
22. M. E. Gross, "A Thermoanalytical Survey of Precursors for Copper Metal-Organic Chemical Vapor Deposition," *J. Electrochem. Soc.,* **138,** 2422 (1991).
23. A. V. Gelatos, S. Poon, R. Marsh, C. J. Mogab, and M. Thompson, "CVD of Copper From a Cu(+1) Precursor and Water Vapor and Formation of TiN-Encapsulated Submicron Copper Interconnects by Chemical-Mechanical Polishing," *1993 Symposium on VLSI Technology—Digest of Technical Papers,* 1993, p. 123.
24. B. Luther, J. F. White, C. Uzoh, T. Cacouris, J. Hummel, W. Guthrie, et al., "Planar Copper-Polyimide Backend of the Line Interconnections for ULSI Devices," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 15.
25. B. L. Crowder and S. Zirinsky, "1 μm MOSFET VLSI Technology: Part VII—Metal Silide Interconnection Technology—A Future Perspective," *IEEE Trans. Electron Dev.,* **ED-26,** 369 (1979).
26. K. C. Saraswat, D. L. Brors, J. A. Fair, K. A. Monnig, and R. Beyers, "Properties of Low Pressure CVD Tungsten Silicide for MOS VLSI Interconnections," *IEEE Trans. Electron Dev.,* **ED-30,** 1497 (1983).
27. T. Ohba, S. Inoue, and M. Maeda, "Selective CVD Tungsten Silicide for VLSI Applications," *Tech. Dig. IEDM,* 213 (1987).
28. A. Kohlhase, M. Mändl, and W. Palmer, "Performance and Failure Mechanisms of TiN Diffusion Barrier Layers in Submicron Devices," *J. Appl. Phys.,* **65,** 2464 (1989).
29. J. Laimer, H. Stori, and P. Rodhammer, "Plasma Assisted Chemical Vapor Deposition of Titanium Nitride in a Capacitively Coupled Radio-Frequency Discharge," *J. Vac. Sci. Technol.,* **A-7,** 2952 (1989).
30. A. Sherman, "Growth and Properties of LPCVD Titanium Nitride as a Diffusion Barrier for Silicon Device Technology," *J. Electrochem. Soc.,* **137,** 1892 (1990).
31. E. O. Travis, W. M. Paulson, F. Pintchovski, B. Boeck, L. C. Parrillo, M. L. Kottke, K.-Y. Fu, M. J. Rice, J. B. Price, and E. C. Eichman, "A Scalable Submicron Contact Technology Using Conformal LPCVD TiN," *Tech. Dig. IEDM,* 47 (1990).
32. J. J. Cuomo, "The Kinetics of Chemical Vapor Deposition of Tungsten," *Proceedings of the 3rd International Conference Chemical Vapor Deposition,* 1972, pp. 270–291.
33. P. A. Gargini and I. Beinglass, "WOS: Low Resistance Self-Aligned Source, Drain and Gate Transistors," *1981 Tech. Dig. IEDM,* 54 (1981).
34. E. K. Broadbent and C. L. Ramiller, "Selective Low Pressure Chemical Vapor Deposition of Tungsten," *J. Electrochem. Soc.,* **131,** 1427 (1984).
35. M. L. Green and R. A. Levy, "Structure of Selective Low Pressure Chemically Vapor-Deposited Films of Tungsten," *J. Electrochem. Soc.,* **132,** 1243 (1985).
36. S. Mehta, S. Mittal, A. Haranahalli, and D. Ranadive, "Blanket CVD Tungsten Interconnect for VLSI Devices," *Proceedings of the 3rd International IEEE VLSI Multilevel Interconnection Conference,* 1986, p. 418.
37. A. S. Sachdev, J. A. Fair, C. Fuhs, and W. Coney, "Blanket Tungsten Applications in VLSI Processing," in R. S. Blewer, ed., *Tungsten and Other Refractory Metals for VLSI Applications,* Materials Research Society, Pittsburgh, PA, 1987, pp. 69–72.
38. P. E. Riley, T. E. Clark, E. F. Gleason, and M. M. Garver, "Implementation of Tungsten Metallization in Multilevel Interconnection Technologies," *IEEE Trans. Semi. Manuf.,* **3,** 150 (1990).
39. C. M. McConica and K. Krishnamani, "The Kinetics of LPCVD Tungsten Deposition in a Single Wafer Reactor," *J. Electrochem. Soc.,* **133,** 2542 (1986).
40. JANAF Thermochemical Tables, 3rd ed., *J. Phys. Chem. Ref. Data,* **14,** Suppl. 1, (1985). D. R. Bradbury, I. E. Turner, K. Nauka, and K. Y. Chiu, "Selective CVD

Micron Ex. 1029, p. 487
Micron v. YMTC
IPR2024-00911

Tungsten as an Alternative to Blanket Tungsten for Submicron Plug Applications on VLSI Circuits," *Tech. Dig. IEDM,* 273 (1991).

41. G. E. Georgiou, J. M. Brown, M. L. Green, R. Liu, D. S. Williams, and R. S. Blewer, "The Influence of Selective Tungsten Deposition on Shallow Junction Leakage," in V. A. Wells, ed., *Tungsten and Other Refractory Metals for VLSI Applications II,* Materials Research Society, Pittsburgh, PA, 1987, p. 227.
42. R. V. Joshi, K. Y. Ahn, and P. Fryer, "The Role of Process Parameters Selective W Deposited by SiH_4, H_2, and WF_6 Chemistry in Terms of Shallow Junction Leakages," in R. S. Blewer and C. M. McConica, eds., *Tungsten and Other Refractory Metals for VLSI Applications IV,* Materials Research Society, Pittsburgh, PA., 1989, pp. 85–92.
43. S. Kang, R. Chow, R. H. Wilson, B. Gorowitz, and A. G. Williams, "Application of Selective CVD W for Low Contact Resistance Via Filling to Aluminum Multilayer Interconnection," *J. Electron. Matls.,* **17,** 213 (1988).
44. Y. Kusumoto, K. Takakuwa, H. Hashinokuchi, T. Ikuta, and I. Nakayama, "A New Approach to the Suppression of Tunneling," in V. A. Wells, ed., *Tungsten and Other Refractory Metals for VLSI Applications II,* Materials Research Society, Pittsburgh, PA, 1987, pp. 103–109.
45. N. Yokoyama, K. Hinode, and Y. Homma, "LPCVD Titanium Nitride for ULSIs," *J. Electrochem. Soc.,* **138,** 190 (1991).
46. M. J. Buiting, A. F. Otterloo, and A. H. Montree, "Kinetical Aspects of the LPCVD of Titanium Nitride from Titanium Tetrachloride and Ammonia," *J. Electrochem. Soc.,* **138,** 500 (1991).
47. T. Suzuki, T. Ohba, Y. Furumura, and H. Tsutikawa, "LPCVD-TiN Using Hydrazine and $TiCl_4$," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 418.
48. K. Ishihara, K. Yamazaki, H. Hamada, K. Kamisako, and Y. Tarui, "Characterization of CVD TiN Films Prepared with Metal-Organic Source," *Jpn. J. Appl. Phys. 1,* **29,** 2103 (1990).
49. R. M. Fix, R. G. Gordon, and D. M. Hoffman, "Synthesis of Thin Films by Atmospheric Pressure Chemical Vapor Deposition Using Amido and Imido Titanium (IV) Compounds as Precursors," *Chem. Mater.,* **2(3),** 235 (1990).
50. I. J. Raaijmakers, R. N. Vrtis, G. S. Sandhu, J. Yang, E. K. Broadbent, D. A. Roberts, and A. Lagendijk, "Conformal Deposition of TiN at Low Temperature by Metal Organic CVD," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 260.
51. L. H. Dubois, B. R. Zegarski, and G. S. Girolami, "Infrared Studies of the Surface and Gas Phase Reactions Leading to the Growth of Titanium Nitride Thin Films from Tetrakis(dimethylamido)titanium and Ammonia," *J. Electrochem. Soc.,* **139,** 3603 (1992).
52. J. T. Hillman, M. J. Rice, Jr., D. W. Studiner, R. F. Foster, and R. W. Fiordalice, "Comparison of Titanium Nitride Barrier Layer Produced by Inorganic and Organic CVD," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 246.
53. K. A. Littau, M. Eizenberg, S. Ghanayem, H. Tran, Y. Maeda, A. Sinha, M. Chang, G. Dixit, M. K. Jain, M. F. Chrisholm and R. H. Haveman, "CVD TiN: A Barrier Metallization for Sub-Micron Via and Contact Applications," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 440.
54. N. J. Ianno, A. U. Ahmed, and D. E. Englebert, "Plasma Assisted Chemical Vapor Deposition of TiN from $TiCl_4/N_2/H_2$ Mixture," *J. Electrochem. Soc.,* **136,** 276 (1989).
55. T. Akahori and A. Tanihara, "Preparation of TiN Films by ECR Plasma CVD," *Proceedings of the International Conference on Solid State Devices and Materials,* Yokohama,

Micron Ex. 1029, p. 488
Micron v. YMTC
IPR2024-00911

1991, pp. 180–182. T. Akahori, A. Tanihara, and M. Tano, "Preparation of TiN Films by Electron Cyclotron Resonance Plasma Chemical Vapor Deposition," *Jpn. J. Appl. Phys.*, **30,** 3558 (1991). ——, "TiN/Ti Films Formation Using ECR Plasma CVD," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 405.

56. A. Intemann, H. Koerner, and F. Koch, "Film Properties of CVD Titanium Nitride Deposited with Organometallic Precursors at Low Pressure Using Inert Gases, Ammonia, or Remote Activation," *J. Electrochem. Soc.,* **140,** 3215 (1993).
57. A. Weber, R. Nikulski, C.-P. Klages, M. E. Gross, W. L. Brown, E. Dons, and R. M. Charatan, "Low Temperature Deposition of TiN Using Tetrakis(dimethylamido)-Titanium in an Electron Cyclotron Resonance Plasma Process," *J. Electrochem. Soc.,* **141,** 849 (1994).
58. J. T. Hillman, R. F. Foster, J. Faguet, R. Arora, M. S. Ameen, C. Arena, and F. Martin, "Titanium Chemical Vapor Deposition," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 365.
59. N. Awaya and Y. Arita, "Selective Chemical Vapor Deposition of Copper," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 103 (1989). Y. Arita, N. Awaya, K. Ohno, and M. Sato, "CVD Copper Metallurgy for ULSI Interconnections," *Tech. Dig. IEDM,* 39 (1990).
60. B. Lecohier, B. Calpini, J.-M. Philippoz, H. van den Bergh, D. Laub, and P. A. Buffat, "Copper Film Growth by Chemical Vapor Deposition," *J. Electrochem. Soc.,* **140,** 789 (1993).
61. D.-H. Kim, R. H. Wentorf, and W. N. Gill, "Film Growth Kinetics of Chemical Vapor Deposition of Copper from $Cu(HFA)_2$," *J. Electrochem. Soc.,* **140,** 3267 (1993).
62. H.-K. Shin, K.-M. Chi, M. J. Hampden-Smith, T. T. Kodas, J. D. Farr, and M. Paffett, "Hot-Wall Chemical Vapor Deposition of Copper from Copper I Compounds 2—Selective, Low-Temperature Deposition of Copper from Copper (I) β-Diketonate Compounds, (β-Diketonate)CuL_n, via Thermally Induced Disproportionate Reactions," *Chem. Mater.,* **4,** 788 (1992).
63. A. Jain, K.-M. Chi, T. T. Kodas, M. J. Hampden-Smith, J. D. Farr, and M. F. Paffet, "Chemical Vapor Deposition of Copper from (Hexafluoroacetylacetonato)(Alkyne)-Copper (I) Complex via Disproportionation," *Chem. Mater.,* **3,** 995 (1991).
64. J. A. T. Norman, B. A. Muratore, P. N. Dyer, D. A. Roberts, and A. K. Hochberg, "New OMCVD Precursors for Selective Copper Metallization," *J. de Phys. IV,* **1,** C2-271 (1991).
65. N. Awaya and Y. Arita, "The Characteristics of Blanket Copper CVD for Deep Submicron Via Filling," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 125 (1993).
66. A. Jain, K.-M. Chi, T. T. Kodas, and M. J. Hampden-Smith, "Chemical Vapor Deposition of Copper from Hexafluoroacetylacetonato Cu(I) Vinyl-Trimethylsilane," *J. Electrochem. Soc.,* **140,** 1434 (1993).
67. L. H. Dubois and B. R. Zegarski, "Selectivity and Copper Chemical Vapor Deposition," *J. Electrochem. Soc.,* **139,** 3295 (1992).
68. K. P. Cheung, C. J. Case, R. Liu, R. J. Schutz, R. S. Wagner, L. F. Tz. Kwakman, D. Huibregtse, H. W. Piekaar, and E. H. A. Granneman, "Improved CVD Al Deposition Using in-situ Sputtered Nucleation Layers," *Proceedings of the 7th International IEEE VLSI Multilevel Interconnection Conference,* 1990, p. 303.
69. L. F. Tz. Kwakman, D. Huibregtse, H. W. Piekaar, E. H. A. Granneman, K. P. Cheung, C. J. Case, W. Y.-C. Lai, R. Liu, R. J. Schutz, and R. S. Wagner, "The Incorporation of Cu in CVD Al by Diffusion from In-Situ Sputtered Sources," *Proceedings of the 7th International IEEE VLSI Multilevel Interconnection Conference,* 1990, p. 282.

Micron Ex. 1029, p. 489
Micron v. YMTC
IPR2024-00911

70. E. Kondoh, Y. Kawano, N. Takeyasu, T. Katagiri, H. Yamamoto, and T. Ohta, "Interconnection Formation by Simultaneous Copper Doping in Chemical Vapor Deposited Aluminum (Al-Cu CVD)," *Tech. Dig. IEDM,* 277 (1993).
71. T. Amazawa and Y. Arita, "A 0.25μm Via Plug Process Using Selective CVD Aluminum for Multilevel Interconnections," *Tech. Dig. IEDM,* 265 (1991).
72. T. Ohta, N. Takeyasu, E. Kondoh, Y. Kawano, and H. Yamamoto, "New Structure of Selective CVD Al Filling Direct on Al for Submicron Plug Applications and Process Using UHV Cluster Tool," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 329.
73. A. K. Sinha, W. S. Lindenberger, D. B. Fraser, S. P. Murarka, and E. N. Fuls, "MOS Compatibility of High-Conductivity $TaSi_2/n^+$ Poly-Si Gates," *IEEE Trans. Electron Dev.,* **ED-27,** 1425 (1980).
74. D. S. Williams, E. Coleman, and J. M. Brown, "Low Pressure Chemical Vapor Deposition of Tantalum Silicide," *J. Electrochem. Soc.,* **133,** 2637 (1986).
75. K. Hieber and F. Neppl, "Possible Applications of Tantalum Silicide for Very Large Scale Integration Technology," *Thin Sol. Films,* **140,** 131 (1986).
76. M. Suzuki, N. Kobayashi, K. Mukai, and S. Kondo, "Characterization of Silane-Reduced Tungsten Films Grown by CVD as a Function of Si Content," *J. Electrochem. Soc.,* **137,** 3213 (1990).
77. Y. Shioya, K. Ikegami, M. Maeda, and K. Yanagida, "High Temperature Stress Measurement on Chemical-Vapor-Deposited Tungsten Silicide and Tungsten Films," *J. Appl. Phys.,* **61,** 561 (1987).
78. T. Hara, T. Miyamoto, H. Hagiwara, E. I. Bromley, and W. R. Harshbarger, "Composition of Tungsten Silicide Films Deposited by Dichlorosilane Reduction of Tungsten Hexafluoride," *J. Electrochem. Soc.,* **137,** 2955 (1990).
79. S. G. Telford, M. Eizenberg, M. Chang, A. K. Sinha, and T. R. Gow, "Chemically Vapor Deposited Tungsten Silicide Films Using Dichlorosilane in a Single-Wafer Reactor—Growth, Properties, and Thermal Stability," *J. Electrochem. Soc.,* **140,** 3689 (1993).
80. P. K. Tedrow, V. Ilderem, and R. Reif, "Low Pressure Chemical Vapor Deposition of Titanium Silicide," *Appl. Phys. Lett.,* **46,** 189 (1985).
81. J. Lee and R. Reif, "Plasma Enhanced Chemical Vapor Deposition of Blanket $TiSi_2$ on Oxide Patterned Wafers II—Silicide Properties," *J. Electrochem. Soc.,* **139,** 1166 (1992).
82. J. Lee and R. Reif, "Plasma Enhanced Chemical Vapor Deposition of Blanket $TiSi_2$ on Oxide Patterned Wafers I—Growth of Silicide," *J. Electrochem. Soc.,* **139,** 1159 (1992).
83. A. Bouteville, C. Attuyt, and J. C. Remy, "Selective RTLPCVD of $TiSi_2$ without Substrate Consumption," *Appl. Surf. Sci.,* **53,** 11 (1991).
84. J. Engqvist, C. Myers, and J.-O. Carlsson, "Selective Deposition of $TiSi_2$ from H_2-$TiCl_4$ Gas Mixtures and Si: Aspects of Thermodynamics Including Critical Evaluation of Thermodynamical Data in the Ti-Si System," *J. Electrochem. Soc.,* **139,** 3197 (1992).
85. K. Saito, T. Amazawa, and Y. Arita, "Selective Titanium Silicide Chemical Vapor Deposition with Surface Cleaning by Silane and Ohmic Contact Formation to Very Shallow Junctions," *J. Electrochem. Soc.,* **140,** 513 (1993).
86. G. E. Georgiou, P. E. Bechtold, H. S. Luftman, and T. T. Sheng, "Selective Electroless Plated Ni Contacts to CMOS Junctions with $CoSi_2$," *J. Electrochem. Soc.,* **138,** 3618 (1991).
87. P.-L. Pai and C. H. Ting, "Copper as the Future Interconnection Material," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 258.

Micron Ex. 1029, p. 490
Micron v. YMTC
IPR2024-00911

88. J. Van den Meerakker, "On the Mechanism of Electroless Plating I—Oxidation of Formaldehyde at Different Electrode Surfaces," *J. Appl. Electrochem.,* **11,** 387 (1981).
89. R. Contolini, S. Mayer, A. Bernhardt, and G. E. Georgiou, "A Copper Via Plug Process by Electrochemical Planarization," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 470.
90. S. P. Murarka, *Silicides for VLSI Applications,* Academic Press, Orlando, FL, 1983.
91. M.-A. Nicolet and S. Lau, "Silicides," in N. G. Einpruch and G. B. Larabee, eds., *VLSI Microstructure Science,* Vol. 6, Academic Press, NY, 1983, p. 329.
92. K. C. Saraswat, F. Mohammadi, and J. D. Meindl, "WSi_2 Gate MOS Devices," *Tech. Dig. IEDM,* 462 (1979).
93. T. Mochizuki, K. Shibata, T. Inoue, and K. Ohuchi, "A New MOS Process Using $MoSi_2$ as a Gate Material," *Jpn. J. Appl. Phys.,* **17-1,** 37 (1977).
94. M. E. Alperin, T. C. Holloway, R. A. Haken, C. D. Gosmeyer, R. V. Karnough, and W. D. Parmanto, "Development of the Self-Aligned Titanium Silicide Process for VLSI Applications," *IEEE Trans. Electron. Dev.,* **ED-32,** 141 (1985).
95. M. Delfino, E. K. Broadbent, A. E. Morgan, B. J. Burrow, and M. N. Norcott, "Formation of $TiN/TiSi_2/p^+/n^-$ Si by Rapid Thermal Annealing (RTA) Silicon Implanted with Boron through Titanium," *IEEE Electron. Dev. Lett.,* **EDL-6,** 591 (1985).
96. T. Brat, C. M. Osburn, T. Finstad, J. Liu, and B. Ellington, "Self-Aligned Ti Silicide Formed by Rapid Thermal Annealing," *J. Electrochem. Soc.,* **133,** 1451 (1986).
97. S. P. Murarka, D. B. Fraser, A. K. Sinha, J. J. Levinstein, E. J. Lloyd, R. Liu, D. S. Williams, and S. J. Hillenius, "Self-Aligned Cobalt Disilicide for Gate and Interconnection and Contacts to Shallow Junctions," *IEEE Trans. Electron Dev.,* **ED-34,** 2108 (1987).
98. L. Van den Hove, R. Wolters, K. Maex, R. F. De Keersmaecker, and G. J. Declerck, "A Self-Aligned $CoSi_2$ Interconnection and Contact Technology for VLSI Applications," *IEEE Trans. Electron. Dev.* **ED- 34,** 554 (1987).
99. J. B. Lasky, J. S. Nakos, O. J. Cain, and P. J. Geiss, "Comparison of Transformation to Low-Resistivity Phase and Agglomeration of $TiSi_2$ and $CoSi_2$," *IEEE Trans. Electron Dev.,* **ED-38,** 262 (1991).
100. I. Sakai, H. Abiko, H. Kawaguchi, T. Hirayama, L. E. G. Johansson, and K. Okabe, "A New Salicide Process (PASET) for Sub-Half Micron CMOS," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 66 (1992).
101. T. Horiuchi, H. Wakabayashi, T. Ishigami, H. Nakamura, T. Mogami, T. Kunio, and K. Okumura, "A New Titanium Salicide Process (DIET) for Sub-Quarter Micron CMOS," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 121 (1994).
102. Y. Matsubara, M. Sekine, N. Nishio, T. Shimura, K. Noguchi, T. Horiuchi, Y. Yamada, and T. Kitano, "A Novel Low Resistance Salicide Technology (SWAN) for Quarter-Micron CMOS," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 103 (1993).
103. M. Sekine, K. Inoue, H. Ito, I. Honma, H. Miyamoto, K. Yoshida, H. Watanabe, K. Mikagi, Y. Yamada, and T. Kikkawa, "Self-Aligned Tungsten Strapped Source/Drain and Gate Technology Realizing the Lowest Sheet Resistance for Sub-Quarter Micron CMOS," *Tech. Dig. IEDM,* 493 (1994).
104. K. Goto, T. Yamazaki, A. Fushida, S. Inagaki, and H. Yagi, "Optimization of Salicide Processes for Sub-0.1μm CMOS Devices," *Dig. Tech. Papers, Symp. on VLSI Technol.* 119, (1994).
105. W. M. Chen, J. Lin, and J. C. Lee, "A Novel $CoSi_2$ Thin Film Process with Improved Thickness Scalability and Thermal Stability," *Tech. Dig. IEDM,* 691 (1994).
106. T. E. Tang, C.-C. Wei, R. A. Haken, T. C. Holloway, L. R. Hite, and T. G. W. Blake, "Titanium Nitride Local Interconnect Technology for VLSI," *IEEE Trans. Electron Dev.,* **ED-34,** 682 (1987).

Micron Ex. 1029, p. 491
Micron v. YMTC
IPR2024-00911

107. D. C. Chen, S. S. Wong, P. V. Voorde, P. Merchant, T. R. Cass, J. Amano, and K. Y. Chiu, "A New Device Interconnect Scheme for Submicron VLSI," *Tech. Dig. IEDM,* 118 (1984).
108. K. K. Young, H. K. Hu, C. H. Lin, D. Peters, and K. Y. Chiu, "TiW Local Interconnect Technology for Advanced ULSI," *International Electronic Devices and Materials Symposium, Taipei, R.O.C.,* 1990, p. 231.
109. R. Beyers, D. Coulman, and P. Merchant, "Titanium Disilicide Formation on Heavily Doped Silicon Substrates," *J. Appl. Phys.,* **61,** 5110 (1987).
110. R. Liu, D. S. Williams, and W. T. Lynch, "A Study of the Leakage Mechanisms of Silicided n^+/p Junctions," *J. Appl. Phys.,* **63,** 1990 (1988).
111. R. Liu, F. A. Baiocchi, L. A. Heimbrook, J. Kovalchick, D. L. Malm, D. S. Williams, and W. T. Lynch, "Formation of Shallow p^+/n and n^+/p Junctions with $CoSi_2$," in *Proceedings of the 1st International Symposium on ULSI Science and Technology,* S. Broydo and C. M. Osburn, eds., Vol. 87-11, The Electrochemical Society, Pennington, NJ, 1987, pp. 446–462.
112. V. Probst, H. Schaber, P. Lippens, L. Van den Hove, and R. De Keersmaecker, "Limitation of $TiSi_2$ as a Source for Dopant Diffusion," *Appl. Phys. Lett.,* **52,** 1803 (1988).
113. C. Y. Lu, J. M. Sung, R. Liu, N. S. Tsai, R. Singh, S. J. Hillenius, and H. C. Kirsch, "Process Limitation and Device Design Trade-Offs of Self-Aligned $TiSi_2$ Junction Formation in Submicrometer CMOS Devices," *IEEE Trans. Electron Dev.* **ED-38,** 246 (1991). H. Jiang, C. M. Osburn, Z.-G. Xiao, G. McGuire, and G. A. Rozgonyi, "Ultra Shallow Junction Formation Using Diffusion from Silicides III—Diffusion into Silicon, Thermal Stability of Silicides, and Junction Integrity," *J. Electrochem. Soc.,* **139,** 211 (1992).
114. R. Liu, D. S. Williams, and W. T. Lynch, "Mechanisms for Process-Induced Leakage in Shallow Silicided Junctions," *Tech. Dig. IEDM,* 58 (1986).
115. K. C. Saraswat, S. Swirhun, and J. P. McVittie, "Selective CVD of W for VLSI Technology," *Proceedings of the 2nd International Symposium on VLSI Science and Technology,* J. E. Bean and G. A Rozgonyi, eds., Vol. 84-7, The Electrochemical Society, Pennington, NJ, 1984, pp. 409–419.
116. D. R. Bradbury and T. I. Kamins, "Effect of Insulator Surface on Selective Deposition of CVD W Films," *J. Electrochem. Soc.,* **133,** 1214 (1986).
117. K. K. Choi, S. B. Hwang, H. L. Park, and C. G. Ko, "A New Selective W-CVD Process Using Poly-Si Glue Layer," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 286.
118. A. Kamgar, R. V. Knoell, F. A. Baiocchi, K. J. Orlowsky, K. P. Cheung, and R. Liu, "Impact of Al Melting on Diode Integrity," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 190.
119. R. Mukai, N. Sasaki, and M. Nakano, "High-Aspect-Ratio Via-Hole Filling with Aluminum Melting by Excimer Laser Irradiation for Multilevel Interconnection," *IEEE Electron Dev. Lett.,* **EDL-8,** 76 (1987).
120. D. Pramanik and S. Chen, "Characterization of Laser Planarized Aluminum for Submicron Double Level Metal CMOS Circuits," *Tech. Dig. IEDM,* 673 (1989).
121. W. Y.-C. Lai, R. Liu, K. P. Cheung, and R. Heim, "The Use of Ti as an Antireflective Coating for the Laser Planarization of Al for VLSI Metallization," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 501, ——, "A Study of Pulsed Laser Planarization of Aluminum for VLSI Metallization," Ibid., 1989, p. 329.
122. C. S. Park, S. I. Lee, J. H. Park, J. H. Sohn, D. Chin, and J. G. Lee, "Al-PLAPH (Aluminum Planarization by Post-Heating) Process for Planarized Double Metal CMOS

Micron Ex. 1029, p. 492
Micron v. YMTC
IPR2024-00911

Applications," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnection Conference,* 1991, p. 326.
123. F. S. Chen, Y. S. Liu, G. A. Dixit, R. Sundaresan, C. C. Wei, and F. T. Liou, "Planarized Aluminum Metallization for Sub-0.5 μm CMOS Technology," *Tech. Dig. IEDM,* 51 (1990).
124. M. Ionoue, K. Hashizume, and H. Tsuchikawa, "The Properties of Aluminum Thin Films Sputter Deposited at Elevated Temperature," *J. Vac. Sci. Technol.,* **A-6,** 1636 (1988).
125. C. J. Case and R. Liu, unpublished.
126. K. Kikuta, T. Kikkawa, and H. Aoki, "0.25 μm Contact Hole Filling By Al-Ge Reflow Sputtering," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 35 (1991).
127. G. A. Dixit, M. F. Chrisholm, M. K. Jain, T. Weaver, L. M. Ting, S. Pourch, K. Mizobuchi, R. H. Havemann, C. D. Dobson, A. I. Jeffryes, P. J. Holverson, P. Rich, D. C. Butler, and J. Hems, "A Novel High Pressure Low Temperature Aluminum Plug Technology for Sub-0.5μm Contact/Via Geometries," *Tech. Dig. IEDM,* 105 (1994).
128. C. S. Wei, V. Mureli, M. L. A. Dass, and D. B. Fraser, "The Use of Selective Silicide Plugs for Submicron Contact Fill," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 136.
129. T. Iijima, A. Nishiyama, Y. Ushiku, T. Ohguro, I. Kunishima, K. Suguro, and H. Iwai, "A Novel Selective Ni_3Si Contact Plug Technique for Deep-Submicron ULSIs," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 70 (1992).
130. A. C. Adams and C. D. Capio, "The Deposition of Silicon Dioxide Films at Low Pressure," *J. Electrochem. Soc.,* **126,** 1042 (1979). A. C. Adams, "Dielectric and Polysilicon Film Deposition," in S. M. Sze, ed., *VLSI Technology,* McGraw-Hill, NY, 1983, p. 107.
131. G. C. Schwartz and P. Johns, "Gap-Fill with PECVD Silicon Dioxide Using Deposition/Sputter-Etch Cycles," *Proceedings of the 3rd International Symposium on ULSI Science and Technology,* The Electrochemical Society, Pennington, NJ, 1991, pp. 720–729.
132. K. Maeda, Y. Nishimoto, N. Tokumasu, and K. Fujino, "Dielectric Film Deposition by Atmospheric Pressure and Low Temperature CVD Using TEOS, Ozone, and New Organometallic Doping Sources," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 382.
133. K. Fujino, Y. Nishimoto, N. Tokumasu, and K. Maeda, "Dependence of Deposition Characteristics on Base Materials in TEOS and Ozone CVD at Atmospheric Pressure," *J. Electrochem. Soc.,* **138,** 550 (1991).
134. K. Fujino, Y. Nishimoto, N. Tokumasu, and K. Maeda, "Silicon Dioxide Deposition by Atmospheric Pressure and Low-Temperature CVD Using TEOS and Ozone," *J. Electrochem. Soc.,* **137,** 2883 (1990).
135. J. Huang, K. Kwok, D. Witty, and K. Donohoe, "Dependence of Film Properties of Subatmospheric Pressure Chemical Vapor Deposited Oxide on Ozone-to-Tetraethylorthosilicate Ratio," *J. Electrochem. Soc.,* **140,** 1682 (1993).
136. J. A. Mucha and J. Washington, "Infrared Studies of Ozone-Organosilicon Chemistries for SiO_2 Deposition," *Mat. Res. Conf. Proc.,* **334,** 31 (1994).
137. J. A. Mucha, "The Effects of Ethoxy-Groups on the Kinetics of O_3-Induced SiO_2 Deposition," submitted for publication (1995).
138. See for example C. J. Adkins, *Equilibrium Thermodynamics,* McGraw-Hill, London, pp. 211–216, 1968.
139. L. Forester, A. L. Butler, and G. Schets, "SOG Planarization for Polysilicon and First Metal Interconnect in a One Micron CMOS Process," *Proceedings of the 6th International IEEE VLSI Multilevel Interconnection Conference,* 1989, p. 72.
140. L. Forester, H. Meynen, and B. Coenegrachts, "Characterization of a Direct-on-Metal Spin-on Glass Planarization Process," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 175.

Micron Ex. 1029, p. 493
Micron v. YMTC
IPR2024-00911

141. B. T. Ahlburn and K. A. Scheibert, "A Non-Etchback Spin On Glass for 0.5 μm Devices Using Hydrogen Silsesquioxane as a Replacement for Methylsiloxane," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 120.
142. L.-J. Chen, S.-T. Hsia, and J.-L. Leu, "Fluorine-Implanted-Treatment (FIT) SOG for the Non-Etchback Intermetal Dielectric," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 81.
143. C. Y. Ting. V. J. Vivalda, and H. G. Schaefer, "Study of Planarized Sputter-Deposited SiO_2," *J. Vac. Sci. Technol.,* **15(3),** 1105 (1978).
144. S. Matsuo and M. Kiuchi, "Low Temperature Chemical Vapor Deposition Method Utilizing an Electron Cyclotron Resonance Plasma," *Jpn. J. Appl. Phys. Lett.,* **22,** L210 (1983).
145. K. M. Kerney, "ECR Finds Applications in CVD," *Semiconductor International,* 67 (Mar. 1989).
146. C. S. Pai, J. M. Miner, and P. D. Foo, "Electron Cyclotron Resonance Microwave Discharge for Oxide Deposition Using Tetraethoxysilane," *J. Electrochem. Soc.,* **139,** 850 (1992).
147. R. Chebi and S. Mittal, "A Manufacturable ILD Gap Fill Process with Biased ECR CVD," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnection Conference,* 1991, p. 61.
148. B. M. Somero, R. P. Chebi, E. O. Travis, H. B. Haver, and W. K. Morrow, "An Integrated, Low Temperature Interlayer Dielectric Using ECR Technology for Advanced CMOS ASIC and Logic Applications," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 72.
149. A. Bose, M. M. Garver, and R. A Spencer, "Advanced Inter-Metal Dielectric Deposition—A Comparative Analysis Between ECR-CVD and O_3/TEOS," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 89.
150. N. Lifshitz and G. Smolinsky, "Water-Related Charge Motion in Dielectrics," *J. Electrochem. Soc.,* **136,** 2335 (1989). ——, "Hot Carrier Aging of the MOS Transistors in the Presence of Spin-On Glass as the Interlevel Dielectric," *IEEE Electron Dev. Lett.,* **EDL-12,** 140 (1991).
151. S. Wolf, *Silicon Processing for the VLSI Era, Volume 2—Process Integration,* Lattice Press, Sunset Beach, CA, 1990, pp. 199–201.
152. T. H. Daubenspeck, J. K. DeBrosse, C. W. Koburger, M. Armcost, and J. R. Abernathey, "Planarization of ULSI Topography over Variable Pattern Densities," *J. Electrochem. Soc.,* **138,** 506 (1991).
153. A. Schiltz and M. Pons, "Two-Layer Planarization Process," *J. Electrochem. Soc.,* **133,** 178 (1986).
154. B. Davari, C. W. Koburger, R. Schulz, J. D. Warnock, T. Furukawa, M. Jost, Y. Taur, W. G. Schwittek, J. K. DeBrosse, M. L. Kerbaugh, and J. L. Mauer, "A New Planarization Technique Using a Combination of RIE and Chemical Mechanical Polish (CMP)," *Tech. Dig. IEDM,* 61 (1989).
155. L. M. Cook, "Chemical Processes in Glass Polishing," *J. Non Cryst. Solids,* **120,** 152 (1990).
156. S. Sivaram, R. Leggett, A. Maury, K. Monnig, and R. Tolles, "Overview of Planarization by Mechanical Polishing of Interlevel Dielectrics," *Proceedings of the 3rd International Symposium on ULSI Science and Technology,* J. M. Andrews and G. K. Celler, eds., Vol. 91-11, The Electrochemical Society, Pennington, NJ, 1991, pp. 607–616.
157. F. W. Preston, "The Theory and Design of Plate Glass Polishing Machines," *J. Soc. Glass Tech.,* **11,** 214 (1927).
158. J. Warnock, "A Two-Dimensional Process Model for Chemimechanical Polish Planarization," *J. Electrochem. Soc.,* **138,** 2398 (1991).

Micron Ex. 1029, p. 494
Micron v. YMTC
IPR2024-00911

159. S. R. Runnels, "Feature-Scale Fluid Based Erosion Modeling for Chemical-Mechanical Polishing," *J. Electrochem. Soc.*, **141,** 1900 (1994).
160. F. B. Kaufman, D. B. Thompson, R. E. Broadie, M. A. Jaso, W. L. Guthrie, D. J. Pearson, and M. B. Small, "Chemical-Mechanical Polishing for Fabricating Patterned W Metal Features as Chip Interconnects," *J. Electrochem. Soc.*, **138,** 3460 (1991).
161. W. J. Patrick, W. L. Guthrie, C. L. Standley, and P. M. Schiable, "Application of Chemical Mechanical Polishing to the Fabrication of VLSI Circuit Interconnections," *J. Electrochem. Soc.*, **138,** 1778 (1991).
162. I. Ali, S. R. Roy, and G. Shinn, "Chemical-Mechanical Polishing of Interlayer Dielectric: A Review," *Sol. State Technol.*, 63 (Oct. 1994).
163. R. Jairath, J. Farkas, C. K. Huang, M. Stell, and S-M. Tzeng, "Chemical-Mechanical Polishing: Process Manufacturability," *Sol. State Technol.*, 71 (July 1994).
164. W. Y.-C. Lai, G. L. Miller, R. J. Schutz, G. Smolinsky, and E. R. Wagner, "Capacitive End-Point Detection for Oxide Planarization by Chemical-Mechanical Polishing," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 147.
165. P. Singer, "Chemical-Mechanical Polishing: A New Focus on Consumables," *Semic. Intl.*, 48 (Feb. 1994).
166. H. Aoki, T. Nakajima, K. Kikuta, and Y. Hayashi, "Novel Electrolysis-Ionized-Water Cleaning Technique for Chemical-Mechanical Polishing (CMP) Process," *Dig. Tech. Papers, Symp. on VLSI Technol.*, 79 (1994).
167. E. K. Broadbent, J. M. Flanner, W. G. Wilbert, G. M. Van den Hoek, and I.-W. Huang Connick, "High Density High Reliability Tungsten Interconnection by Filled Interconnect Groove Metallization," *IEEE Trans. Electron Dev.*, **ED-35,** 952 (1988).
168. C. W. Kaanta, W. J. Cote, J. E. Cronin, K. Holland, P. Lee, and T. Wright, "Submicron Wiring Technology with Tungsten and Planarization," *Tech. Dig. IEDM,* 209 (1987).
169. C. W. Kaanta, S. G. Bombardier, W. J. Cote, W. R. Hill, G. Kerszykowski, H. S. Landis, D. J. Poindexter, C. W. Pollard, G. H. Ross, J. G. Ryan, S. Wolff, and J. E. Cronin, "Dual Damascene: A ULSI Wiring Technology," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnection Conference,* 1991, p. 144.
170. J. Givens, S. Geissler, O. Cain, W. Clark, C. Koburger, and J. Lee, "A Low-Temperature Local Interconnect Process in a 0.25μm-Channel CMOS Logic Technology with Shallow Trench Isolation," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 43.
171. K. D. Beyer, W. L. Guthrie, S. R. Markarewicz, E. Mendel, W. J. Patrick, K. A. Perry, W. A. Pliskin, J. Riseman, P. M. Schiable, and C. L. Standly, "Chemical-Mechanical Polishing Method for Producing Coplanar Metal/Insulator Films on a Substrate," U.S. Patent 4,944,836, 1990.
172. A. Krishnan, C. Xie, N. Kumar, J. Curry, D. Duane, and S. P. Murarka, "Copper Metallization for VLSI Applications," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 226.
173. N. Misawa, S. Kishii, T. Ohba, Y. Arimoto, Y. Furumura, and H. Tsutikawa, "High Performance Planarized CVD-Cu Multi-Interconnection," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 353.
174. H. M. Dalal, R. V. Joshi, H. S. Rathore, and R. Fillipi, "A Dual Damascene Hard Metal Capped Cu and Al-Alloy for Interconnect Wiring of ULSI Circuits," *Tech. Dig. IEDM,* 273 (1993).
175. Y. Hayashi, K. Kikuta, and T. Kikkawa, "A New Abrasive-Free, Chemical-Mechanical-Polishing Technique for Aluminum Metallization of ULSI Devices," *Tech. Dig. IEDM,* 976 (1992).

Micron Ex. 1029, p. 495
Micron v. YMTC
IPR2024-00911

176. T. Shinzawa, K. Sugai, A. Kobayashi, Y. Hayashi, T. Nakajima, S. Kishida, H. Okabayashi, T. Yako, K. Tsunenari, and Y. Murao, "Adhesion-Layerless Submicron Al Damascene Interconnection Using Novel Al-CVD," *Dig. Tech. Papers, Symp. on VLSI Technol.* 77 (1994).
177. K. Kikuta, Y. Hayashi, T. Nakajima, K. Harashima, and T. Kikkawa, "Aluminum-Germanium-Copper Multilevel Damascene Process Using Low Temperature Reflow Sputtering and Chemical Mechanical Polishing," *Tech. Dig. IEDM,* 101 (1994).
178. S. Roehl, L. Camiletti, W. Cote, E. Eckstein, K. H. Froehner, P. I. Lee, D. Restaino, G. Roeska, V. Vynorius, S. Wolff, and B. Vollmer, "High Density Damascene Wiring and Borderless Contacts for 64M DRAM,"*Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 22.
179. R. N. Hall, D. M. Brown, R. H. Wilson, and D. W. Skelly, "Electromigration Reliability Studies of Intermetal Contact Having CVD Tungsten Via Plugs," *Proc. Tungsten and Other Refractory Metals for VLSI Applications III,* V. A. Wells, ed., Materials Research Society, Pittsburgh, PA, 1987, pp. 231–237.
180. F. Matsuoka, H. Iaai, K. Hama, H. Itoh, R. Nakata, T. Nakakubo, K. Maoguchi, and K. Kanzaki, "Electromigration Reliability for Tungsten Plug Filled Via Hole Structure," *IEEE Trans. Electron Dev.,* **ED-37,** 562 (1990).
181. S. Fujimura, K. Shinagawa, M. T. Suzuki, and M. Nakamura, "Resist Stripping in an $O_2 + H_2O$ Plasma Downstream Reactor," *J. Vac. Sci. Technol.,* **A-9,** 357 (1991).
182. J. S. H. Cho, H. K. Kang, I. Asano, and S. S. Wang, "CVD Cu Interconnection for ULSI," *Tech. Dig. IEDM,* 297 (1992).
183. S. Lakshminarayanan, J. Steigerwald, D. Price, M. Bourgeois, T. P. Chow, R. J. Gutmann, and S. P. Murarka, "Dual Damascene Copper Metallization Process Using Chemical-Mechanical Polishing," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 49.
184. J. S. H. Cho, H.-K. Kang, C. Ryu, and S. S. Wong, "Reliability of CVD Cu Buried Interconnect," *Tech. Dig. IEDM,* 265 (1993).
185. T. Nitta, T. Ohmi, T. Hoshi, S. Sakai, K. Sakaibara, S. Imai, and T. Shibata, "Evaluating the Large Electromigration Resistance of Copper Interconnects Employing a Newly Developed Accelerated Life-Test Method," *J. Electrochem. Soc.,* **140,** 1131 (1993).
186. K. Urabe, T. Katoh, and Y. Murao, "Fully Planarized Multilevel Gold Metallization with Gap Filled Wires," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference,* 1993, p. 366.
187. Y. Ushiku, H. Ono, T. Iijima, N. Ninomiya, A. Nishiyama, H. Iwai, and H. Hara, "Planarized Silver Interconnect Technology with a Ti Self-Passivation Technique for Deep Submicron ULSIs," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 121 (1993).
188. M. Maeda, T. Makino, E. Yamamoto, and S. Konaka, "A Low Permittivity Interconnection Using an SiBN Interlayer," *IEEE Trans. Electron Dev.,* **ED-36,** 1610 (1989).
189. J. Ida, M. Yoshimaru, T. Usami, A. Ohtomo, K. Shimokawa, A. Kita, and M. Ino, "Reduction of Wiring Capacitance with New Low Dielectric SiOF Interlayer Film for High Speed/Low Power Sub-Half Micron CMOS," *Dig. Tech. Papers, Symp. on VLSI Technol.,* 59 (1994).
190. T. Homma, R. Yamaguchi, Y. Murao, "Room Temperature Chemical Vapor Deposition SiOF Film Formation Technology for the Interlayer in Submicron Multilevel Interconnections," *J. Electrochem. Soc.,* **140,** 687 (1993).
191. R. M. Geffken, "An Overview of Polyimide Use in Integrated Circuits and Packaging," in *Proceedings of the 3rd International Symposium on ULSI Science and Technology,* J. M. Andrews and G. K. Celler, eds., Vol. 91-11, The Electrochemical Society, Pennington, NJ, 1991, pp. 667–677.

Micron Ex. 1029, p. 496
Micron v. YMTC
IPR2024-00911

192. R. V. Joshi, L. Hsu, H. Dalal, P. Klymco, M. Jaso, and H. Ng, "A Novel Application of Polyimide-W-Al/Cu for VLSI Interconnect," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnection Conference,* 1991, p. 75.
193. S. Babrad, X. Zhang, B. J. Howard, C. Chiang, G. Cuan, K. Huang, R. Olson, H. Bakhru, C. Steinbruchel, T.-M. Lu, and J. F. McDonald, "Aliphatic Tetrafluorinated Parylene (PA-f) as a Conformal Insulator for Submicron Multilevel Interconnections," *Proceedings of the 9th International VLSI Multilevel Interconnection Conference,* 1992, p. 86.
194. X. Zhang, R. Tacito, D. S. Yaney, C. Chiang, C. Steinbruchel, and J. F. McDonald, "A Submicron Parylene Etch Process for Multilevel Interconnection," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference,* 1994, p. 87.
195. J. Paraszczak, D. Edelstein, S. Cohen, E. Babich, and J. Hummel, "High Performance Dielectrics and Processes for ULSI Interconnection Technologies," *Tech. Dig. IEDM,* 261 (1993).
196. C. Mazure, J. Fitch, and C. Gunderson, "Facet Engineered Elevated Source/Drain by Selective Epitaxy for 0.35 Micron MOSFETs," *Tech. Dig. IEDM,* 853 (1992).
197. M. Moslehi, L. Velo, A. Paranjpe, R. Chapman, Y. J. Lee, S. Huang, J. Kuehne, C. Schaper, T. Breedjik, D. Yin, D. Anderson, and D. Davis, "Microelectronic Manufacturing Science and Technology (MMST): Single-Wafer RTP-Based 0.35 μm CMOS IC Fabrication," *Tech. Dig. IEDM,* 649 (1993).

PROBLEMS

1. (*a*) Estimate the capacitance per unit length of two parallel Al wires 0.5 μm × 0.5 μm in cross section and separated by a SiO_2 dielectric layer $d = 0.5$ μm. Use $f = 1.5$ for fringe field effects ($\varepsilon_{ox} = 0.345$ pF/cm). (*b*) Estimate the capacitance per unit length for one Al wire at the distance $d = 0.5$ μm over a ground plane, where $f = 3$.

2. If we assume no external resistance, what is the "intrinsic" *RC* delay for the configuration in Problem 1(*b*), (*a*) for an Al wire length of 1 mm, and (*b*) for a wire of 2-mm length? The resistivity of Al is $\rho = 2.7$ μΩ-cm.

3. Show that the intrinsic *RC* delay for two parallel Al wires is independent of the metal height. How does the intrinsic *RC* scale with the scaling factor $S \approx \sqrt{2}$ (*a*) if the die size also scales down as $1/S$, (*b*) if the die size stays constant, and (*c*) if the die size scales up as S?

4. For an nMOSFET with $L = 0.35$ μm, $W = 1$ μm, and a gate oxide thickness = 9 nm, operating at $V_d = V_g = 3.3$ V, the on-current $I_{on} \approx 0.5$ mA is achieved. (*a*) If a $V_{dd} = 3.3$ V is applied to the drain of the MOSFET through a long 1 mm Al wire with a cross section of 0.5 μm × 0.5 μm, what is the V_d? Will the MOSFET operate in the saturation mode? (*b*) Do the same calculations for a wide MOSFET with $W = 10$ μm. (*c*) Do the same calculations for a wider MOSFET with $W = 40$ μm and a 2 mm Al wire. (Let $V_{dsat} = 0.8$ V; the resistivity for Al is $\rho = 2.7$ μΩ-cm.)

5. When a MOSFET is operating in the saturation region it can be considered as a current source. Let us define the node delay time as the time required to charge the capacitance of a node, $\Delta t_n = C_n \Delta V / i_n$, where C_n is the node capacitance, i_n the charging current, and $\Delta V = V_{dd}$ for full-swing operation. (*a*) Consider the case $C_n = 20$ fF, $V_{dd} = 3.3$

Micron Ex. 1029, p. 497
Micron v. YMTC
IPR2024-00911

V, and a MOSFET as in Problem 3 with $W = 10$ μm; what is the node delay time? (*b*) Consider the case with $W = 40$ μm, and V_{dd} is connected to the drain of the MOSFET through a 2 mm Al wire, as in Problem 3(*c*). Ignore the *RC* delay in the Al wire, and assume a $V_{sat} = 0.8$ V; what is the node delay time now? (*c*) What is the node delay time for $W = 80$ μm?

6. The rule of thumb for the electromigration limit in Al wiring is to restrict the current density below 1×10^5 A/cm^2. Consider the device and metallization parameters in the table below. At what generation of technology will the current density in the Al wire exceed the design limit from the current of a minimum dimension transistor?

	nMOSFET				Metallization	
Technology	L(μm)	W(μm)	G_{ox}(nm)	I_{on}/μm	w(μm)	h(μm)
0.5 μm (3.3 V)	0.50	0.70	12	0.3 mA	0.70	0.50
0.35 μm (3.3 V)	0.35	0.50	8	0.5 mA	0.50	0.50
0.25 μm (2.5 V)	0.25	0.35	6	0.5 mA	0.35	0.40
0.18 μm (1.8 V)	0.18	0.25	4	0.5 mA	0.25	0.35

7. The mean free path of a molecule in a gas is given by $\lambda = kT/(\sqrt{2}\pi d^2 p)$, where p is the gas pressure and d the diameter of the molecule. Estimate the mean free path for (*a*) a sputtering system with an Ar pressure of 4 mtorr ($d = 0.36$ nm); (*b*) a CVD W deposition system with hydrogen pressure of 30 mtorr ($d_{WF_6} = 0.56$ nm, $d_H = 0.28$ nm); (*c*) an O_3-TEOS system with a N_2 pressure of 600 torr ($d_{N_2} = 0.37$ nm, $d_{TEOS} = 2$ nm); and (*d*) an evaporation system with an H_2O pressure of 10^{-6} torr, ($d = 0.47$ nm).

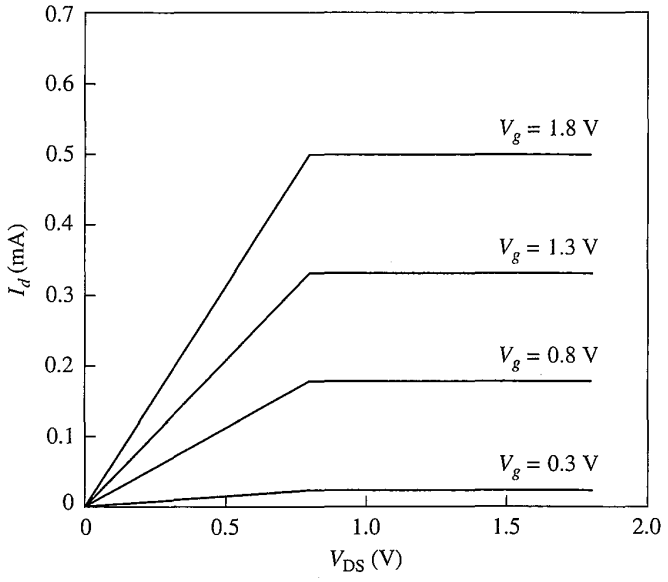
8. A collimator with an aspect ratio of 1 : 1 allows only about 20% of the sputtered material to pass. Let us assume that the conventional sputtering is 100% transparent and that the bottom coverage in the contact holes is 10% for conventional sputtering and 33.33% for collimated sputtering. (*a*) Calculate the thickness of collimated TiN barrier metal needed if 100 nm is used by conventional sputtering. (*b*) Estimate the cost of collimated TiN relative to conventional TiN sputtering, assuming that the only relevant factor is the equipment throughput. (*c*) How can the economy of collimated sputtering be improved?

9. Consider three CVD metal deposition processes: (*a*) CVD Al using TIBA, or tri-isobutyl-Al [$(C_4H_9)_3Al$]; (*b*) CVD W using WF_6; and (*c*) CVD Cu using $Cu(hfa)_2$. If we assume a utilization rate of 5% in all cases, how many grams of each chemical precursor are needed to deposit a 0.5-μm coating of metal on 200-mm wafers? For a 5000 wafer-start/week factory and three applications of the metal per wafer, how much precursor material (in weight) is needed per year (50 weeks)? How much waste material needs to be disposed of or burned? (The densities of Al, W, and Cu are 2.7, 19.3, and 8.9 g/cm^3, and the atomic weights for W, Cu, Al, and F are 184, 63.5, 27, and 19, respectively. Assume deposition occurs only on the wafer.)

10. Derive Eq. (8.41).

11. An nMOSFET with $L = 0.18$ μm and $W = 1$ μm can produce a $I_{on} = 0.5$ mA at $V_d = 1.8$ V, with characteristics as shown in the *I-V* curve. (*a*) If we assume there is no current crowding, what is the output current if a single 0.2-μm (round) contact is

Micron Ex. 1029, p. 498
Micron v. YMTC
IPR2024-00911



made on both the source and the drain of the transistor? Use $\rho_{sc} = 1 \times 10^{-7}$ Ω-cm^2, 2×10^{-7} Ω-cm^2, and 5×10^{-7} Ω-cm^2 and compare the results. (*b*) Repeat (*a*) for $W = 10$ μm. (*Hint:* Solve graphically.)

12. Consider a wide nMOSFET with $L = 0.25$ μm and $W = 40$ μm. Calculate the IR drop along the width of the transistor assuming $I = 10^{-12}$A/μm. Also calculate the RC delay along the width of the transistor assuming a $C = 1$ fF/μm (*a*) for a polysilicon gate with $R_s = 100$ Ω/□; (*b*) for a polycide gate with $R_s = 15$ Ω/□; and (*c*) for a salicide gate with $R_s = 3$ Ω/□.

13. Consider a MOSFET with $L = 0.25$ μm and $W = 10$ μm with a single pair of contacts at one end of the source and drain, as shown in Fig. 12*b*. The current crowding effect can be roughly simulated by a set of parallel resistors with values $R_{ch} + 2kR_{sh}$, where R_{ch} is the channel resistance per unit width, R_{sh} the diffusion sheet resistance, and k the number of squares for the kth parallel current path. (*a*) If $R_{ch} = 3.5$ kΩ/μm and $R_{sh} = 100$ Ω/□, what is the ratio I_d/I_{do}, where the I_{do} is the drain current without current crowding (i.e., current flows through the entire width of the MOSFET)? (*b*) What is I_d/I_{do} if $R_{sh} = 200$ Ω/□? (*Hint:* Divide the wide transistor into 10 parallel 1-μm wide transistors.)

14. To refine the model in Problem 13, consider a similar MOSFET with $L = 0.25$ μm and $W = 20$ μm. Let the current distribution along the width of the transistor be $i_d(x) = i_{do}x_o/(x_o + \alpha x)$, where x is the lateral distance in μm from the center of a contact window, $X_o = 1$ μm, and $\alpha = 2R_{sh}/R_{ch} = 200/3500 = 0.057$ for $R_{sh} = 100$ Ω/□ and $R_{ch} = 3.5$ kΩ/μm. Calculate the drain current ratio I_d/I_{do} for (*a*) a single contact to the source/drain at the center of the MOSFET; (*b*) a single contact to the source/drain at one end of the transistor; (*c*) 19 contact windows, separated by 1 μm, to the source/drain; and (*d*) repeat (*a*), (*b*), (*c*) for silicided source/drain with $R_{sh} = 3$ Ω/□.

15. Let us assume that the cost for a lithography step is $10; for etching, metal deposition, and dielectric deposition is $5; the cost for sacrificial material (resist, SOG) is $2; and for CMP is $8, per wafer, per application. Assume that the lithography depth of focus is 1.0

Micron Ex. 1029, p. 499
Micron v. YMTC
IPR2024-00911

μm and poly and metal thicknesses are $M_0 = M_1 = M_2 = M_3 = M_4 = 0.6$ μm. What is the most economical four-level-metal metallization process? What is the most economical process if the cost per application for CMP is $20? You may use any combination of global and local planarization processes.

16. Show that during a CMP process with $\omega_2 = 2\omega_1$ the locus of a point at the edge of the wafer on the platen is an ellipse with the major axis $= R + r$ and the minor axis $= R - r$. (*Hint:* Solve graphically.)

17. Show that during CMP, the locus of a point on the edge of the wafer for $\omega_2 = \frac{1}{2}\omega_1$ traces a shorter path than that for $\omega_2 = -\frac{1}{2}\omega_1$. (*Hint:* Solve graphically.)

18. The current crowding at a metal-via contact can be calculated by a finite element method using commercial tools. Consider a very simple 2-D case with a resistance network as shown below. Assign $R_{21} = 2R_{11}$ and $R_{31} = 3R_{11}$ to account for the difference in the length of current paths for I_1, I_2, and I_3. (*a*) Let $R_{12} = R_{22} = R_{32} = 4R_{11}$ for a CVD W plug. Find the ratios I_1/I_2 and I_1/I_3. (*b*) Repeat for a TiN overcoat between the W plug and Al; let $R_{12} = R_{22} = R_{32} = 10R_{11}$. (*c*) Repeat for an Al plug where $R_{12} = R_{22} = R_{32} = R_{11}$.

19. To use Cu for wiring one must overcome several obstacles: (1) the diffusion of Cu through SiO_2, (2) adhesion of Cu to SiO_2, and (3) corrosion of Cu. One way to overcome these obstacles is to use a cladding/adhesion layer (e.g., Ta or TiN) to protect the Cu wires. Consider a cladded Cu wire with a square cross section of 0.5 μm × 0.5 μm and compare it with a layered TiN/Al/TiN wire of the same size, with the top and bottom TiN layers 40 nm and 60 nm thick, respectively. (*a*) What is the maximum thickness of the cladding layer if the resistance of the cladded Cu wire and the TiN/Al/TiN wire is the same? (*b*) What is the maximum cladding layer thickness if 20% lower resistance is desired? (*c*) What is the maximum cladding layer thickness if 40% lower resistance is desired? (*d*) Keeping the cross section as a square and decreasing the wire size to 0.3 μm × 0.3 μm, repeat (*b*). What is the sheet resistance of the structures (for both Al and Cu) for (*b*) and for (*d*)? (The resistivity of Al is 2.7 μΩ-cm, and the resistivity of Cu is 1.7 μΩ-cm.)

20. Consider a long wire with resistance R and capacitance (to ground) C driven by a transistor with maximum output current I_o. The voltage across the capacitor at time t can be expressed by $V(t) = V_o + (V_{dd} - V_o)\{1 - \exp[-(t - t_o)/RC]\}$ where $V_o = V_{dd} - i_o R$ is the voltage across the capacitor at time t_o. The total time required to charge this line to a voltage $V_f \approx 0.9V_{dd}$ can be derived as $t = t_o + RC[\log_e(V_{dd} - V_o)/(V_{dd} - V_f)]$, and $t_o = CV_o/I_o$. (*a*) For a 2-mm long Al wire with a cross section of 0.5 μm × 0.5 μm at 0.5 μm from the ground plane, as in Problem 2(*b*), connected to a transistor with a maximum output of 2.5 mA, what is the time required to charge this line to a $V_f = 3.0$ V ($V_{dd} = 3.3$ V)? (*b*) What is the fraction of this time caused by the wire *RC* delay? (*c*) What is the time required to charge this line to 3.0 V if it is Cu? (*d*) What is the minimum time required to charge this wire to 3.0 V if the driving transistor can deliver unlimited current? (*e*) In case (*d*), what fraction of the delay is caused by the wire *RC* delay? (*f*) If Cu wire is used in (*d*) instead of Al, what is the minimum delay time? What fraction of it is due to the wire *RC* delay?

Micron Ex. 1029, p. 500
Micron v. YMTC
IPR2024-00911

CHAPTER 9

Process Integration

C. Y. Lu and W. Y. Lee

9.1 INTRODUCTION

Integrated circuit (IC) technology has produced dramatic advances over the past 20 years. Since the early 1970s technical progress has propelled the integrated circuit industry from small-scale integration (SSI), at less than 30 devices per chip, to medium-scale integration (MSI), 30 to 10^3 devices per chip, to LSI, 10^3 to 10^5 devices per chip, to VLSI, 10^5 to 10^7 devices per chip, and now to ULSI, 10^7 to 10^9 devices per chip. The present period started in the early 1990s, roughly corresponding to the introduction of four-Mbit dynamic random-access memory (DRAM) devices. The minimum line width span for ULSI, from 0.80 to 0.15 micrometers, puts us squarely in the submicrometer regime. In this chapter, we discuss various integration issues for different technologies and circuits. These technologies have much in common; they differ only in special blocks of processes particular to their applications. The mainstream of ULSI technology is in complementary metal-oxide-semiconductor (CMOS)-based silicon processes.[1] These processes are shown schematically in Fig. 1. Bipolar technology, the major technology in the past, is becoming a niche application technology in ULSI integration. In Sections 9.3 through 9.5 we treat the CMOS, bipolar, and bipolar CMOS (BiCMOS) technologies in turn. For each technology, we describe a typical process flow in some detail, along with any special process modules needed. These descriptions are not intended to standardize the processes because there are many options; rather, we present baseline process sequences that are useful primarily as vehicles for illustrations.

Throughout the years, a considerable number of process modules have been developed for process integration. In Section 9.2 we describe some of the important process modules, or building blocks, that can be used, in general, for all circuits, and we discuss the special processing modules relevant to their particular technologies in

Micron Ex. 1029, p. 501
Micron v. YMTC
IPR2024-00911

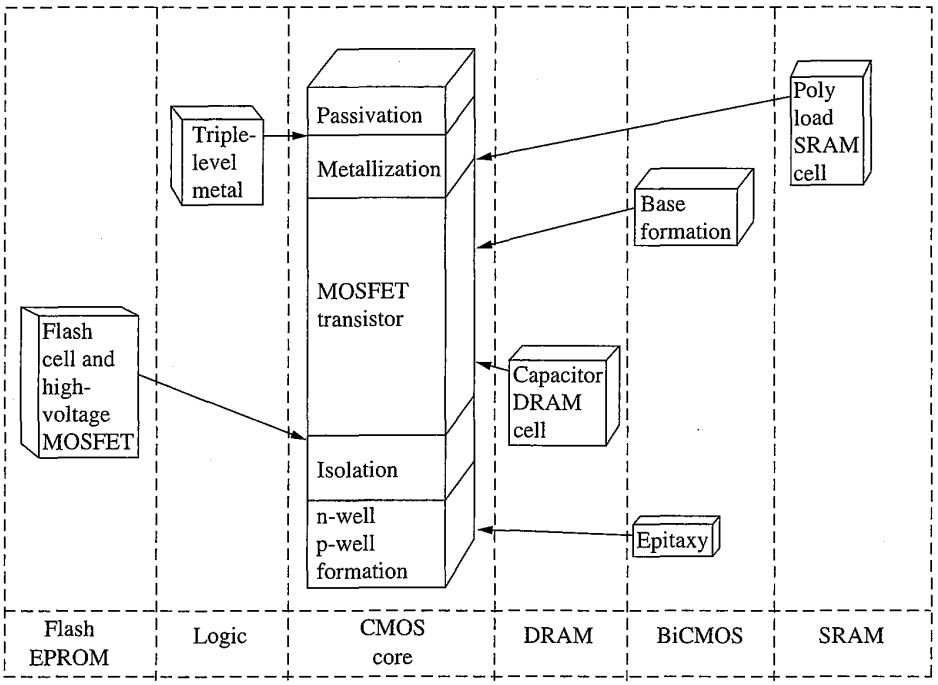


FIGURE 1
Process harmonization for CMOS core-based technologies. (*After Chatterjee, Ref. 1.* ©1991 IEEE.)

Sections 9.3 through 9.6. At the end of these sections we give special consideration to some of the fabrication and device issues of that particular technology. Standard memory technologies of DRAM and static random-access memory (SRAM) are covered in detail in Section 9.6. Section 9.7 describes some major considerations for process integration in ULSI fabrication. Finally, Section 9.8 summarizes the chapter and explores future trends for this industry.

The purpose of this chapter is to present fundamentals of integrating silicon processing steps to create silicon devices in the ULSI context. Since device performance is an important criterion for process optimization, we discuss how changes in processing steps affect the important device parameters. A detailed treatment of device physics is beyond the scope of this book, and we assume that readers are familiar with these fundamentals. However, readers may refer to several excellent textbooks on this subject, such as those by Sze[2] and Shur,[3] in the reference list at the end of the chapter.

9.2 BASIC PROCESS MODULES AND DEVICE CONSIDERATIONS FOR ULSI

9.2.1 Introduction

As the complexity and performance of integrated circuits increase, more processing steps are needed to fabricate them. Four or five mask levels were quite adequate for

Micron Ex. 1029, p. 502
Micron v. YMTC
IPR2024-00911

primitive ICs in the 1970s, whereas 16-Mb DRAM (ULSI) memory chips require more than twenty mask levels. Often it was found that, as dimensions shrank, it became more difficult to make smaller-geometry devices with pattern transfer techniques alone; single-unit process steps that served well before were no longer sufficient because the devices thus obtained would have degraded performance. It takes a group of operations, sometimes involving a dozen individual steps, to replace the original unit process step in order to maintain or improve the device performance and manufacturability. An example is the salicidation process described in subsection 9.2.6. This process replaces the simple source and drain diffusion to reduce device series resistances and increase circuit density. Another example is the intermetal layer dielectric spin-on-glass (SOG) planarization steps, which smooth out the topography caused by smaller line and space pitches. As Fig. 1 shows, a modular approach to managing the design of multiple technologies, with a maximum of common modules, or building blocks, and a common set of manufacturing equipment, is essential for ULSI technology development.[1] Because of their importance, we describe some common process modules below. For special-purpose ICs, additional modules are covered in subsequent sections.

9.2.2 Well Structures

The CMOS device is the most popular choice in ULSI applications. The first consideration in the CMOS process is to determine how the substrates will be formed for the two types of transistors. The n-channel transistors need a p-substrate, and the p-channel transistors require an n-substrate. The three approaches to forming the two different substrates are referred to as p-well, n-well, and twin-well processes (see Fig. 2).

The p-well (also called p-tub) process involves implanting or diffusing p-type dopant into an n-substrate at a concentration high enough to overcompensate the n-substrate and to give good control over the desired p-type doping (Fig. 2*a*). The concentration of n-substrate dopant must be high enough to ensure that the p-channel device characteristics are adequate. The concentration of p-well dopant must typically be five to ten times higher than that in the n-substrate to ensure this control. However, excessive p-well doping produces deleterious effects in the n-channel transistor, such as increased back-gate bias sensitivity, reduction in mobility, and increased source/drain to p-well capacitance.

The n-well (also called n-tub) process is an alternative approach in which an n-well is formed in a p-type substrate. Because the n-channel device is formed in the p-type substrate, this approach is compatible with standard n-channel MOS (NMOS) processing (Fig. 2*b*). In this case the n-well overcompensates the p-substrate, and the p-channel transistor suffers from excessive doping effects.

The twin-well (twin-tub) process allows two separate wells to be implanted into very lightly doped silicon. This allows the doping profiles in each well region to be tailored independently so that neither type of device will suffer from excessive doping effects. The lightly doped silicon is usually an epitaxial layer grown on a heavily doped silicon substrate (Fig. 2*c*). The substrate can be either n-type or p-type.

Micron Ex. 1029, p. 503
Micron v. YMTC
IPR2024-00911

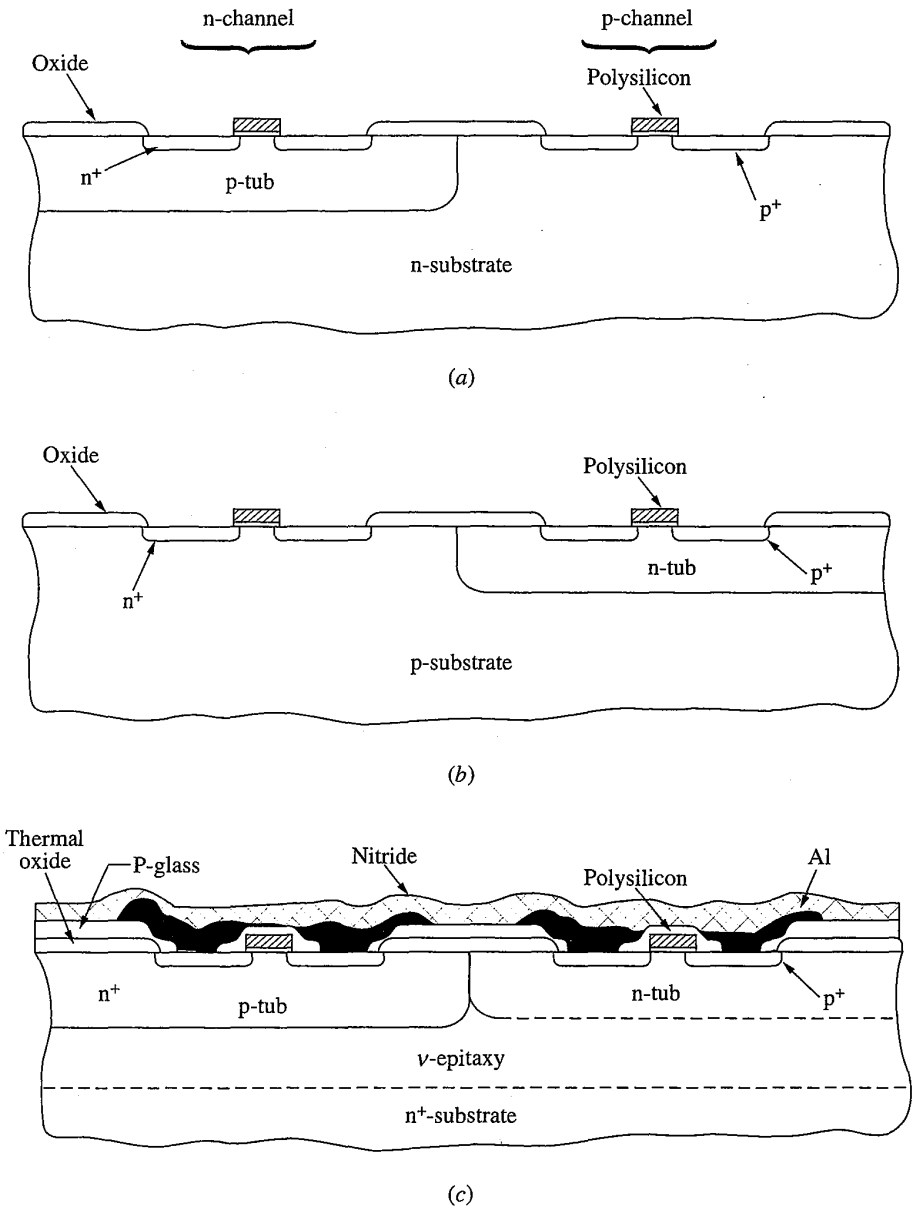


FIGURE 2
Various CMOS structures: (*a*) p-well, (*b*) n-well, (*c*) twin-well. (*After Parrillo et al., Ref. 4.* ©1980 IEEE.)

Twin well

Figure 3 shows the twin-well process module. Although the process is a little complicated, both its n- and p-channel devices can be optimized without compromising their performance.[4,5] First, the wafer is oxidized and capped with a nitride layer, which is selectively removed from the n-well regions by a patterning step. Phosphorus for n-well dopant is then implanted. After the photoresist mask is stripped, a thick (350 nm) oxide is grown over the n-well regions. The nitride

Micron Ex. 1029, p. 504
Micron v. YMTC
IPR2024-00911

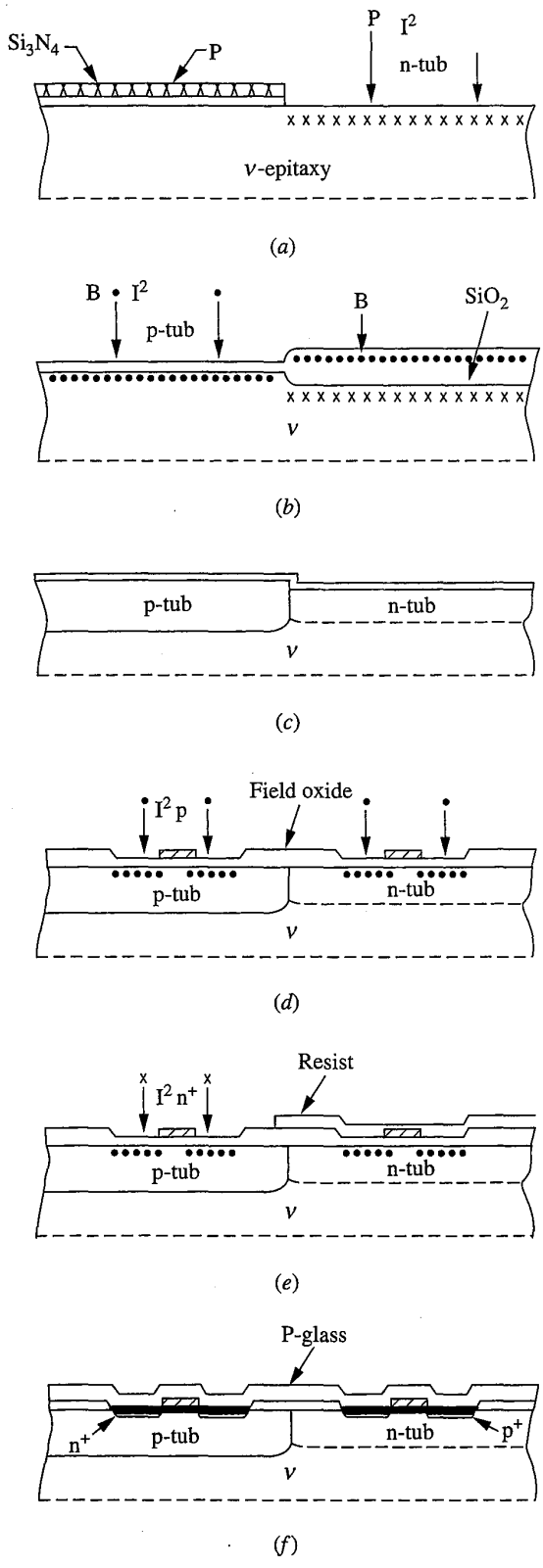


FIGURE 3
Twin-well CMOS structure at several stages of the process: (*a*) n-well ion implant (I^2); (*b*) p-well implant; (*c*) twin-well drive-in; (*d*) nonselective p^+ source/drain implant; (*e*) selective n^+ source/drain implant using photoresist mask; (*f*) P-glass deposition. (*After Parrillo et al., Ref. 4.* ©1980 IEEE.)

Micron Ex. 1029, p. 505
Micron v. YMTC
IPR2024-00911

layer is then etched off, and p-well regions are exposed, while the n-well regions are covered with the implant-blocking oxide. Thus, during the p-well implant, boron is implanted self-aligned to the n-well edge; the n-well oxide protects the n-well from counter-doping. Next, both wells are driven in simultaneously (at 1100°C for 500 min). At the end of the drive-in step, the concentration on the surface may be 1×10^{16} cm^{-3} for the p-well, and 3×10^{16} cm^{-3} for the n-well. The higher concentration of the n-well is designed to improve the punch-through performance of the p-channel metal-oxide-semiconductor field-effect transistor (pMOSFET) and to eliminate the need for a separate masking step to increase the surface impurity concentration and, therefore, the threshold voltage for the n-well field isolation (channel stop).

Retrograde well

Conventional wells are formed by implanting dopants and diffusing them to the desired depth. However, they diffuse laterally as well as vertically, which reduces packing density. If a high-energy implant is used to place the dopants at the desired depth without further diffusion, much less lateral spread will occur. Such high-energy implants also bury the peak of the impurity profile to a certain depth within the substrate, and the concentration of the impurity decreases as it approaches the wafer surface. Wells with such a profile are called retrograde wells.[6]

Besides the potential benefit of packing density, such wells also have the following advantages:

- A retarded field is created in the parasitic bipolar transistor in the CMOS structure, thereby reducing the possibilities of latch-up.[7]
- Susceptibility to vertical punch-through is reduced.
- The conductivity in the bottom of the well is increased, which also provides some further latchup protection.[7]
- Sometimes boron at high energies (such as 400 keV) is implanted after the field oxide is grown; this may provide higher field threshold voltage and less channel encroachment by the field dopants.

9.2.3 Isolation

For MOS circuits, threshold voltage for the field-oxide areas must be higher to isolate individual devices. Many isolation methods have been proposed. We consider three of the most well-known structures suitable for ULSI applications: LOCOS, poly-buffered LOCOS, and shallow trench isolation.

LOCOS

The localized oxidation isolation method (LOCOS) is certainly the most dominant isolation process used in IC technologies.[8] It utilizes the property that oxygen diffuses through Si_3N_4 very slowly. When silicon is covered with silicon nitride, Si_3N_4, no oxide can grow. In addition, nitride itself oxidizes at a very slow rate; thus it will remain as an integral oxidation barrier layer throughout the entire oxidation step. Figure 4 shows the major process steps. After a wafer with a bare silicon surface is cleaned, a 20 to 60 nm SiO_2 layer is thermally grown on its surface. Next, a 100

Micron Ex. 1029, p. 506
Micron v. YMTC
IPR2024-00911

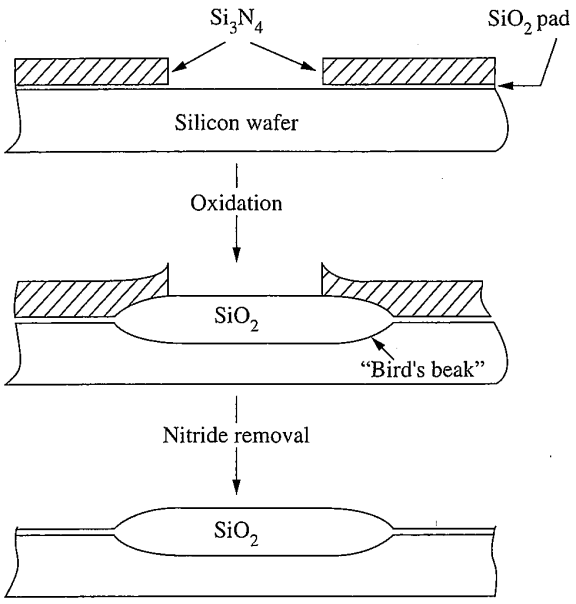


FIGURE 4
Cross-section depiction of process sequence for local oxidation of silicon LOCOS process.

to 200-nm-thick layer of chemical vapor deposition (CVD) silicon nitride, which functions as an oxidation mask, is deposited. The active regions are then defined with a photolithographic step so that they are protected by the photoresist patterns. The composite oxide/nitride layers are then plasma-etched as a stack. With the photoresist patterns in place, the wafer is subsequently implanted with a 10^{12} to 10^{13} cm^{-2} dose of boron and with energies in the range of 60 to 100 keV. This channel stop is now self-aligned to the n-channel devices. The dosage is chosen to provide adequate field-inversion voltage but not so high as to cause high n^+ junction-to-substrate capacitance and excessively low n^+ breakdown voltage. The nitride films have a drawback; they exhibit very high tensile stress when deposited on silicon. The underlying layer of oxide, called a pad or buffer oxide, is used to cushion the transition of stresses between the silicon substrate and the nitride film. In general, the thicker the pad oxide is, the less likely it is that the edge force transmitted to the substrate could cause crystal dislocations. On the other hand, a thicker pad oxide will render the nitride layer less effective as an oxidation mask by allowing lateral oxidation to take place. The lateral regions at the nitride edges, which span the transition from the thick field oxide to the thinner pad oxide, are called a *bird's beak* because they are shaped like the beak of a bird.

After the channel-stop implant, the field oxide with a thickness of 500 to 900 nm is thermally grown by wet oxidation at temperatures of 900 to 1000°C for 4 to 8 hours. This and subsequent high-temperature steps have the following device consequences: Because of preferential impurity segregation, boron redistributes at the interface and becomes more depleted. The lateral diffusion of boron also causes it to encroach into the NMOS active areas, and the resultant transistor will behave as if it is a narrower device. Another phenomenon, named the Kooi effect after its discoverer, occurs during the growth of the field oxide.[8] It was found that a thin layer of silicon oxy-nitride can form on the silicon surface near the bird's beak as a result

Micron Ex. 1029, p. 507
Micron v. YMTC
IPR2024-00911

of the reaction of NH_3 with the silicon substrate.[9, 10] The NH_3 is generated from the reaction between water vapor and the masking nitride during the field oxidation step and then diffuses through the pad oxide and reacts with the substrate to form silicon oxy-nitride spots or ribbons (thus it is sometimes called the white-ribbon or black-belt effect). When the gate oxide is grown, its growth rate may be more impeded at these locations than elsewhere, and is thus thinner at these spots, causing low-voltage breakdown and reliability problems with the gate oxide.

In the next stage of the process sequence, the masking nitride layer is stripped with phosphoric acid at 180°C using a reflux boiler. Then the pad oxide is etched. An extra "sacrificial" gate oxide layer is then grown to eliminate the black belts and white ribbons of the Kooi effect by lifting the micromasking oxy-nitride skin off the interface and washing it away when this sacrificial oxide is etched off.[10]

Poly-buffered LOCOS

Although conventional LOCOS has enjoyed long staying power over a number of years, it runs into serious limitations in the submicrometer regime because of its relatively long bird's beak (approximately 0.4 to 0.6 μm on each side). Many schemes have been modified to correct this shortcoming. The simplest way is to etch back a portion of the field oxide after it is grown. The most promising approach is the poly-buffered LOCOS, or PBLOCOS.[11,12] We know that the bird's beak can be substantially reduced if we increase the nitride and reduce the pad thicknesses, but severe dislocations may be generated in the substrate because of the stress exerted by the nitride film during oxidation.[9] However, it was found that if we substitute the oxide pad with a poly-buffered pad layer (50 nm poly/10 nm oxide) and use a much thicker nitride (approximately 400 nm), good results can be achieved. The polysilicon acts as a cushion against damaging stresses, and a bird's beak of 0.1 to 0.15 μm is possible. With some modification, PBLOCOS can even achieve a bird's beak of almost 0 μm without leakage and gate oxide degradation problems.[12]

Shallow trench

No matter how we modify the LOCOS process, it will probably be difficult to reduce the bird's beak length to much less than 0.1 μm per side with totally flat topology. Therefore, for sub-quarter-micrometer technology, new approaches to isolation with totally flat topology are needed. In shallow-trench isolation,[13] trenches about 0.3 to 0.8 μm deep are anisotropically etched into the silicon substrate through dry etching. Active regions are those that are protected from the etch when the trenches are created. Next, a CVD oxide is deposited on the wafer surface and then etched back so that it remains only in the recesses, with its top surface at the same level as the original silicon surface. Etchback is performed using a sacrificial photoresist method. This technique has the advantages of having no bird's beak and no encroachment. Also, when two devices are separated by a trench, the electrical field lines have to travel a longer distance and change direction twice, so they are considerably weakened. Therefore, trenches of submicrometer dimensions are adequate for isolation to prevent punch-through and latch-up phenomena.[14] However, because it is a complicated process, and because up to now LOCOS and PBLOCOS have been satisfactory methods, shallow-trench isolation has not been used extensively in the industry.

Micron Ex. 1029, p. 508
Micron v. YMTC
IPR2024-00911

9.2.4 Drain Engineering

Lightly doped drain (LDD)

In scaling down devices, thinner gate oxide and higher-doped channels are needed to boost the punch-through voltage of the short channeled devices. Both measures drastically increase the electric field (both the horizontal and vertical components) near the drain regions where charge carriers, accelerated by this field and becoming hot, can overcome the oxide barrier and inject into the gate. Hot carriers induce damage to the gate oxide and may be trapped there, thereby degrading the device performance. By reducing this drain field, the lightly doped drain (LDD) alleviates this problem.[15]

In the LDD structure, the drain is formed by two implants (Fig. 5). One of these is self-aligned to the gate electrode and the other is self-aligned to an oxide spacer at the edge of the gate. The spacer is formed by depositing an oxide layer about 200 nm thick and then etching it with a blanket anisotropic reactive ion etch. All the oxide in the field will be cleared except at regions next to sharp steps where the deposited oxide is the thickest. This oxide spacer has a lateral dimension roughly equal to 75 to 85% of the oxide's original thickness and can be controlled to a high degree of precision.

The first, lighter dose forms a lightly doped section of the drain and source at the edges near the channel. In the NMOS devices, this dose is normally 1 to 5×10^{13} cm^{-2} of phosphorus. The second, heavier dose forms low-resistivity regions of the source and drain that merge into the lighter doped regions. In NMOS devices, this implant is typically arsenic with a dose of 1 to 5×10^{15} cm^{-2}. These two implants are separated and displaced in a self-aligned manner by the spacers. As a result, the electric field can be reduced to an acceptable level near the gate edge because the impurity concentrations there are much lighter than those in a conventional drain. Since the heavier, second dose is further away from the gate edge than would be the case in a conventional drain structure, its depth can be made somewhat greater without adversely affecting the device operation. The increased junction depth also has the advantages of lowering both the sheet and contact resistances and providing better protection against junction spiking by the aluminum contact.

Large-angle-tilt implanted drain (LATID)

In the sub-half-micrometer regime, the usefulness of the LDD structure itself may be reaching a limit, because the spacer interface degradation from hot carrier injection is not subjected to gate control. Better results can be obtained by implementing the LATID.[16,17] It is made by implanting phosphorus for a total flux of 1 to 8×10^{13} cm^{-2} with a tilt angle θ (20 to 60 degrees) from the source and drain sides and by rotating the wafer stepwise using a reduced spacer length L_s (0.05 to 0.15 μm) (Fig. 6). This drain is thus fully compatible with the conventional process and the usual LDD masking steps. In the example given, $L_s = 0.08$ μm and the n^- dose $= 10^{13}$ cm^{-2} for an optimized 0.25-μm device structure. Devices of better reliability can be obtained using this scheme of off-axis implant, because there is better gate control action over the LDD regions.

Micron Ex. 1029, p. 509
Micron v. YMTC
IPR2024-00911

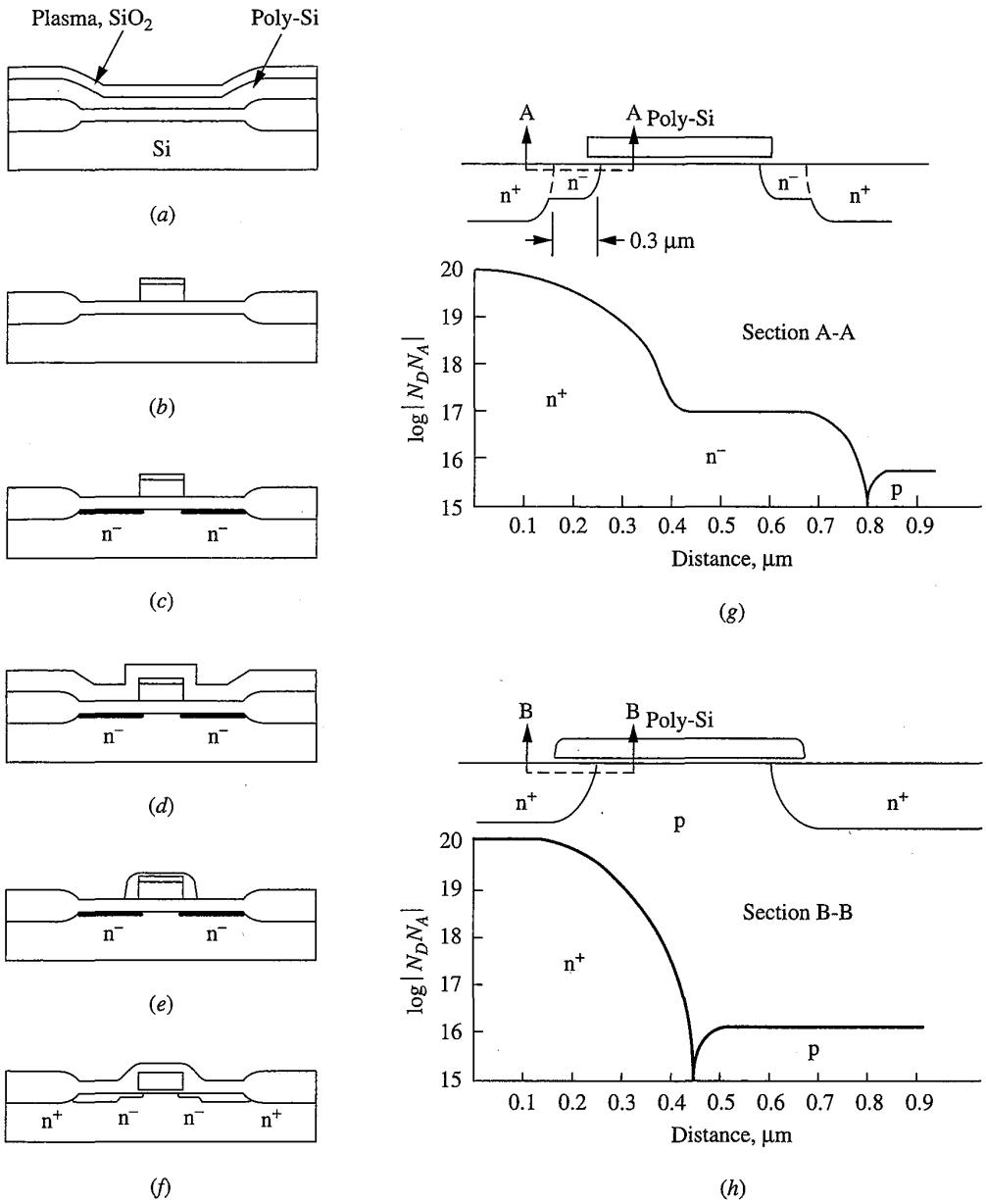


FIGURE 5
(*a–f*) Process sequence used to form lightly doped drain (LDD) structures. (*g*) Doping profile in an LDD structure taken through section A-A. (*h*) Doping profile in a conventional drain structure taken through section B-B. (*After Ogura et al., Ref. 15.* ©1980 IEEE.)

Double-implant lightly doped drain (DI-LDD), or halo

To improve the punch-through voltage of a MOSFET, a "halo" implant can be applied to the drain of the device.[18] For an n-channel device, boron may be implanted with a dosage of about 1×10^{13} cm^{-2} at the drain side. Because of the faster diffusion of boron, its junction is pushed past that of the n^+ even if its concentration is

Micron Ex. 1029, p. 510
Micron v. YMTC
IPR2024-00911

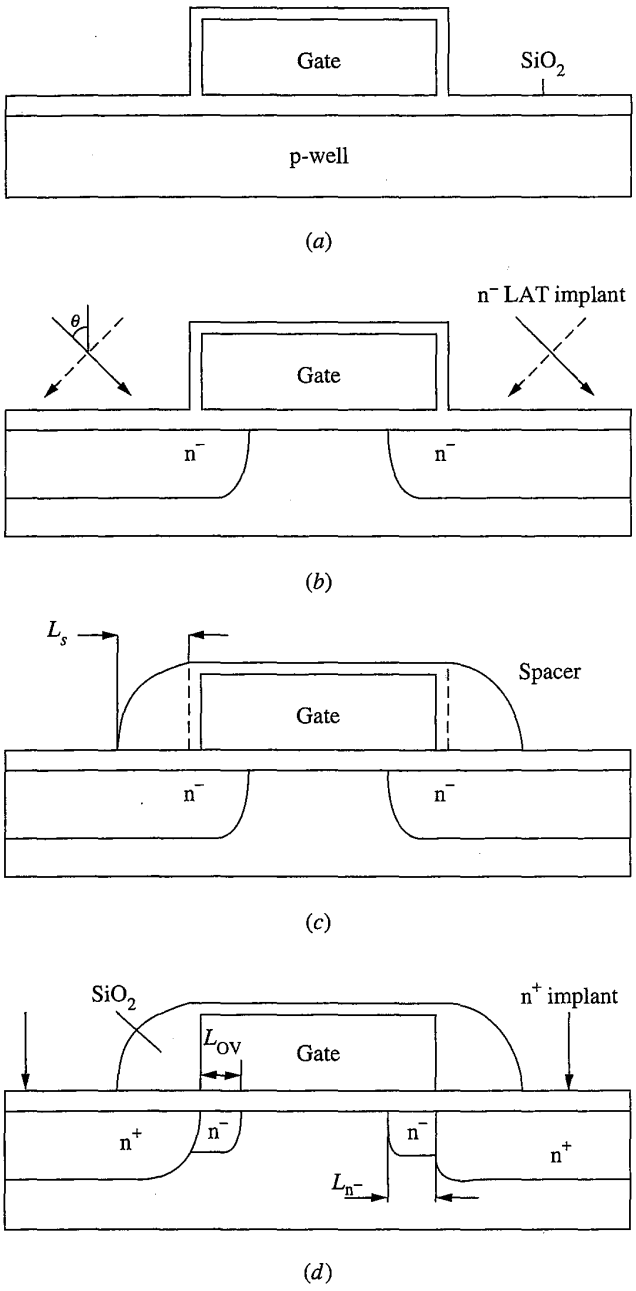


FIGURE 6
Process sequence for LATID FETs having sidewall spacers. (*a*) Gate etching and sidewall oxidation. (*b*) Phosphorus n^- LAT implant. (*c*) Formation of sidewall spacers. (*d*) Arsenic n^+ implant. (*After Hori, Odake, and Yasui, Ref. 16.* ©1992 IEEE.)

much less. The p-type impurities form a ring-shaped halo around the n^+ junction. It has the advantage that its enhanced concentration would increase the punch-through voltage, and this feature takes much less space because it is self-aligned. Also, since the source side is not affected, the body factor[2,3] of the device will not be increased.[18]

9.2.5 Shallow Junctions

As devices shrink, the vertical source and drain junction depths of MOSFET devices must also shrink to reduce short channel effects and improve isolation properties.[19]

Micron Ex. 1029, p. 511
Micron v. YMTC
IPR2024-00911

It is much harder to make a shallow p^+/n junction than a shallow n^+/p junction. Because of its lighter mass, the boron-implant impurity profile does not follow a strict Gaussian distribution; rather, it has a long tail formed by excessive channeling of boron atoms in the silicon lattice, which makes it hard to produce p^+/n junctions shallower than 0.25 μm by conventional method. To reduce p^+/n junctions to 0.25 μm, BF_2 is more suited than boron because of its higher molecular weight and lower tendency of channeling in silicon. But even at very low implant energies such as 15 keV, the channeling tail is still considerable. This method can only be regarded as a stop-gap measure and not as a satisfactory way to obtain shallower junctions.

Extensive work has been done to attain very shallow junctions with high activation and low defect. The process steps of a typical example follow:

1. Do preamorphization of the silicon substrate with a Si^+ or Ge^+ implant; Ge is preferable to Si because it is more efficient in amorphization. Preamorphization randomizes the crystal surface and reduces channeling.[20,21]
2. Implant boron or BF_2 with very low energy (< 10 keV); boron is preferable to BF_2 because it is more difficult to anneal out damages in the presence of fluorine.[22]
3. Anneal with rapid thermal annealing (RTA).

It is important to set the right energies and depths of the implants to minimize high diode leakage, which is governed by the following considerations:

The depth of the amorphous/crystal (a/c) interface (X_a) must be shallow since the end-of-range defects of the a/c interface cannot be annealed out completely.

The initial boron implant range R_p must be considerably smaller than the a/c depth so that no channeling will occur and so that the boron tail will not receive enhanced diffusion as it would if it were in the interstitial-rich region under the a/c interface.[20]

The junction can then be pushed out during annealing to about 70 nm beyond the a/c interface, so that all the crystal defect regions are completely enclosed by the p^+/n region. This technique minimizes excessive diode leakage.[20]

Since the enhanced diffusion is only a transient phenomenon aided by the presence of interstitials, by the time the junction reaches the a/c interface, most of the interstitials will be gone and no enhanced diffusion will occur.

The following lists a few examples from the literature[20] of conditions under which shallow p^+/n junctions were achieved. X_j is the final p^+/n junction depth.

Case 1:

Ge preamorphization for $X_a = 40$ nm
BF_2 implant at 10 keV, $2 \times 10^{15} cm^{-2}$
RTA anneal at 950°C for 10 sec, $\rightarrow X_j = 80$ nm

Case 2:

Ge preamorphization, implant at 27 keV, 3×10^{14} cm^{-2}
Boron implant at 1.35 keV, 5×10^{14} cm^{-2}
RTA anneal at 1050°C for 10 sec, $\rightarrow X_j = 110$ nm

Micron Ex. 1029, p. 512
Micron v. YMTC
IPR2024-00911

Case 3:

Si preamorphization, implant at 15 keV, 2×10^{15} cm^{-2}
BF_2 implant at 15 keV, 2×10^{15} cm^{-2}
Low-temperature anneal at 600°C, 1 hour
RTA anneal at 1000°C for 10 sec $\rightarrow$ $X_j = 120$ nm

9.2.6 Salicidation

In the salicidation, or self-aligned silicidation, process, the entire source and drain regions and the top of the poly-silicon gate of a MOSFET are covered with a low-resistivity metallic silicide film. This approach is attractive because such a film is formed using a self-aligned process that does not require any additional masking steps.[23,24] The conventional salicidation scheme of post-junction-formation silicide (PJS) process sequence of $TiSi_2$ is described as follows (see Fig. 7):

1. After poly-Si sidewall oxide spacers are formed, the source/drain (S/D) regions are implanted to form junctions (Fig. 7*a*).

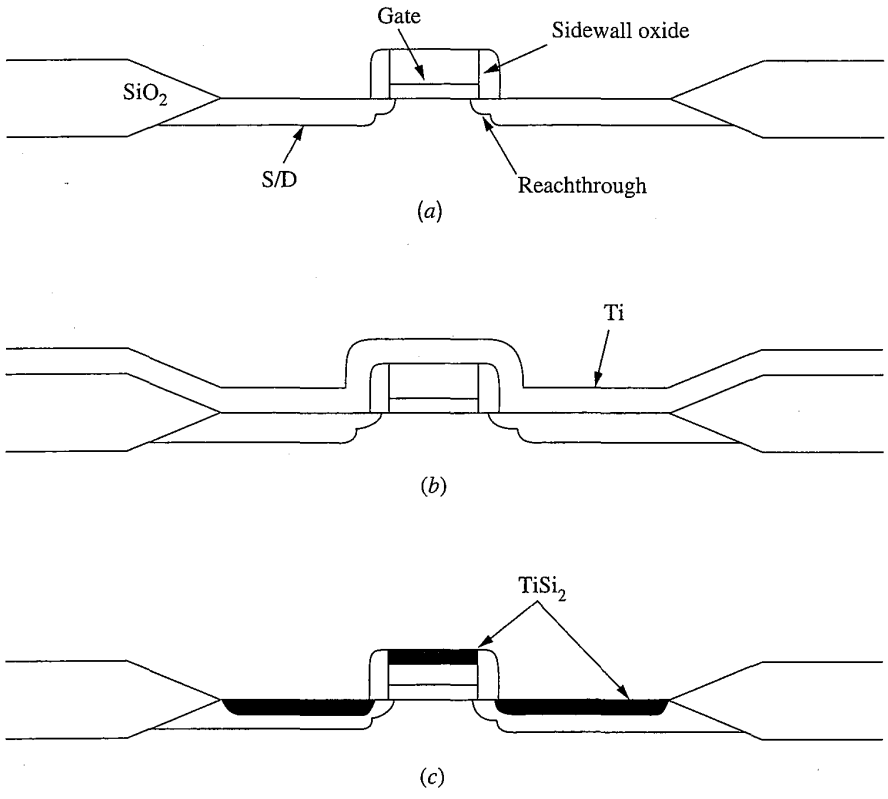


FIGURE 7
Fabrication of MOS transistors with low-resistivity gates and junctions using the self-aligned $TiSi_2$ process. (*After Tang et al., Ref. 28.* ©1987 IEEE.)

Micron Ex. 1029, p. 513
Micron v. YMTC
IPR2024-00911

2. A HF dip is applied to ensure a clean interface on the wafer and promote uniform reaction. Immediately afterward, 50 to 100 nm of metal titanium film is deposited (Fig. 7*b*).
3. The wafer is then heated to 600 to 800°C in nitrogen, which causes the silicidation reaction to occur wherever metal is in contact with silicon or poly; everywhere else, the metal remains unreacted. Special precaution must be taken to avoid oxygen contamination in the furnace tube, since Ti reacts readily with oxygen to form unwanted titanium oxides. Nitrogen diffuses into and reacts with Ti over the oxide regions to form a stable layer of TiN, which acts as a diffusion barrier. Silicide formation is faster than TiN formation at higher temperature; therefore, at higher anneal temperatures silicon diffuses laterally from the substrate and gate regions and converts Ti over the oxide spacer regions into silicides. This can cause a silicide film to form on top of the oxide sidewall, which can short the gate to the substrate. However, since a lower temperature is chosen, no advance effect occurs at this step.
4. The unreacted metal is then selectively removed by an etchant that does not attack silicide, the silicon substrate, or the oxide. As a result, the poly-gate and the source and drain exposed regions are now completely covered by a silicide film, but there is no silicide elsewhere.
5. A subsequent annealing of 800 to 900°C is then used to further reduce the sheet resistance to its final values, typically 10 to 1.5 $\Omega/\square$ (Fig. 7*c*).

In order to integrate the salicide process with a shallow junction in ULSI CMOS application, several novel schemes have been proposed, such as dopant drive-out (DDO),[25] an implant through metal film (ITM),[26] and an implant through prior-formed silicide film (ITS).[27] For a detailed discussion of these various approaches, see Chapter 8 of this book or Ref. 24 at the end of this chapter.

9.2.7 Local Interconnect (LI)

An interconnection scheme derived from the salicidation process described above uses TiN to replace buried contacts by making direct connections from the diffusion to the poly gates and, as a result, increases the packing density. This TiN local interconnect approach requires no extra deposition steps since it utilizes a layer that is normally discarded.[28] Figure 8 illustrates how the salicidation process has been modified. After the silicide reaction step, the conductive TiN is patterned and etched. Following resist strip, an 800°C anneal in argon reduces the TiN and $TiSi_2$ resistivities to their final values. This process is capable of realizing LI linewidth of less than 0.8 μm and specific contact resistance as low as 2×10^{-8} Ω-cm^2. LI also allows junctions to metal contact to be extended over isolation regions so that minimum-geometry junctions can be realized to reduce circuit areas and capacitances. An example of how LI technology increases circuit packing density is that the cell area of a six-transistor SRAM cell, using the same design rules, is reduced by 25%.[28]

Local interconnect is not limited to using TiN. For the same purpose, a polysilicon layer, with or without a silicide on top, can be used as landing pads on top of source or drain diffusion junctions. They can be extended to make metal connections on top of field oxide or directly to other diffusions or gates to form a sub-level

Micron Ex. 1029, p. 514
Micron v. YMTC
IPR2024-00911

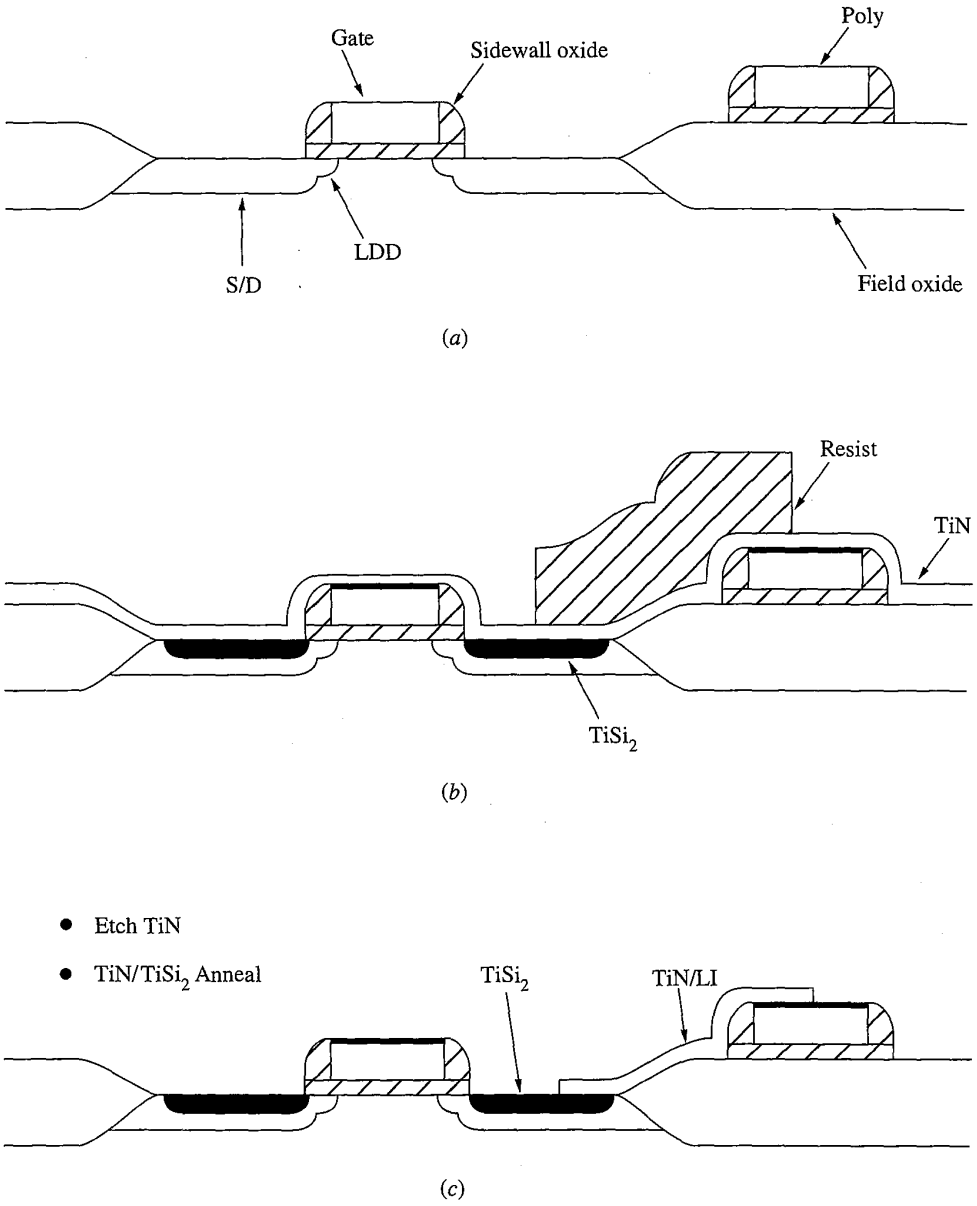


FIGURE 8
Modification of the self-aligned titanium silicide process to realize a level of local interconnect using titanium nitride. (*After Tang et al., Ref. 28.* ©1987 IEEE.)

local interconnect. A representative example is described by the folded extended window structure MOSFET (FEWMOS) with raised source and drain.[29] For its implementation, as depicted in Fig. 9, the polysilicon (or polycide) and its top deposited dielectric layers are patterned as a stack, and then oxide sidewall spacers are formed at this poly/dielectric stack by anisotropic etchback. Then a doped polysilicon layer is deposited and contacted with the exposed source and drain regions. This poly layer is then patterned to form a metal-contact landing pad and local interconnect. This doped polysilicon landing pad on top of the source and drain areas can serve as a

Micron Ex. 1029, p. 515
Micron v. YMTC
IPR2024-00911

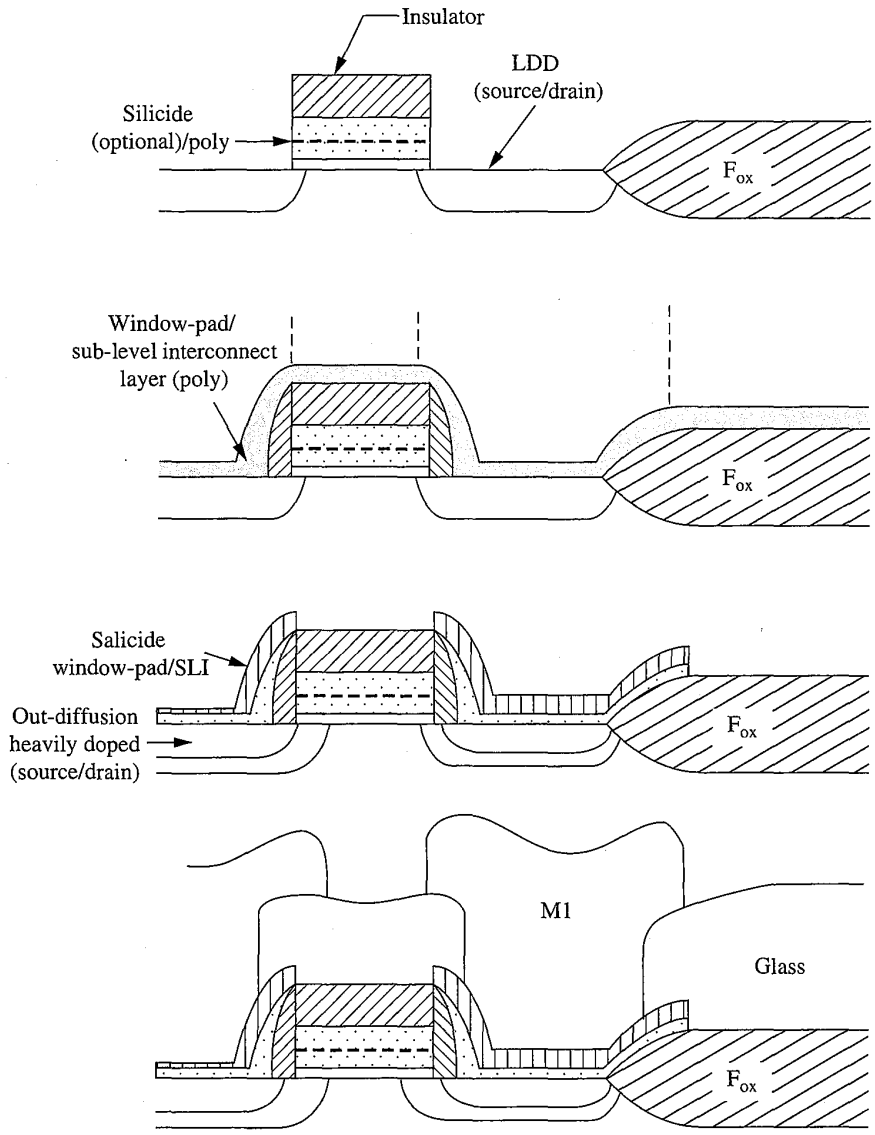


FIGURE 9
The process sequence of a FEWMOS structure. (*After Lu et al., Ref. 29.* ©1988 IEEE.)

diffusion source to form ultra-shallow junctions in bulk silicon with raised source and drain areas. This raised source and drain landing pad can also minimize metallization spiking with shallow junction as a barrier. This device and LI structures could provide a fundamental solution to the shallow-junction issues of scaled MOSFET devices in ULSI.[29]

9.2.8 Self-Aligned Structures and Contact

Self-alignment is a technique in which multiple levels of regions on the wafer are formed using a single mask, thereby eliminating the alignment tolerances required

Micron Ex. 1029, p. 516
Micron v. YMTC
IPR2024-00911

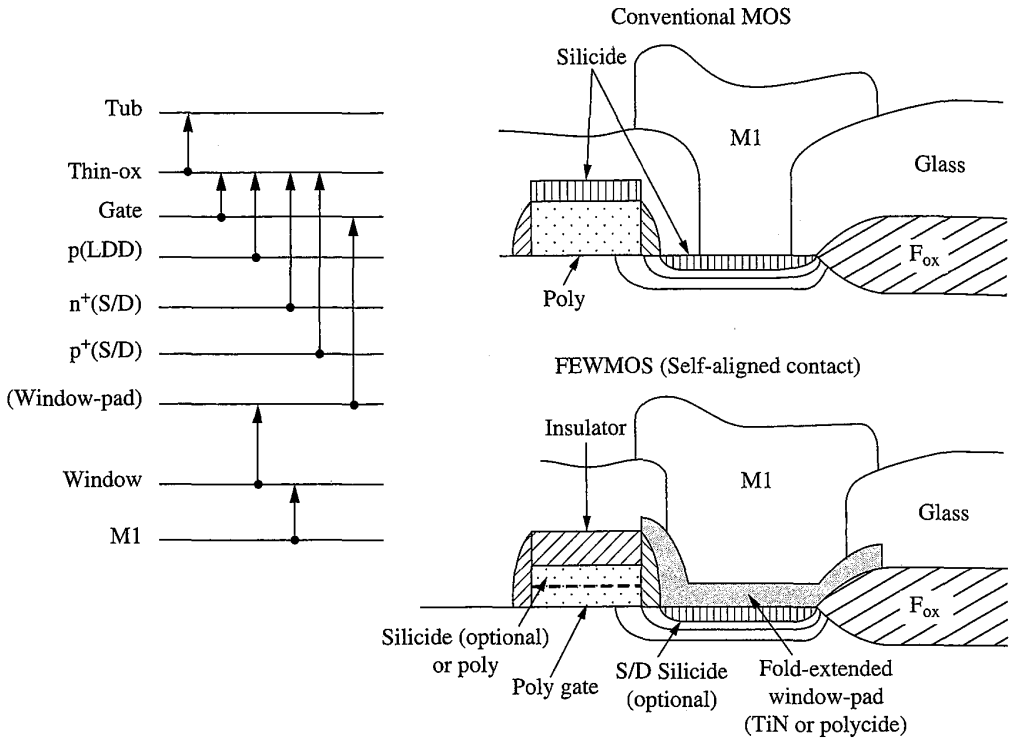


FIGURE 10
Cross-section comparison of the conventional LDD MOS structure and the FEWMOS LDD structure, and the alignment sequence of various levels. (*After Lu et al., Ref. 29.* ©1988 IEEE.)

by additional masks. This powerful approach is being used more often as circuit sizes decrease. There are many examples of this technique; one of the earliest and most widely applied is the self-aligned source and drain implant to the poly gate. Self-aligned contacts are often used in memory cells where contacts are limited only by the spacers and field oxide bird's beaks or a contact window landing pad. Therefore, the mask contact window can be oversized to underneath contact area, and no contact borders are needed, resulting in significant space saving[29, 30] (see Fig. 10).

9.2.9 Multilevel Interconnect

A key requirement of submicrometer technology is a high-density, high-reliability metallization system to interconnect the many active devices on a single substrate. Usually more than one level of interconnect is needed in ULSI circuits. There are three main challenges in the implementation:

Planarization of intermetal dielectric layers.
Filling of high-aspect-ratio gaps between metal space, contact holes, and vias.
Integration of new conductor materials that exhibit low resistance, high reliability, and process compatibility.

Micron Ex. 1029, p. 517
Micron v. YMTC
IPR2024-00911

The important subject of multilevel metallization is treated thoroughly in Chapter 8.

9.3 CMOS TECHNOLOGY

9.3.1 Introduction

The pairing of complementary n- and p-channel (CMOS) transistors to form low-power ICs was originally proposed in the early 1960s.[31] For much of its history, the CMOS transistor was considered to be only a runner-up for the design of MOS ICs. It was not until the severe limitations in power density and dissipation occurred in NMOS circuits that CMOS became the dominate technology for IC manufacturing. CMOS technology employs both NMOS and PMOS transistors to form logic elements. The advantage of CMOS technology is that its logic elements draw significant current only during the transition from one state to another and draw very little current between transitions, allowing power to be conserved.

The circuit diagram of a CMOS inverter is illustrated in Fig. 11*a*. The cross section of the inverter structure (Fig. 11*b*) shows the n-channel transistor formed in a p-region called a tub or well. The p-channel transistor is formed in the n-substrate. The gates of the transistors are connected to form the input. If we define the threshold voltage of n-channel and p-channel transistors as $V_{T\mathrm{n}}$ (e.g., 0.6 V) and $V_{T\mathrm{p}}$ (e.g., -0.6 V), respectively, we can track the output voltage V_o, shown in Fig. 11*c*, as a function of the input voltage V_i. The transistor current for the individual transistors is shown in Fig. 11*d* as a function of the input voltage. At $V_i = 0$, the n-channel transistor is off, since $V_i < +0.6$ V; but the p-channel is on, since $V_i = 0 > V_{T\mathrm{p}}$ (-0.6 V). As the gate voltage of the n-channel transistor is raised above $V_{T\mathrm{n}}$, it begins to conduct current, which flows into the p-channel transistor. The p-channel transistor is on at this point since the magnitude of its gate voltage is well above the threshold voltage. Continuing to increase V_i will bring the gate voltage of the p-channel device closer to the p-channel threshold voltage and eventually turn the device off. The important point to remember is that in either logic state, where V_o is either V_{DD} or V_{SS}, one of the two transistors is off; since the transistors are in series, negligible current is conducted through the inverter. Significant current is conducted through the inverter only when both transistors are on during the transition, as shown in Fig. 11*d*. Therefore, in either logic state, very little current flows from V_{SS} to V_{DD} and very little power is consumed. The low power consumption of CMOS is one of its most important attributes for high-density applications.

9.3.2 Fabrication Process Sequence

There are many variations in the design of a CMOS process.[32] The version presented here is the mainstream approach and illustrates the general sequence of process steps and mask layers used in ULSI applications.[5]

Micron Ex. 1029, p. 518
Micron v. YMTC
IPR2024-00911

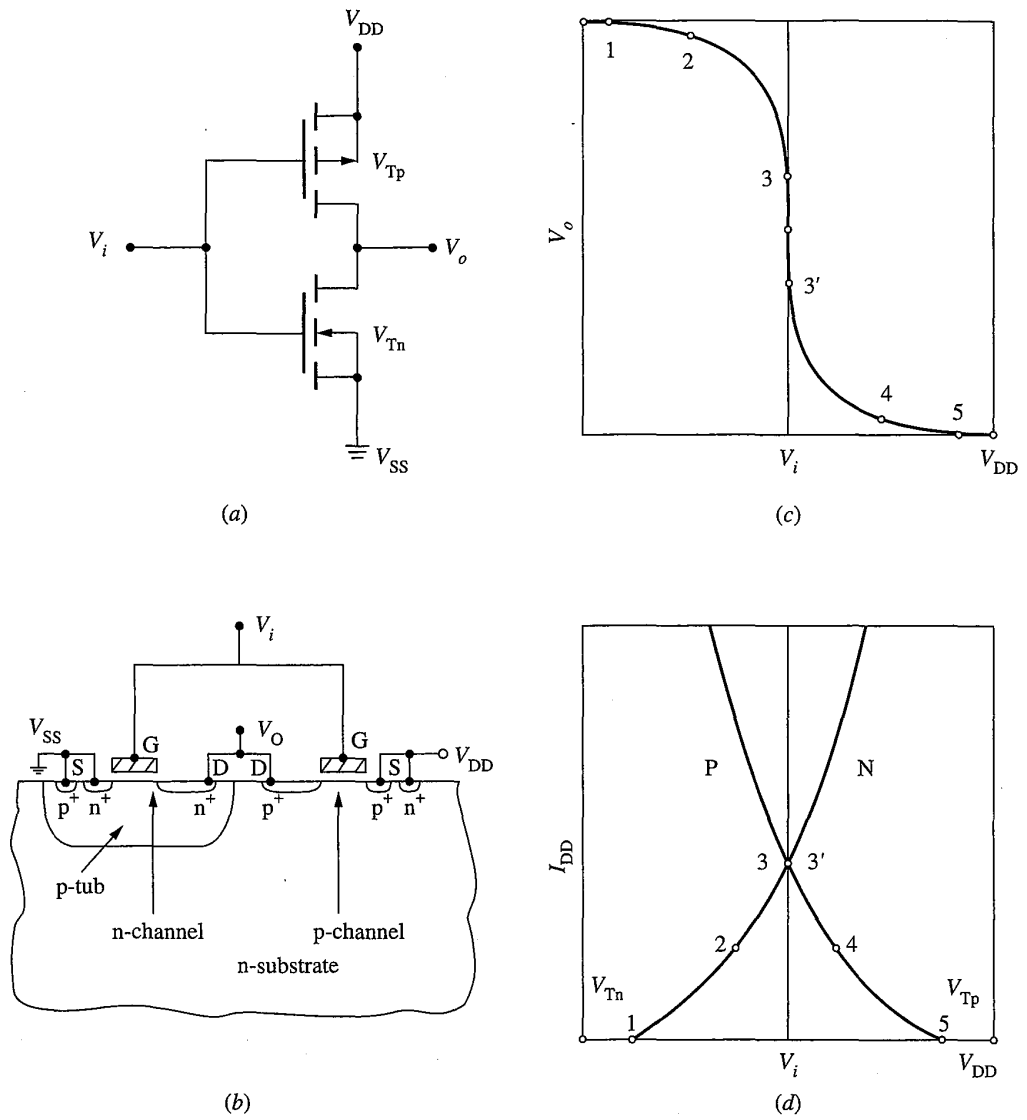


FIGURE 11
CMOS inverter. (*a*) Circuit schematic; V_{DD} and V_{SS} are the highest and lowest circuit potentials, respectively. (*b*) Device cross section. (*c*) Output (V_o) versus input (V_i) voltage of inverter. (*d*) Current through inverter as a function of input voltage (solid curve); *I-V* characteristics of n- and p-channel transistors (dashed curves). The numbers correspond to different points on the inverter transfer characteristic. (*After* B. Hoefflinger and G. Zimmer, "New CMOS Technologies," in J. Carrol, ed., *Solid-State Devices 1980,* from *10th European Solid-State Device Research Conf., Sept. 1980,* Institute of Physics Conference Series **57,** 1980.)

The AT&T Twin-Tub VI 0.6-μm CMOS process[5] sequence is shown in Fig. 12. The CMOS devices are fabricated with p^--epi on p^+ substrate wafers. Conventional twin-well (twin-tub) processing using high-pressure oxidation (HIPOX) to grow a LOCOS field oxide is used.[33] A selective arsenic implant and a deep phosphorus implant are performed to form the n-well followed by a selective well steam oxida-

Micron Ex. 1029, p. 519
Micron v. YMTC
IPR2024-00911

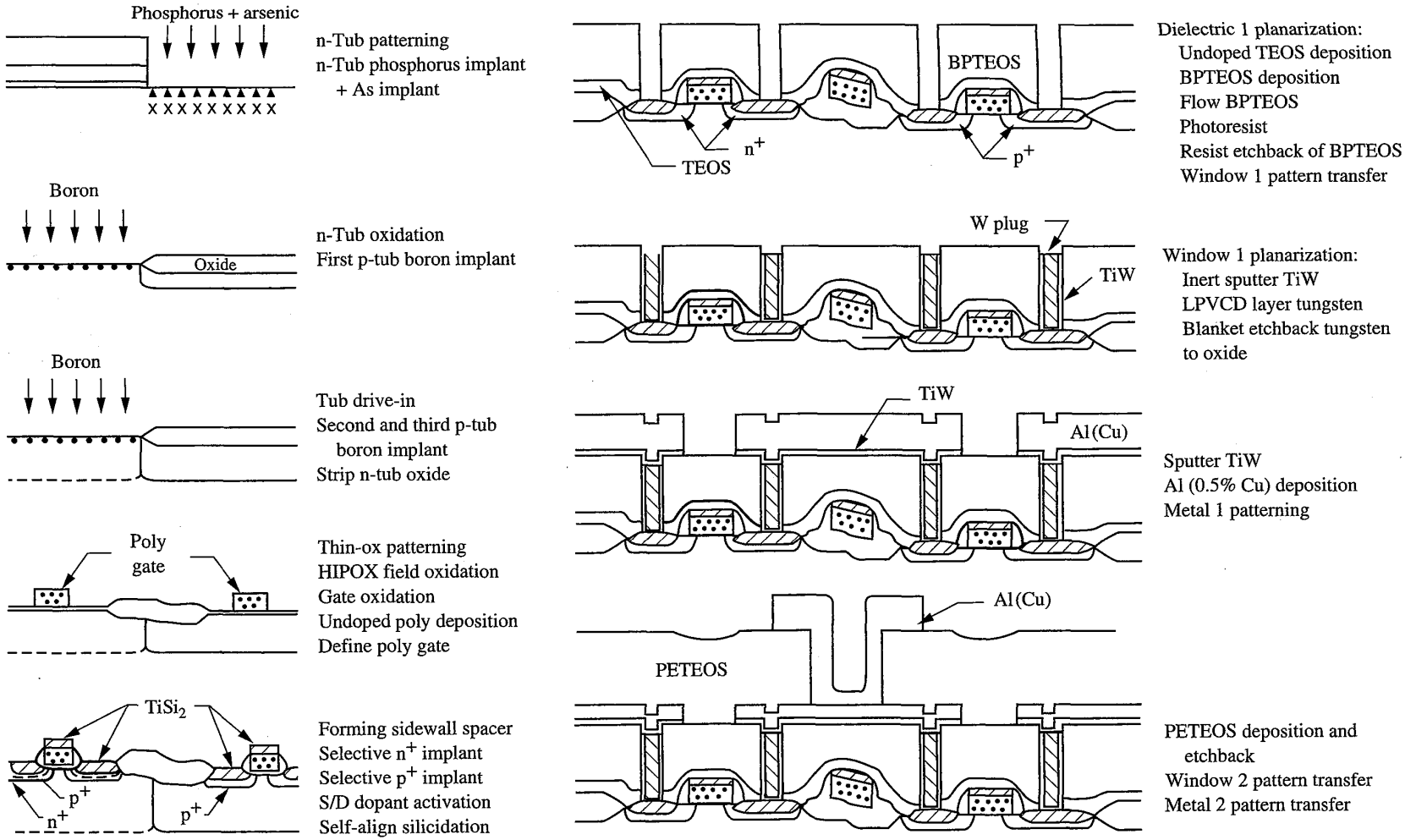


FIGURE 12
Process sequence of the AT&T Twin-Tub VI CMOS process. (*After Lu et al., Ref. 5.* ©1989 IEEE.)

Micron Ex. 1029, p. 520
Micron v. YMTC
IPR2024-00911

tion at 1050°C. By using the well oxide as a mask, a shallow p-well boron implantation is done followed by the 1050°C-well drive-in in an N_2 ambient. After the well drive-in, a deep boron implant is performed to prevent device punch-through. The final depth of both wells is about 1.8 μm. No channel stop implant is needed.[33]

After the 12.5-nm gate oxidation, a layer of undoped low-pressure chemical vapor deposition (LPCVD) polysilicon is deposited as the gate material. For the gate doping, either a blanket arsenic implant is performed followed by a selective low-energy boron implant to counter-dope the poly on p-channel device regions, or the polysilicon is left undoped. After gate patterning, a thin layer of dry oxide is grown on the poly and the substrate. The lightly-doped drain implantation may be performed to reduce hot-carrier effect. A sidewall spacer is then formed by depositing LPCVD tetraethylorthosilicate (TEOS) oxide and then anisotropically etching it back by reactive ion etch (RIE).

The n^+ region is then patterned and formed by a high-dose arsenic implant followed by a lower-dose phosphorus implant. The high-dose arsenic implant amorphizes the Si substrate and, hence, eliminates the phosphorus implantation tail. After n^+ photoresist mask stripping, the p^+ region is patterned and doped by a 30-keV BF_2 implant or by a 20-keV boron implant with a Si substrate preamorphization by Si implant. If a BF_2 implant is chosen to form the p^+ junction, the high-temperature junction drive of an 875°C anneal could be shared by n^+ and p^+ junctions to anneal the implantation damage and push the junctions away from the residual damage. However, if a preamorphized boron implant is used, the 875°C anneal is too high for the p^+ shallow junction; therefore, this high-temperature (875°C) anneal should be performed immediately after the n^+ implant. After silicon preamorphizing and boron implanting, an additional, brief 800°C anneal is performed to drive both junctions slightly. In order to fully activate the dopants without pushing the junction too deep, a high-temperature rapid thermal anneal (RTA) is used. For the BF_2-implanted p^+n junction, the RTA is essential for low parasitic series resistance ($\leq$ 2.2 kΩ/μm) of p-channel devices. However, for a preamorphized boron p^+ junction, the RTA is not as critical to obtain low parasitic resistance for pMOSFETs. In general, the process with the BF_2 implant is simpler, gives a shallower junction, and has better junction leakage properties. This activation RTA step also alleviates the gate depletion problems for implant doped gates.[34]

Following the junction formation, a brief HF cleaning is performed to remove the residue oxide on the source/drain/gate regions, and a 60-nm titanium film is sputtered followed by Si implant for ion mixing. Due to shallow junctions, this Si-implant energy needs to be adjusted to be optimum for the least junction leakage and to maximize the ion-mixing effects for salicide formation. RTA is used to form the $TiSi_2$ of phase C49 at 625°C. The unreacted Ti is then etched away by a wet Ti etch solution followed by a 900°C RTA to change C49 $TiSi_2$ into a low-resistivity and stable phase (C54) of $TiSi_2$. A nominal $TiSi_2$ thickness of 60 nm and a sheet resistance of 2.5 Ω/□ has been obtained. The actual junction depths obtained from SIMS (secondary-ion mass spectroscopy) analysis are 140 and 180 nm underneath the $TiSi_2$ interface for the n^+p and p^+n junctions, respectively. About 30 nm of Si was consumed in the $TiSi_2$ process. Then passivation glass is deposited, window holes are opened, and metallization is performed.

Micron Ex. 1029, p. 521
Micron v. YMTC
IPR2024-00911

9.3.3 Special Considerations

Latch-up

A major problem in CMOS circuits is the inherent self-destructive phenomenon known as latch-up, in which a very-low-resistance path is established between the V_{DD} and V_{SS} power lines, allowing large currents to flow through the circuit.[35] This can cause the circuit to cease functioning or even destroy itself from heat damage. The susceptibility to latch-up arises from the presence of complementary parasitic bipolar transistors contained in the CMOS structures (see Fig. 13); when in close proximity to one another, they can interact electrically to form pnpn device structures, which behave like thyristors. Such devices normally do not conduct; however,

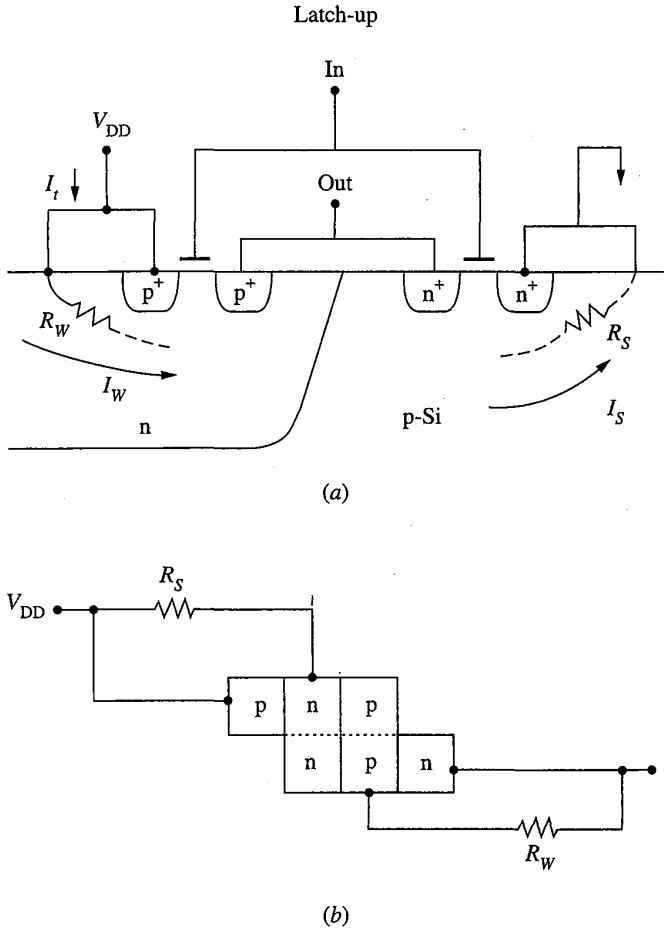


FIGURE 13
(*a*) Cross section of a CMOS inverter. R_w and R_s represent the resistances associated with the well and the substrate, respectively. (*b*) Circuit and schematic representation of the cross-coupled parasitic npn and pnp transistors.

Micron Ex. 1029, p. 522
Micron v. YMTC
IPR2024-00911

a triggering current may forward bias a pn junction, and then the structure will fire into its highly conductive mode and remain so even after the trigger ceases.[2,3]

There are various measures that can minimize the risks of latch-up:[35]

To reduce the current gain of the CMOS vertical parasitic bipolar transistor, create a deeper well or a retrograde well.

To reduce the resistances of the parasitic current paths, use a heavily doped substrate with devices fabricated on a lightly doped epi-layer.

To put current sinks in the parasitic current paths, implement changes through design rules, such as putting p^+ guard rings around n-channel devices and increasing the separations between the n- and p-channel devices.

Dual polysilicon gates

For ULSI CMOS circuits in the submicrometer range, a major issue is the scaling of the PMOS transistors. For ease of fabrication, n^+ doped polysilicon gates are used for both n-channel and p-channel devices. Asymmetry results because of work-function differences for the two types of devices; this creates a large native p-channel threshold voltage for pMOSFET,[36] which requires adjustment to the appropriate magnitude with a boron compensation implant. This implant could be large enough to counter-dope the surface region of the n-well, creating the "buried channel" structure. Such a structure can produce drain-to-source punch-through current at short channel lengths because there is less voltage control of the gate. This situation has been tolerated down to 0.35-μm pMOSFET or in the case where PMOS is used as loads. A great improvement can be made if a p^+-doped gate is used for the p-channel devices. This makes both the n-channel and p-channel transistors completely symmetrical in their threshold voltages, device dimensions, and dopings; and the p-channel transistor becomes a surface channel device. Such a device has been made successfully by a number of workers[5,34,37] and has even been put in production, although at the cost of more process complexity. To produce such a device, undoped polysilicon gates are patterned first, and they are doped at the same time the source and drain of each type of device are implanted.[5] Some special processing problems need to be considered: one problem is that, at high processing temperatures, boron tends to diffuse through the gate oxide very rapidly into the channel region, especially in the presence of hydrogen-related species such as water vapor.[38] If this happens, the threshold of the device will change, or it may even become a depletion mode device. Therefore, special care must be taken to ensure that the furnace operations are free of hydrogen.[39]

Another concern is that, since the individually doped poly gates are formed by implantation, the doping species must be sufficiently activated, otherwise a depletion layer may form at the poly/oxide interface during device operation. This can result in a significant reduction of the transconductance of the device.[34]

In the dual-polygate CMOS fabrication process, BF_2 implantation was usually used simultaneously for pMOSFETs' source, drain, and gate implant doping; however, fluorine incorporated into p^+-gate devices can jeopardize the PMOS stability even under moderate annealing conditions.[40,41] It was demonstrated that fluorine can severely influence the p^+-gate PMOS device characteristics in addition to the hydrogen effects. These threshold voltage instability and anomalous edge

Micron Ex. 1029, p. 523
Micron v. YMTC
IPR2024-00911

effects were proven to be caused by the enhanced boron penetration through the thin gate oxide[40,41] and by the generation of negative-charge interface-traps in the gate oxide.[40,42] These effects have an impact on any ULSI process steps for dual-polygate MOSFET device fabrication that involve fluorine in addition to the H_2/H_2O-containing ambient anneal.[39]

Another concern in the dual-polygate CMOS process is the interdiffusion of dopants between the two different doping-type regions through the neighboring horizontal junctions. The interdiffusion is especially significant through the shorting overlayer of metal silicide, if a salicidation structure is used. Dopant lateral diffusions in n^+/p^+ dual-polygate CMOS were studied in $CoSi_2$,[37] $TaSi_2$,[43] and $TiSi_2$,[44] and it was found that lateral diffusion occurs even at 800°C annealing in $TiSi_2$ and at higher temperatures in other silicide layers. Such lateral diffusion of dopants can compensate the normal dopant concentration or even convert the dopant type of the underlying polysilicon of the neighboring gate that shares the common polycide runner, resulting in large variations in threshold voltages of n- and p-channel MOSFETs.

9.4 BIPOLAR TECHNOLOGY

9.4.1 Introduction

Bipolar technology was the only available technology in early IC production. With better understanding and improvements in the oxide-silicon interface, bipolar technology was surpassed by PMOS and NMOS technologies. Subsequently, CMOS has become the mainstream technology since the early 1970s. Although they have long held a distant second place to CMOS ICs with their high circuit density, bipolar ICs have nevertheless been progressing rapidly, approaching densities of 100 k gates per chip.[45] Bipolar IC density still falls far short of ULSI density, but bipolar ICs offer unique advantages; they have faster speed and higher current-driving capability than CMOS ICs, and they are supplanting CMOS ICs as the random logic of choice in top-of-the-line digital electronic equipment.

Interestingly, the advances in bipolar ICs result from adaptations of the strategies of rival technologies; most of the new developments in MOS process equipment and technology are also applicable to bipolar technology. New developments in ion implantation, epitaxial growth, and polysilicon deposition have made extremely thin-base transistors possible, with their base diffusion extending less than 150 nm into the silicon.

9.4.2 Standard Bipolar Process

The three types of planar structures for the early bipolar IC all used a reverse-biased pn junction to provide isolation between collectors on the same chip. The three structures are (1) standard buried-collector transistor (SBC),[46] (2) collector-diffused isolation transistor (CDI),[47] and (3) triple-diffused transistors (3D).[48]

Micron Ex. 1029, p. 524
Micron v. YMTC
IPR2024-00911

A transistor fabricated with the SBC process has a collector region consisting of a heavily doped n^+ buried layer and a lightly n-doped epitaxial layer. The collectors are isolated from one another by p-diffusions that extend all the way through the epi-layer to contact the p-substrate. The base and emitter regions are formed

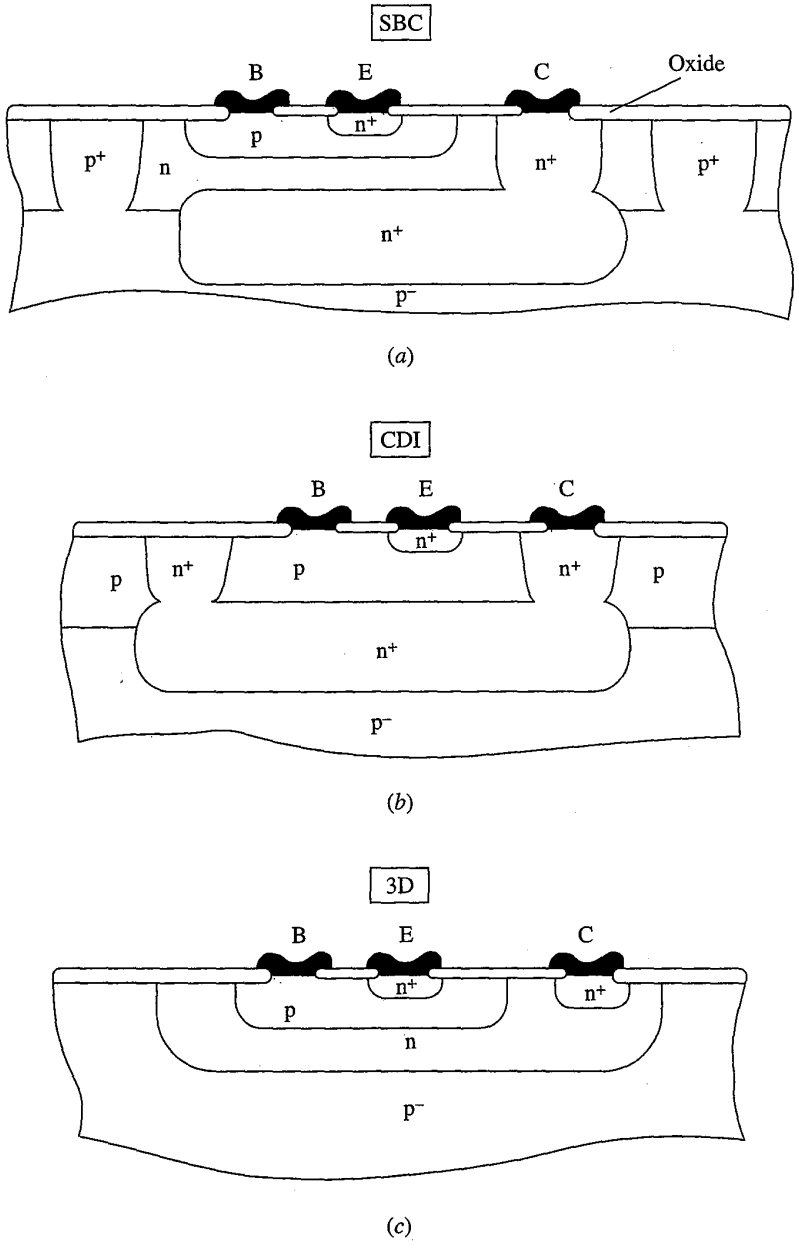


FIGURE 14
Cross sections of three junction-isolated processes: (*a*) Standard buried-collector (SBC) process, (*b*) collector-diffused isolation (CDI) process, and (*c*) triple-diffused (3D) process. (*After Wolf, Ref. 52.*)

Micron Ex. 1029, p. 525
Micron v. YMTC
IPR2024-00911

by diffusion after the isolation structures have been created (Fig. 14*a*). In the CDI process, an n^+ ring completely encircles the device and merges with the bottom buried n^+ layer so that the transistor is completely isolated by the collector/substrate junction (Fig. 14*b*). The 3D process has no buried or epitaxial layers (Fig.14*c*). Instead, the collector as well as the emitter and base regions are formed by means of implanting ions directly into the substrate. Figure 14 shows the cross-sections of these junction-isolated processes.

The large sizes of these junction-isolated transistors make it difficult for them to meet the functional density requirement of ULSI applications and also cause them to have high parasitic capacitances. By replacing junction isolation with oxide isolation, it is possible to shrink the transistor size and hence the propagation-delay times. In this section, only the oxide-isolated structures that have self-aligned poly emitters will be discussed, since these structures are more suitable for ULSI applications.[49–52]

9.4.3 Self-Aligned Bipolar Structures and Fabrication

The most widely used state-of-the-art bipolar approach employs a double-polysilicon process to form an emitter that is self-aligned to the base, as shown in Fig. 15. The first layer of polysilicon (poly 1) is doped p^+ and serves as the base electrode, whereas the second layer (poly 2) is doped n^+ and makes contact with the transistor's emitter region. The emitter and the extrinsic base regions are formed by means of dopant outdiffusion from poly 2 and poly 1, respectively.

Figure 16 shows the sequence of steps that form this self-aligned structure.[51,52] After the field oxide is grown and the oxide is removed from the surface of the active regions, poly 1 is deposited and heavily p-doped with boron. This layer is then covered with a CVD oxide (Fig. 16*a*). The emitter mask is used to pattern the emitter-area regions, and a dry-etch process is used to produce openings in the CVD oxide and poly 1 (Fig. 16*b*). A thermal oxide is then grown over the etched structure, and a relatively thick oxide (approximately 0.1 to 0.4 μm) is grown on the vertical

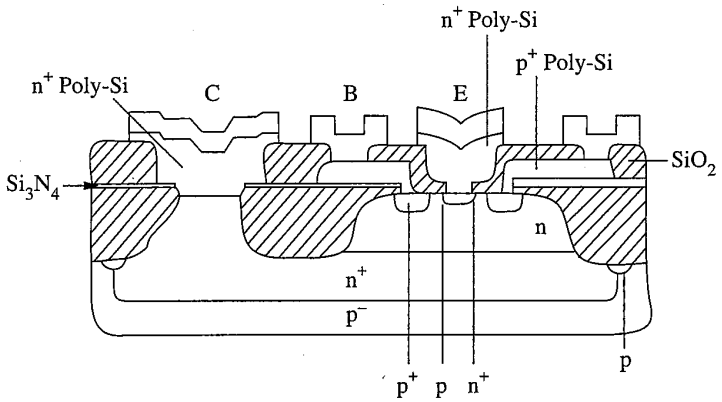


FIGURE 15
Cross section of a double-polysilicon self-aligned bipolar transistor. (*After Wieder, Ref. 49.* ©1986 IEEE.)

Micron Ex. 1029, p. 526
Micron v. YMTC
IPR2024-00911

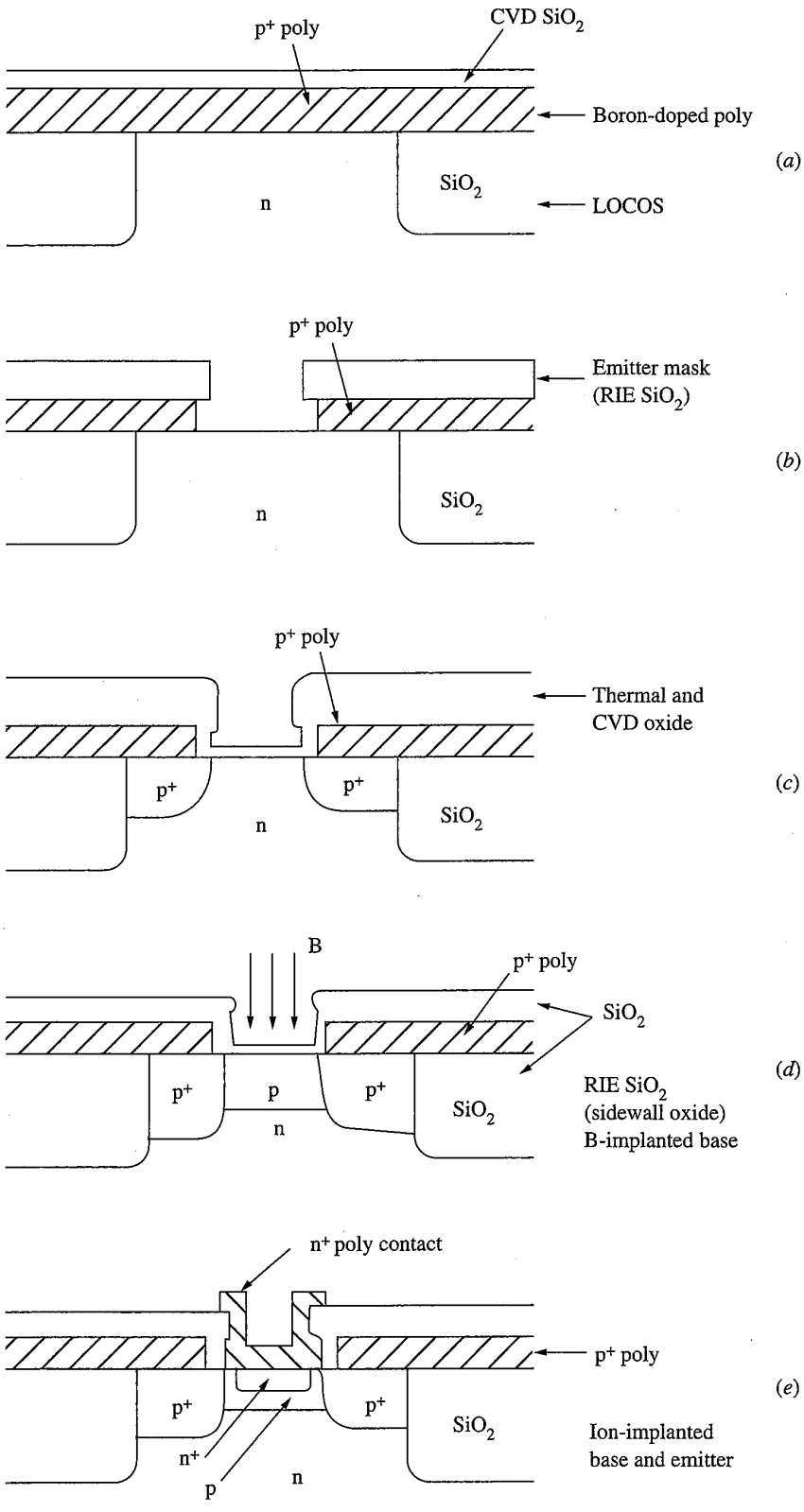


FIGURE 16
Process sequence for the fabrication of double-polysilicon self-aligned npn bipolar transistor. (*After Tang et al., Ref. 50.* ©1980 IEEE.)

Micron Ex. 1029, p. 527
Micron v. YMTC
IPR2024-00911

sidewalls of the heavily doped poly. The thickness of this oxide determines the spacing between the edges of the base and emitter contacts. The extrinsic base regions are also formed during the thermal-oxide growth step as a result of the outdiffusion of boron from the poly 1 into the substrate (Fig. 16*c*). Because boron diffuses laterally as well as vertically, the extrinsic base region will be able to make contact with the intrinsic base region that is formed next, under the emitter contact.

The parasitic series resistance of the extrinsic base region, R_{BI}, will now consist of the sum of the resistances of the poly 1 electrode, R_{BIpoly}, and of the path between the edges of the base and emitter contact, R_{BIsub}. Since the latter distance is so small, R_{BIsub} is negligibly small. The doping of poly 1 can be much heavier than that of the extrinsic base region in the substrate, and the value of R_{BIpoly} can therefore be made much smaller than R_{BI} in a nonself-aligned bipolar transistor.

Following the oxide-growth step, the intrinsic base region is formed using ion implantation of boron (Fig. 16*d*). This serves to self-align the intrinsic and extrinsic base regions. After the contact is cleaned to remove any oxide layer, poly 2 is then deposited and implanted with As or P. A shallow emitter region is then formed through dopant outdiffusion from poly 2 (Fig. 16*e*). The use of rapid thermal anneal for the base and emitter outdiffusion steps has also been reported to facilitate the formation of shallow emitter-base and collector-base junctions.

This self-aligned structure allows the fabrication of emitter regions smaller than the minimum lithographic dimension. When the sidewall-spacer oxide is grown, it fills the contact hole to some degree because the thermal oxide occupies a larger volume than the original volume of polysilicon. Thus, a 0.8-μm-wide opening will shrink to about 0.4-μm if a 0.2-μm-thick sidewall oxide is grown on each side. The relatively conventional process steps of borophosphosilicate glass (BPSG) deposition, contact opening, reflow, and metallization follow.

9.4.4 Special Considerations

High-performance bipolar transistor

An optimal bipolar device requires a minimum number of masking steps, very few critical alignments, and has very high yields. It should also have the following characteristics:

Low junction capacitances
Low series emitter and collector resistances
Low base transit time
Good dc gain
High breakdown voltages

Scaling

Bipolar performance is improved by shrinking the device in both the horizontal and vertical directions. The planar surface geometry shrinking has been spurred by advances in photolithography, etching, and various self-alignment techniques. Vertical scaling is very important, because junction depth must be as shallow as possible to minimize capacitances and transit time, while maintaining adequate breakdown

Micron Ex. 1029, p. 528
Micron v. YMTC
IPR2024-00911

voltages and dc gain. To get a feel for this problem, note that the transit time is proportional to the electrical base-width squared, W^2, and a small increase in W can have a large and detrimental effect on transit time. However, reducing the base junction X_j has the undesired effect of reducing the punch-through voltage, because it would be easier for the collector depletion region to touch the emitter. Increasing the base impurity concentration will reduce the depletion spread but will increase capacitances and reduce gain. A very shallow emitter junction would give lower emitter efficiency, lowering the dc current gain. This problem is alleviated by using the poly emitters,[53] as explained later.

Vertical impurity profile

Figure 17 gives a general doping profile of a bipolar transistor along a cross section perpendicular to the surface, passing through the emitter, base, and collector. The collector region consists of a heavily doped n^+ buried layer and a lightly doped epitaxial layer. The base and emitter regions are formed by implants and subsequent diffusions after isolation structures are created. The lightly doped side of the collector determines the breakdown voltage of the collector-base junction, BV_{CBO}, caused by avalanche breakdown, except in the case of punch-through. The breakdown voltage of the emitter-base junction, BV_{EBO}, is limited by the lightly doped base

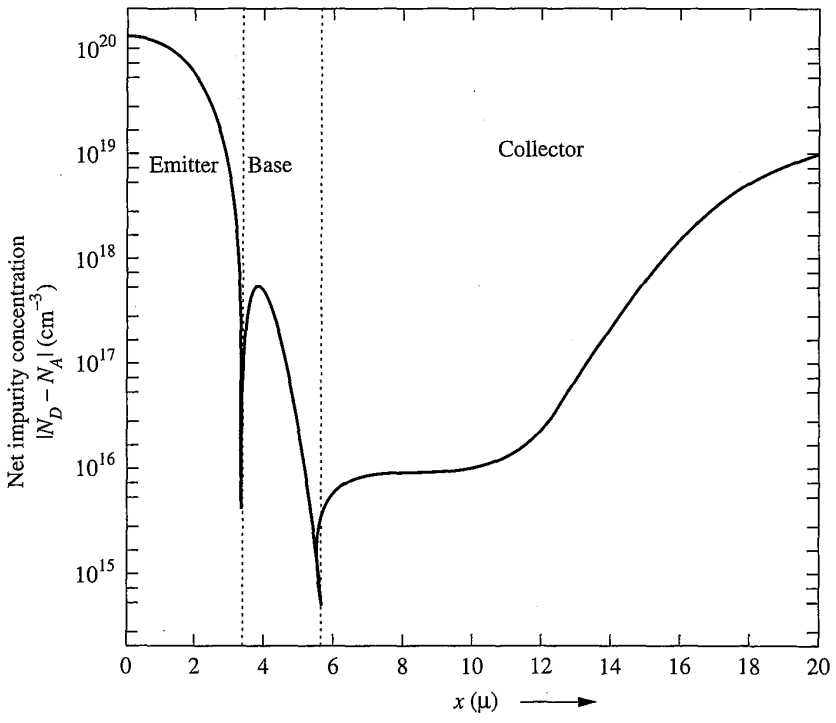


FIGURE 17
The impurity distribution in an SBC bipolar transistor. $|N_d - N_A|$ is the net impurity concentration, and x is the depth into silicon surface. (*After Wolf, Ref. 52.*)

Micron Ex. 1029, p. 529
Micron v. YMTC
IPR2024-00911

region. Since the base is more heavily doped than the collector, BV_{EBO} is much smaller than BV_{CBO}. (For a general discussion on bipolar device physics, see Ref. 54.) The combined structure of the n^+ buried collector and the lightly doped epi allows high BV_{CBO} and low R_c.

Poly emitter structure

The poly emitter has the following advantages:

The lateral and vertical dimensions of the emitter can be scaled more easily.
Since the emitter is formed by outdiffusion of dopant from the polysilicon in direct contact with the substrate, it does not have the severe yield problem that occurs in implanted emitters because of the residual damage of the implant.
Aluminum contact spiking in the emitter-base junction is minimized.

The current gain of transistors fabricated with a poly emitter is three to seven times that of transistors made with conventional emitters (e.g., emitter depth is approximately 200 nm, and base depth is approximately 100 nm). Transistors with poly emitters also have a higher cut-off frequency, f_T, of 17 to 30 GHz and less than 50 ps gate delay.[55] An interfacial oxide tunneling model has been postulated to explain this increase in current gain: An unavoidable, thin native oxide always forms on the substrate before polysilicon deposition; this acts as a tunneling barrier to hole injection into the polysilicon, and thus increases the emitter efficiency.[56] Another reason for the higher current gain may be enhanced impurity segregation at the substrate/poly interface. SIMS analysis confirms that such a dopant pileup does occur.[57] This impurity-segregation effect gives rise to a potential-energy barrier that hinders the injection of holes into the emitter during forward bias and decreases the slope of the hole concentration in the emitter, leading to a smaller hole current and an increased current gain.[58]

In processing, the nonuniform, variable nature of the unintentionally grown interfacial oxide has caused poor reproducibility, especially as devices are being scaled down. A manufacturable process flow is described as follows: The emitter driving-in step is done by RTA at 1050°C, preceded by a careful cleaning and a HF dip just before poly deposition. This results in a reproducible native oxide of approximately 0.5 to 1 nm, and allows a shallow (approximately 50-nm) As-doped emitter region to be regrown without disrupting the continuous grain boundary at the poly/mono silicon interface. By following this procedure the emitter resistance can be reduced to below 30 Ω-μm^2. Better results can also be obtained by depositing poly at 275°C using a plasma-enhanced (PE) CVD process.

Base regions

A major obstacle to small base width, W_B, in advanced bipolar devices is the channeling tail of the intrinsic-base boron implant. One way to remedy this is to eliminate the direct implantation of boron into the substrate. Instead, after the sidewall spacer oxide is grown, poly 2 is deposited and implanted with boron. After the intrinsic base is formed by outdiffusion, As^+ is implanted. This approach has produced npn transistors with a base width of 100 nm, an emitter junction depth of 50 nm, and cut-off frequencies of 16 GHz.[59]

Micron Ex. 1029, p. 530
Micron v. YMTC
IPR2024-00911

The connection between the extrinsic and intrinsic base regions in the double-poly self-aligned bipolar structure is critical. The space between the edge of the base and the emitter contacts is determined by the thickness of the sidewall oxide spacer. The amount of overlap between the two base regions depends on the spacer thickness and the lateral diffusion from the p^+ poly. Control of these two parameters is vital in order to fabricate the structures with high yield. If the overlap between the intrinsic and extrinsic regions is too large, low values of BV_{EBO}, β, and f_T result. On the other hand, if the space between the edges of the two regions is too great, the result will be anomalously high R_B values and low values of perimeter punch-through voltage between the emitter and collector through the lightly doped link region of the base.

Injection of minority carriers into the base occurs at the sidewalls as well as at the bottom of the emitter. It can be seen that the current gain resulting from transistor action from the sidewall injection is much smaller than that from the bottom injection. This is because both the base width and the doping concentration are smaller in the base region beneath the emitter than they are along its sidewalls. For devices with emitter dimensions smaller than 1 μm, the perimeter-to-area ratio may increase to such a degree that common emitter current gain, β, can be significantly degraded by the perimeter injection effect.

Super-high-performance SiGe HBT

One of the promising new developments for bipolar technology is the heterojunction bipolar transistor (HBT) using an alloy of silicon and germanium as base material. The Ge_xSi_{1-x} alloy has a narrower band-gap for the base region than silicon and allows an appreciable increase in emitter efficiency. However, these devices are not even in the LSI stage yet.

9.5 BiCMOS TECHNOLOGY

9.5.1 Introduction

BiCMOS is a technology that integrates both CMOS and bipolar device structures on the same chip. CMOS can offer low power and high density to a digital IC, but it is usually slower than the emitter-coupled logic (ECL)-based IC. Bipolar transistors, on the other hand, can deliver large drive currents, operate with small logic swings, and have high noise immunity; however, they exhibit high power consumption, poor density, and limited circuit options. BiCMOS offers the benefits of both bipolar and CMOS circuits.[60] By appropriately trading off the characteristics of each technology, speed and power can be balanced. This desirable result, however, is attained with the penalty of adding more process complexity. So far BiCMOS has penetrated the application-specific IC (ASIC) and very-high-speed SRAM market. Even though it may not become the dominant IC technology, it will carve out a sizable niche in the high-performance digital and analog arena.

Micron Ex. 1029, p. 531
Micron v. YMTC
IPR2024-00911

9.5.2 Different Process Design Approaches

There have been two major approaches to BiCMOS process design. The low-end BiCMOS process incorporates a moderate-speed bipolar transistor into an aggressive CMOS process. The high-end BiCMOS process extends an aggressive bipolar process to include MOS devices. In low-end BiCMOS, the overall objective is to improve performance with a minimum increase in cost. CMOS attributes, large component density and low power dissipation, are retained while performance is boosted by the application of bipolar devices used selectively on the chip. The high-end BiCMOS usually incorporates large memory blocks using MOS elements of high-speed ECL peripheral circuits[54] to implement mass memory with fast access time.

9.5.3 Fabrication Process Sequence

Since BiCMOS is a mixture of two technologies in various proportions, one can be dazed by the sheer variety. First we consider the simplest kind of BiCMOS, and then we add on more features and finally discuss the full-fledged technology.

Triple-diffused (3D) BiCMOS

The simplest way to add an npn bipolar transistor is to use the PMOS n-well as the collector of the bipolar device and introduce an additional mask level for the p-base region.[61] Figure 18 shows a cross section of a typical structure. The NMOS device is built in a 15 to 10-μm-thick p-epitaxial layer on top of a p^+ substrate, whereas the PMOS transistor is built in an implanted n-well 4 to 5 μm deep. The p^+ substrate is used to reduce latch-up susceptibility by providing a low-impedance current path through a vertical pnp parasitic device. Poly gates are used for both n- and p-channel devices. The p-type base region of the bipolar device is 1 μm deep, with a doping level of 1×10^{17} cm^{-3}. Many of the processing steps can be

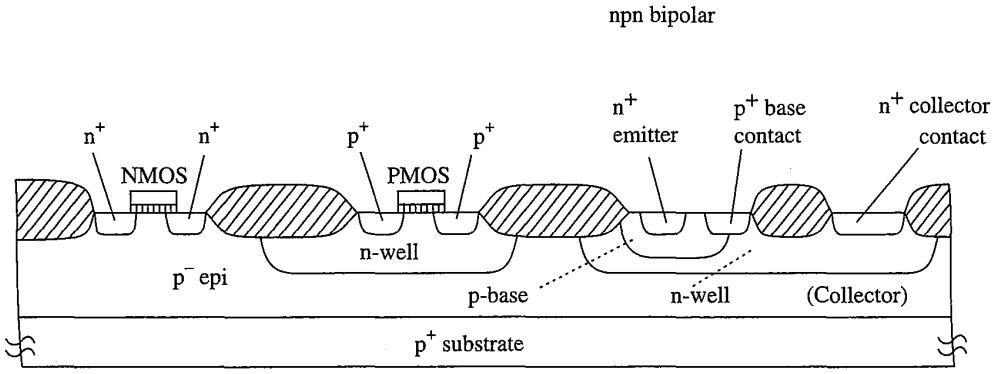


FIGURE 18
BiCMOS structure formed by the simple addition of an npn bipolar transistor to the basic n-well CMOS process. (*After Haken et al., Ref. 62.*)

Micron Ex. 1029, p. 532
Micron v. YMTC
IPR2024-00911

shared between the MOS and bipolar devices. For example, the n^+ source/drain (S/D) doping step is used for the emitter and collector contact of the bipolar structure. The p-channel S/D implant is the same as that used for the p^+ base contact. This simple process is called the 3D BiCMOS process.

Standard buried-collector (SBC) BiCMOS

From a bipolar performance standpoint, the simple approach described above has a number of limitations. The most significant of these is that the lightly doped n-well leads to a high collector resistance, which limits the usefulness of the bipolar transistor. Figure 19 illustrates how this problem is overcome by introducing a buried n^+ layer under the n-well.[60] The buried n^+ layer is first formed in the substrate by arsenic implantation and then capped with a 2-μm-thick n-type epi layer. This n^+ layer not only reduces the collector resistance, but also reduces the susceptivity to latch-up. The p-wells for the NMOS device and the bipolar transistor are then fabricated in the thin n-epitaxial layer. The resistivity of the n-epi is typically 1 Ω-cm and is chosen so that it can support both the PMOS and the bipolar transistors. The collector resistance is further reduced by adding a deep n^+ connection to the buried n^+ layer with an additional mask; this may be either a deep n^+ implant or a poly plug. The p-well step is also used to form the junction isolation structures of the bipolar transistor collector. Thus, with two added mask levels, this approach merges the process steps needed to achieve low bipolar collector resistance with those required to reduce CMOS latch-up susceptibility.[60]

Twin-well BiCMOS

The previous approach produces a bipolar device that, despite much-improved characteristics, still has a number of drawbacks. Its packing density is limited by the p^- substrate doping level and by the spacings that must be used to prevent punch-through from one collector to another. Raising this doping level would increase collector-to-substrate capacitance. Also, the p-type epi-layer has to be counter-doped to form the n-well regions for the PMOS devices. Counter-doping the 2-μm-thick 1 Ω-cm material can cause process control difficulties and can reduce PMOS perfor-

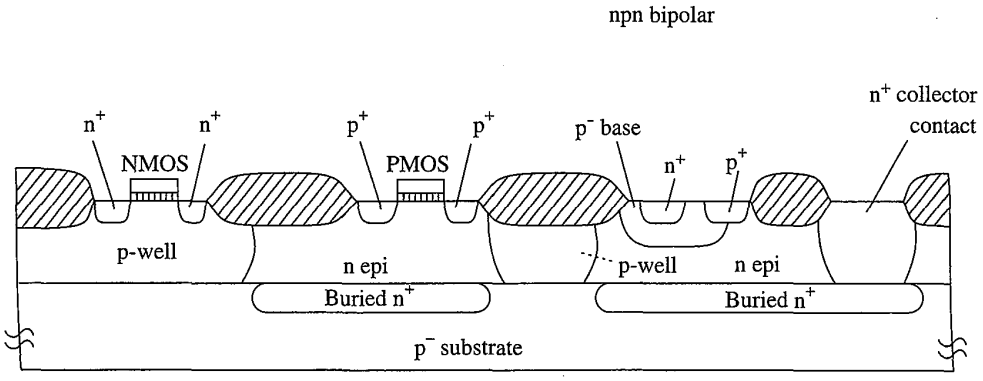


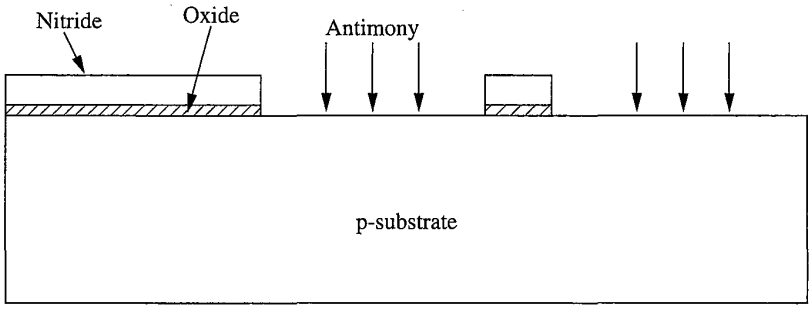
FIGURE 19
BiCMOS structure showing the introduction of a buried n^+ layer and deep n^+ top side contact to reduce collector resistance. (*After Haken et al., Ref. 62.*)

Micron Ex. 1029, p. 533
Micron v. YMTC
IPR2024-00911

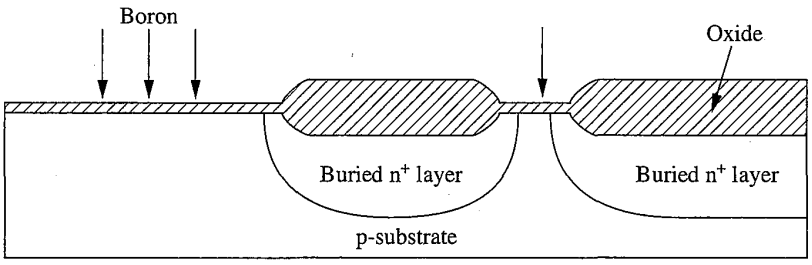
mance by degrading mobility. Figure 20 (*a*)–(*e*) shows how bipolar packing density can be improved by self-aligning buried p-layers to the buried n^+ regions at the cost of higher collector-to-substrate capacitance.[62]

A lightly doped (10 Ω-cm) p^--type substrate is used for the starting wafer. An oxide pad and nitride composite layers are first deposited and patterned for the buried n^+ layer. Antimony is implanted into the windows and annealed at 1250°C to diffuse the antimony and remove defects (see Fig. 20*a*). At the same time, a thick oxide is grown over the buried n^+ regions. This thick oxide serves as a blocking mask for the self-aligned boron-buried p-layer implant, as illustrated in Fig. 20*b*. All oxide is removed from the surface prior to epitaxy, and a short HCl etch is performed in situ to remove native oxides before growing a thin (1 to 1.5 μm) n-epitaxial layer with a resistivity of more than 15 Ω-cm.

The wafer is then oxidized and capped with a nitride layer, which is selectively removed from the n-well regions by a patterning step. After the n-well dopant of phosphorus is implanted, a thick oxide (350 nm) is grown over the n-well regions, and nitride is then stripped from the p-wells. The subsequent p-well implant is self-aligned to the n-well edges (see Fig. 20*c*). A well drive is then performed at 1000°C with the oxide cap in place. All oxide is again removed from the surface; another thin oxide pad and nitride composite layer are deposited, and the active areas are defined (see Fig. 20*d*). Normal CMOS processing steps are then followed using a LOCOS-type isolation process to isolate adjacent active regions. Next, standard NMOS and



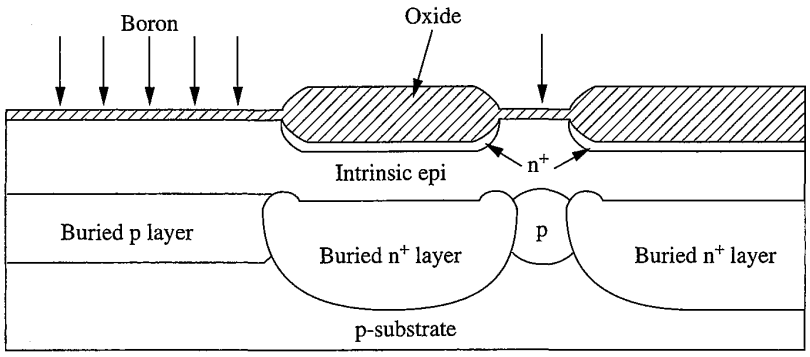
(*a*)



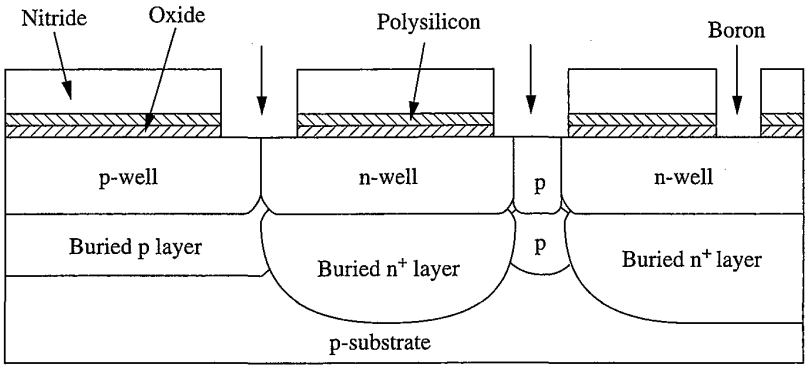
(*b*)

FIGURE 20
(*a*) Device cross section for BiCMOS process flow showing buried n^+ implant. (*b*) Device cross section for BiCMOS process flow showing buried p implant self-aligned to buried n^+. (*Continued*)

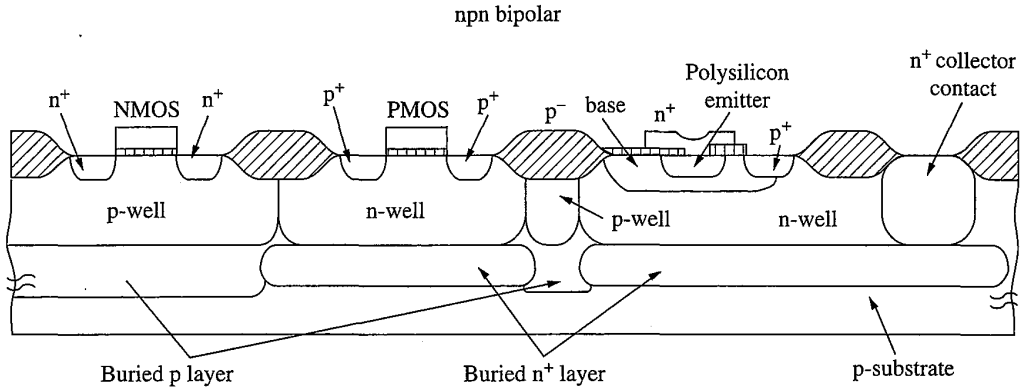
Micron Ex. 1029, p. 534
Micron v. YMTC
IPR2024-00911



(*c*)



(*d*)



(*e*)

FIGURE 20 (***continued***)
(*c*) Device cross section for BiCMOS process flow showing p-well implant self-aligned to oxide-masked n-well. (*d*) Device cross section for BiCMOS process flow showing channel-stop implant after etch of field isolation regions. (*e*) Optimized BiCMOS device structure. Key features include self-aligned p and n^+ buried layers for improved packing density, separately optimized n and p-wells (twin-well CMOS) formed in an epitaxial layer with intrinsic background doping, and a polysilicon emitter for improved bipolar performance. (*After Haken et al., Ref. 62.*)

Micron Ex. 1029, p. 535
Micron v. YMTC
IPR2024-00911

PMOS channel implants are performed to set the MOS device threshold voltages. A contact plug mask is patterned and a high dose of phosphorus is implanted into the npn collector region to reduce its resistance and prevent the device from entering saturation. After gate oxidation, a poly layer is deposited and n^+ doped to form the MOS electrodes (see Fig. 20*e*).

The process described above shows how a twin-well CMOS structure can be implemented without heavily counter-doping the epi-layer. Instead of using an epi-layer in which the doping level is determined by the bipolar and PMOS requirements, a near-intrinsic epi-layer is deposited. Up to this point, four additional mask levels, buried n^+, deep n^+ contact, p-base, and emitter, are required to merge this high-performance bipolar process with a baseline CMOS flow, resulting in a total mask count of fourteen compared to a typical ten-mask CMOS process. Figure 21 compares typical CMOS and bipolar technology process flows to a high-performance BiCMOS technology process flow that has these additional masking steps. All flows are assumed to have LDD MOSFETs, twin-wells, and double-level metal.[62]

Further enhancement

For the cost of an extra mask level, the bipolar performance can be further improved by replacing the diffused emitter (which is formed by the n^+ source/drain implant) with a polysilicon emitter (see Section 9.4). Higher bipolar performance is obtained because shallower emitters and narrower base-widths are achieved; this reduces both transit time and parasitic capacitance by reducing emitter junction area.

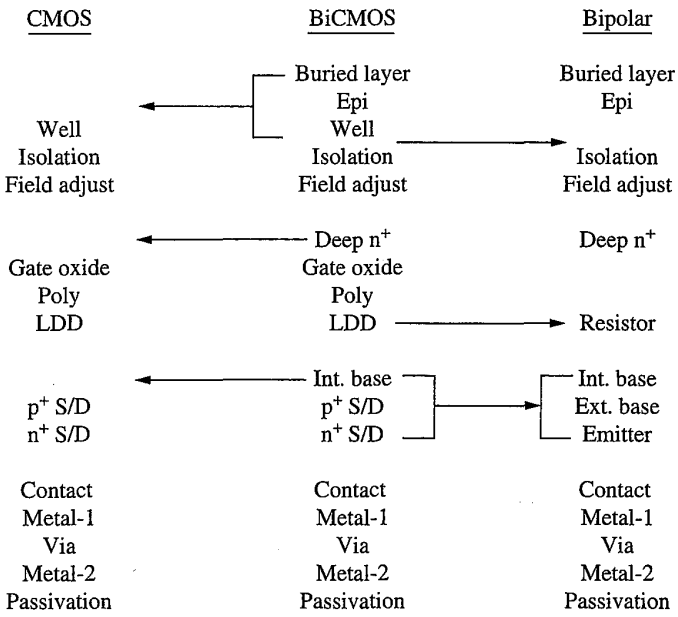


FIGURE 21
Schematic representation of CMOS, bipolar, and BiCMOS process steps. (*After Alvarez, Ref. 60.*)

Micron Ex. 1029, p. 536
Micron v. YMTC
IPR2024-00911

After poly gate patterning, the NMOS and PMOS source/drain are implanted. In this process, LDD structures are used for both NMOS and PMOS transistors. The n^- and n^+, and the p^- and p^+ S/D implants are also used for the collector and base contacts, respectively, of the npn device.

The npn intrinsic-base implant of boron is performed next, followed by the deposition of an interpoly undoped oxide layer. Next, the emitter opening is patterned in the deposited oxide layer. A second polysilicon layer is then deposited and patterned to form the poly emitter. Care should be exercised to minimize damage and contamination of the crystal surface during RIE oxide etching of the emitter windows. It is also important to minimize the thickness of the interfacial oxide, which may be grown on the silicon immediately before the poly 2 deposition; excessive thickness of the interfacial oxide results in low current gain and high emitter resistance. After the poly 2 is doped, a final high-temperature anneal is used to drive arsenic from the poly layer into the silicon to form the emitter. Since metal contacts are placed on top of the emitter to reduce resistance, an undercoating barrier material is used to prevent aluminum diffusion through the poly into the shallow emitter-base junction.

To further enhance performances, a salicide process may be implemented in which the gates, emitters, and diffusions are silicided in a self-aligned manner to reduce their sheet resistances from 20 to 50 Ω/□ to 1.5 Ω/□. The p^+ extrinsic base can also be self-aligned to the sidewall of oxide spacers next to the emitters from the outdiffusion of an extra poly layer. Local-interconnect technology may be used to connect the gates and diffusions without the need for contact holes and metal borders. As examples of these approaches, the cross sections of a high-performance twin-well submicrometer BiCMOS[49] and a trench-isolated BiCMOS[63] are shown in Figs. 22 and 23, respectively.

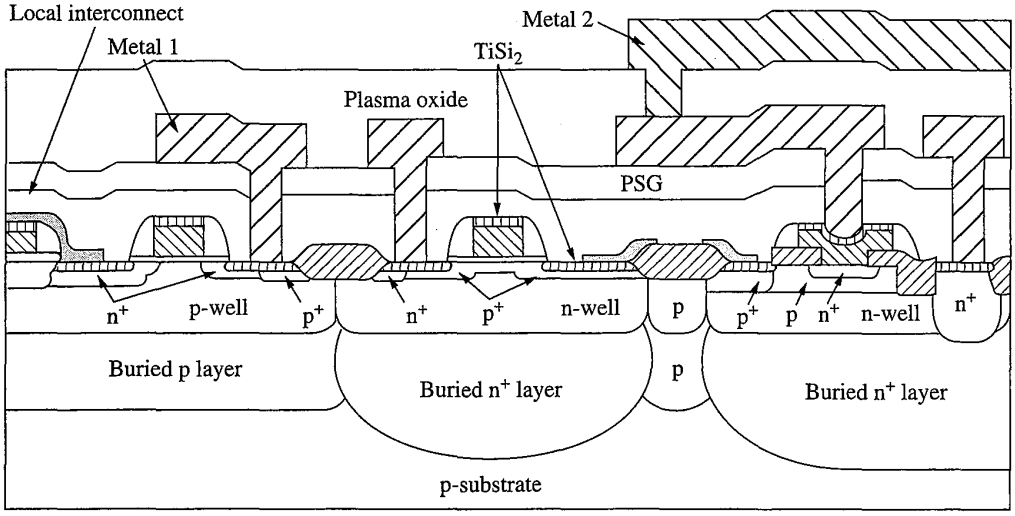


FIGURE 22
Device cross section for BiCMOS process flow following completion of double-level metal interconnect. (*After Haken et al., Ref. 62.*)

Micron Ex. 1029, p. 537
Micron v. YMTC
IPR2024-00911

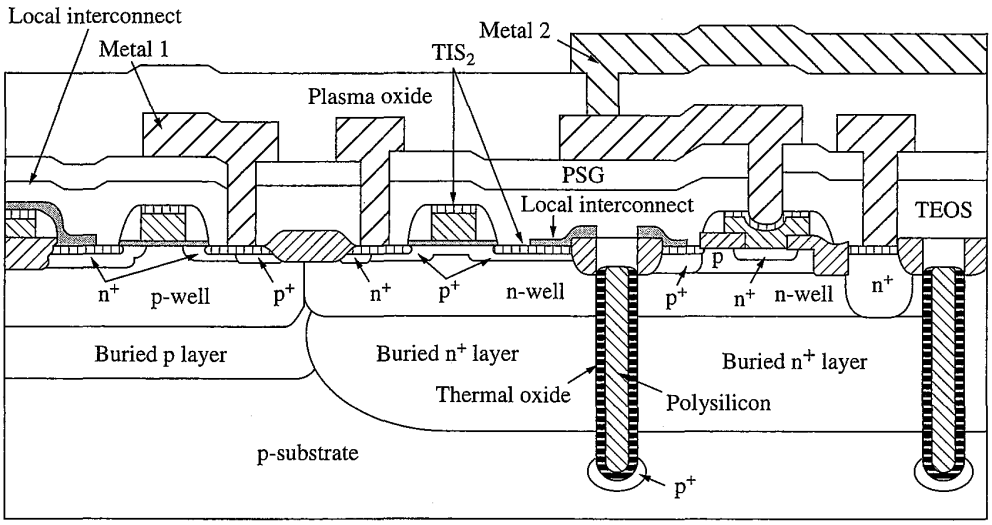


FIGURE 23
Cross section of high-performance submicrometer BiCMOS process showing both trench-isolated bipolar devices and silicided gates, emitters, and diffusions for low sheet resistance. (*After Havemann et al., Ref. 63.* ©1987 IEEE.)

9.5.4 Special Considerations

Buried layers

For BiCMOS, process requirements for the CMOS well and the bipolar collector need to be considered. One major process design trade-off involves the characteristics of the epi and well profiles. For bipolar devices, minimum epi thickness is determined by the emitter breakdown voltage (BV_{CEO}), the collector capacitance, and the manufacturing controllability. On the other hand, if the epi is too thick, f_T will be reduced and R_C will increase. From the CMOS perspective, the use of a p^+ buried layer beneath the NMOS devices dictates that the epi must be thick enough to keep the NMOS back-bias body effect[2,3] and junction capacitance from being excessive.

The n-well in the twin-well structure not only affects the PMOS devices but also serves as the collector of the bipolar devices. Thus, besides giving adequate CMOS characteristics, the n-well should be doped heavily enough to prevent the Kirk effect,[2] yet lightly enough to provide an adequately high BV_{CBO} for the bipolar devices.

Epitaxy and autodoping

The growth of a lightly doped, thin epi layer over two types of buried layers presents a challenge to the epitaxial deposition process.[64] Autodoping for both types of dopants must be minimized in both the vertical and lateral directions to avoid the need for excessive counter-doping in the wells. It has been shown that arsenic autodoping can be suppressed through the use of reduced-pressure epitaxy, but unfortunately, boron autodoping increases at lower pressures.[65] A novel two-step epi

Micron Ex. 1029, p. 538
Micron v. YMTC
IPR2024-00911

process has been reported in which a high-temperature (1150°C), low-pressure (10 torr) SiH_2Cl_2-epi cap is deposited first, followed by a low-temperature (900°C), low-pressure SiH_4 epi-layer deposition. This process produces an abrupt epi substrate transition and minimal arsenic and boron autodoping. Lateral autodoping can also be reduced if the time and temperature of the prebake period (the period between in-situ etching of the surface and the actual deposition) is increased, since the buried-layer dopants diffuse to the surface and evaporate into the gas stream during this cycle.[64]

9.6 MOS MEMORY TECHNOLOGY

9.6.1 DRAM Processing Technology

Dynamic random access memory is so named because its cells can retain information only for a limited time before they must be read and refreshed at periodic intervals. A DRAM cell consists of one transistor and one storage capacitor. For bit densities of up to one megabit, planar-type storage capacitors are used. However, as component density has increased, the amount of charges needed for a sufficient noise margin remains fixed. Therefore, in order to increase the specific capacitance, two different routes have been taken (see Fig. 24): the first solution is to store charges vertically in a trench; the second solution, which allows the cell to shrink in size without losing storage capacity, is to stack the capacitor on top of the access transistor. The majority of manufacturers have taken the stacking approach. In Fig. 24, we can see that as the memory density increases, the capacitor structure becomes more intricate and grows in the vertical direction.

Representative process flow

We now describe a typical process flow for a 16-Mb DRAM. A stacked capacitor structure is chosen because most products use it.[52] To fabricate this device, we use a modified CMOS process with a minimum geometry of 0.5 μm. A cross-sectional view of this cell is depicted in Fig. 25.

Up to the gate-oxide growth, the process uses conventional CMOS fabrication steps. Two hundred nm of poly 1 is deposited and doped with $POCl_3$, followed by a deposition of 250 nm of TEOS oxide. The poly 1 photomask is applied and the composite film is plasma-etched as a stack. Selective LDD implants are performed. Next, a layer of 150 nm of TEOS is deposited and plasma-etched, so that spacers are formed at each poly step. At this point, the gates of the peripheral devices, as well as those of the cell transfer devices, are formed and covered by a thick layer of oxide, both on top and at the sides of the poly 1 gates. The peripheral NMOS source and drain mask is then applied, and an arsenic dose of 5×10^{15} cm^{-2} is implanted.

The photomasking level of the cell contact comes next. This layer consists of oversized contact openings at the drain regions of the transfer devices. Plasma-etch is used to rid the silicon substrate of any thin oxide layer. Three hundred nanometers of poly 2 layer is then deposited, doped, and patterned. This will be the lower

Micron Ex. 1029, p. 539
Micron v. YMTC
IPR2024-00911

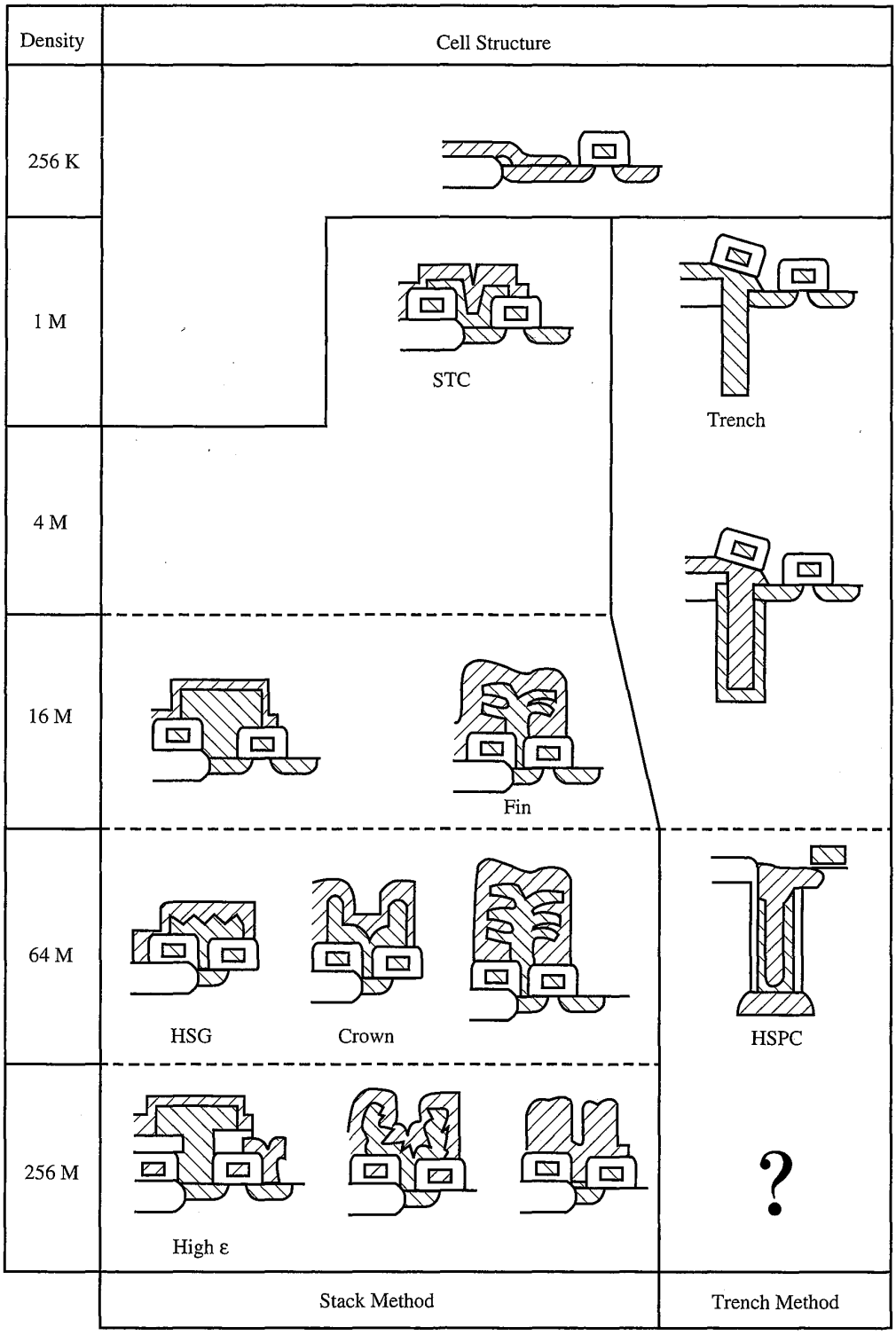


FIGURE 24
Two methods for fabricating high-density DRAM storage capacitors. STC = stack capacitor; Fin = finlike stack capacitor; HSG = hemisphere-grain stack capacitor; Crown = crownlike stack capacitor; HSPC = high-substrate plate capacitor. (*After* M. Ogirima, *1993 Symposium on VLSI Technology, Dig. of Tech. Papers.* Kyoto, Japan, 1993, p. 1. ©1993 IEEE.)

Micron Ex. 1029, p. 540
Micron v. YMTC
IPR2024-00911

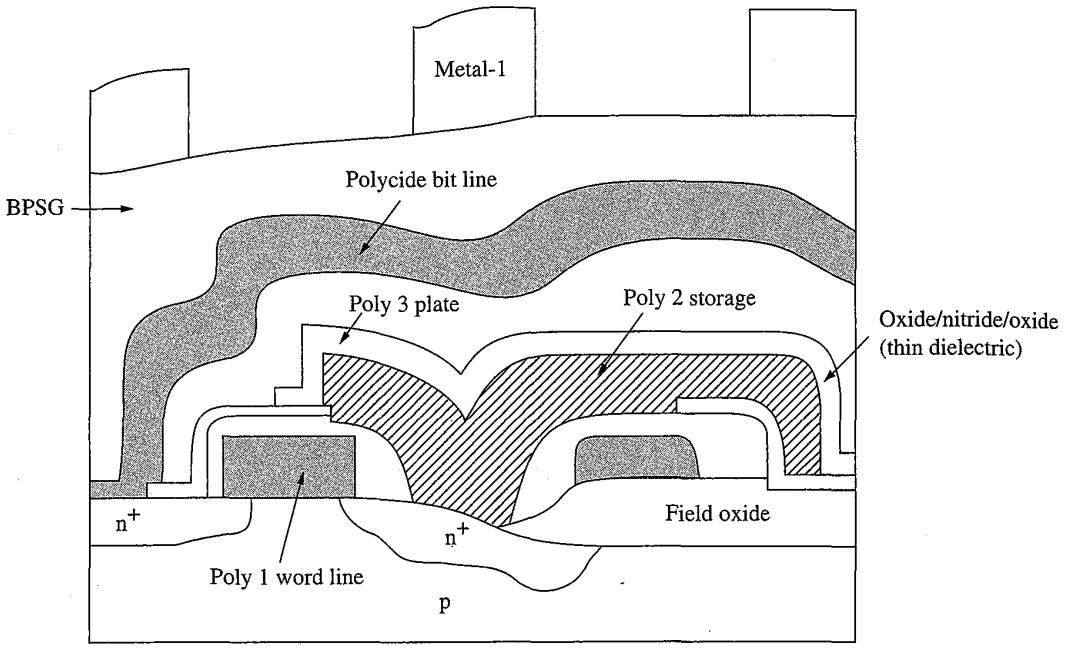


FIGURE 25
16-Mb DRAM cell cross section—stacking capacitor structure.

capacitor electrode. Note that, because the poly 1 gates are completely protected by a thick oxide, no alignment borders are needed for the cell contacts, and the contact mask could be oversized. The boundaries of the contacts are limited by the poly 1 spacers and the bird's beak of the field oxide, not by the mask dimensions. This self-aligned feature significantly reduces the cell area and increases its density.

For the capacitor dielectric, a composite film of oxide/nitride/oxide with an equivalent oxide thickness of 6 nm is deposited. The native oxide on the poly 2 is used as the first oxide layer, on top of which 5 nm of nitride is deposited. The nitride layer is then reoxidized in steam at 920°C to grow a top oxide layer 2 nm thick. The oxidation step seals any potential pinholes in the oxide/nitride/oxide composite film, making the film very reliable and high yielding. A counter electrode of 150 nm of polysilicon film is deposited next, and then it is doped and patterned. This electrode extends throughout the memory array area and has openings cut out over the source regions of each transfer device.

At this point, the p^+ masking and BF_2 implants are implemented for the peripheral devices. This step is deliberately placed after the nitride reoxidation step in order to reduce p^+ junction depth. Three hundred and fifty nanometers of BPSG is then deposited over the capacitors and reflowed at 850°C in order to smooth out the topography and reduce step height (see Fig. 25). Contact openings are then made in the source areas by using a photomasking step and etching through the BPSG layer. Seventy nanometers of poly is then deposited and doped by phosphorus implantation, and a layer of tungsten silicide is chemical-vapor deposited. Using another masking step, the composite poly/silicide (polycide) layers are patterned to define the bit lines for the cells.

Micron Ex. 1029, p. 541
Micron v. YMTC
IPR2024-00911

Next come the conventional steps of the back-end processing of metallization. The word lines in the memory arrangement are stitched with metal lines at equal intervals in order to reduce their resistances.

Special considerations for deep-trench capacitors

An alternative way to obtain larger specific capacitance is to bury the capacitors in deep trenches. The process technology making these structures possible is anisotropic etching of silicon by RIE. The requirements for etching deep, narrow trenches—3 to 4 microns for the 0.8-μm DRAM, and 5 to 7 microns for the 0.5-μm DRAM technologies—are very exacting for devices with acceptable yield:[52]

1. The etched trench should have sidewalls that are smooth and tapered with an angle of roughly 87°. Too much taper would waste cell space, but too little could cause void formation during the refilling step by CVD oxide or polysilicon.
2. The sidewalls should have no bowing or kinks, and the trench should have a smooth bottom with rounded corners, so as to prevent stress-induced defects from being generated during oxide growth and to avoid large fields at sharp corners.
3. Etch damage in the substrate must be minimized, so as not to have excessive diode leakage.
4. The trench depth must be uniform both across the wafer and from wafer to wafer.

As a common capacitor plate for all the cells, in-situ doped polysilicon is deposited to fill the trenches. Excessive material on the surface is then etched back or planarized by the chemical mechanical polishing (CMP) process.

Sometimes in deep trenches an inversion problem, due to positive charges either at the trench surface or in the trench-refill material, forms an undesirable leakage path.[66] The polysilicon that fills the trench must also be highly doped to prevent depletion effects.

9.6.2 SRAM Processing Technology

Static random access memory is called *static* because it does not require periodic refresh signals to retain stored data. The bit states in a SRAM are stored in a pair of cross-coupled inverters, which form a circuit known as a flip-flop. Figure 26*a* is a circuit schematic of a six-transistor SRAM cell. Figure 26*b* is a circuit schematic of a four-transistor cell where the active loads are replaced by passive poly resistors. As SRAM has evolved, it has increased in density. At present, with a 0.5-μm minimum line width, it has reached a size of four Mb.

BiCMOS SRAM offers densities and power dissipation levels close to those of CMOS SRAM, but with higher operating speeds. Typically, the memory array in a BiCMOS SRAM consists of an NMOS cell in a p-well. The decoders, word line, and write drivers are implemented with BiCMOS logic gates. The sense amplifiers, which require high input sensitivity, use pure bipolar circuits.

CMOS SRAM process flow

Now we examine a 1-Mb CMOS SRAM with 0.8-μm technology as an example of CMOS SRAM process flow.[67] The processing steps that precede the gate oxidation and follow the metal contact hole formation are fairly conventional and are omitted.

Micron Ex. 1029, p. 542
Micron v. YMTC
IPR2024-00911

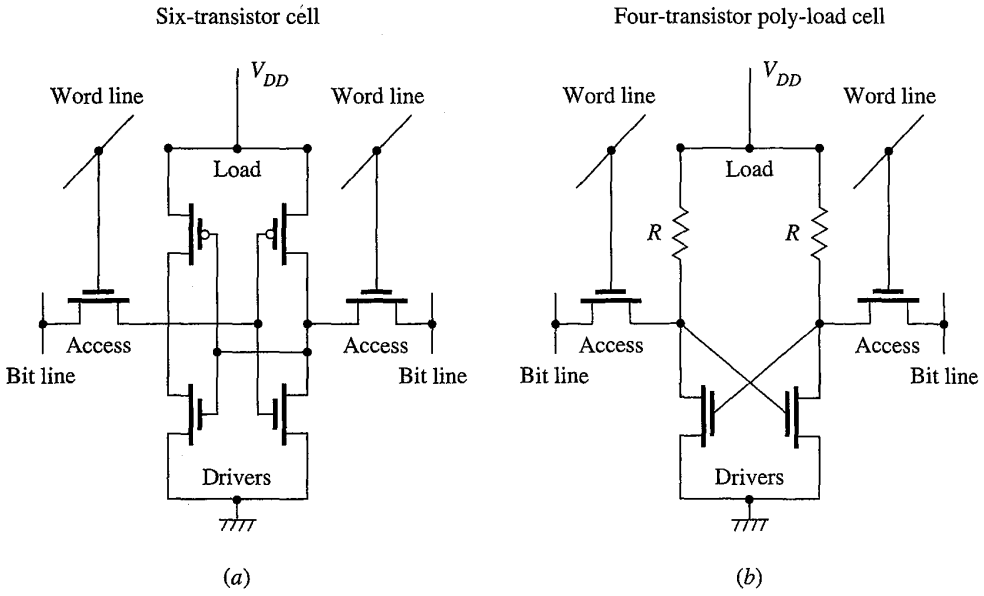


FIGURE 26
(*a*) The full-CMOS six-transistor cell has a lower power consumption and a greater immunity to transient noise and temperature fluctuations than poly load four-transistor cells. (*b*) The four-transistor cell requires less space on the chip than the six-transistor cell because it is fabricated in a separate poly layer above the transistors.

Figure 27 shows the layout of a SRAM memory cell. For maximum density in a memory device, the cell must be laid out as small a size as possible. The size is determined by the cell's topography and by the design rules of the IC fabrication technology. For this process, we use an advanced six-transistor cell. Using a PMOS thin-film transistor (TFT) on isolator load, the memory cell size is 5.2 μm × 7.9 μm. The SRAM is fabricated with a single aluminum and a triple-level polysilicon, twin-well CMOS technology. The first polysilicon level forms the gate electrodes of the cell driver NMOS transistors and the word lines. The second polysilicon level, crystallized from amorphous silicon, is used for the PMOS TFT loads and the V_{CC} lines. The third interconnect layer, tungsten silicide, is used as ground lines and upper word lines. The bit lines are formed by aluminum. A cross-sectional view of the memory cell is shown in Fig. 28. The p^+ drain of the PMOS TFT is directly connected to the n^+ polysilicon (poly 1) through a contact hole formed above the buried contact region.

Both n-channel and p-channel MOSFETs have an LDD drain structure. The gate oxide thickness of the bulk silicon CMOSFET is 20 nm. The gate oxide thickness for the TFT is 40 nm, on which 60 nm of amorphous polysilicon is deposited. The body of the TFT is annealed in an oxidizing atmosphere and crystallized to polysilicon with a grain size in the range of 100 to 500 nm. An offset at the drain of the TFT reduces its gate-drain electric field and thus reduces its off current.

Special considerations for buried contact, poly load, and thin-film transistor load

In order to establish contacts from the drain regions of the load devices to those of the n-channel devices, poly lines are used to connect the diffusion directly and

Micron Ex. 1029, p. 543
Micron v. YMTC
IPR2024-00911

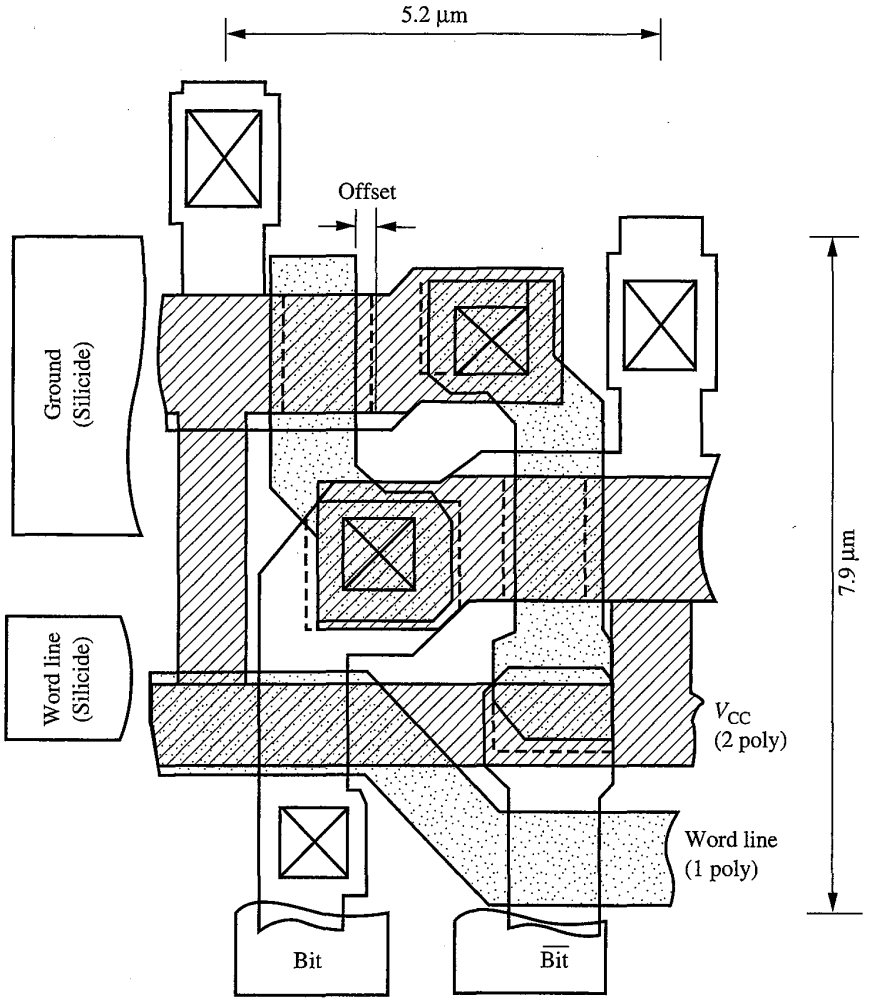


FIGURE 27
Typical SRAM memory cell layout with a PMOS TFT load. (*After Ando et al., Ref. 67.* ©1989 IEEE.)

serve as local interconnects. The poly lines are from the same layer as that of the poly gate. These contacts are buried deep under the interlayer dielectric and are thus called *buried contact.* Because it is borderless, a buried contact takes up less space and also can be used as a local interconnect between the drain of the access transistors and the gate of the driver transistors.

A representative structure of the buried-contact module is depicted in Fig. 29. Immediately after gate oxide growth, a thin layer of doped polysilicon of 50 nm is deposited. A mask layer is applied that opens the buried contact areas but protects the poly gate areas. Plasma etching is used to etch this composite layer until the substrate is exposed. At this stage, the gate oxide is protected by the thin poly so that it is sealed from contamination. Three hundred and fifty nanometers of doped polysilicon is then deposited before a second poly gate mask is applied. This mask again defines the poly gate areas and the buried contacts, but allows a small offset of the contacts to account for misalignment from the previous mask. During the second RIE etching, this offset

Micron Ex. 1029, p. 544
Micron v. YMTC
IPR2024-00911

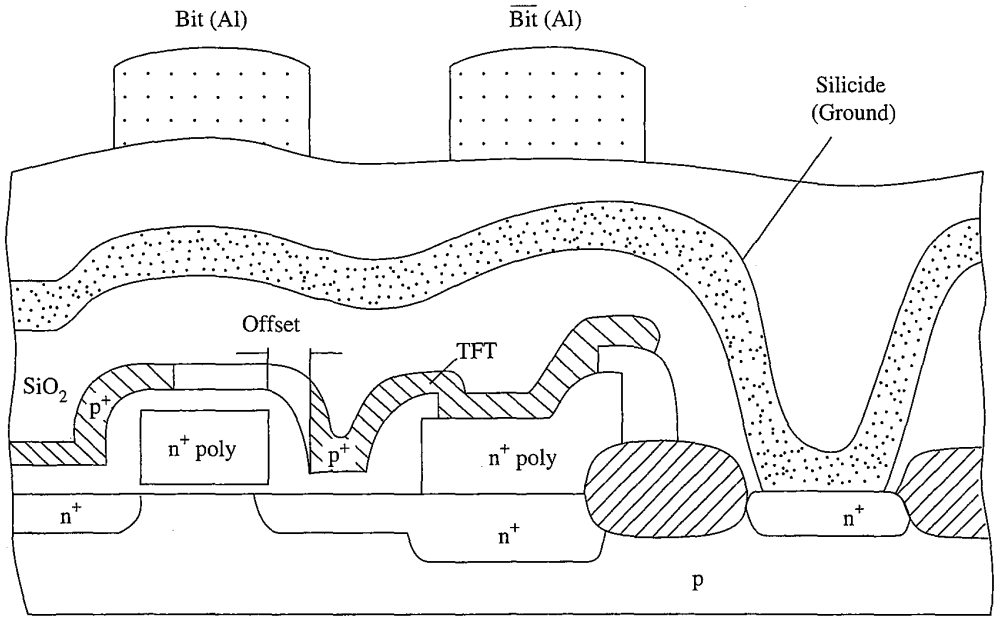


FIGURE 28
Cross section of the SRAM memory cell with a PMOS TFT load. (*After Ando et al., Ref. 67.* ©1989 IEEE.)

causes the digging of a shallow dimple in the substrate because there is no gate oxide in this area to protect it from the poly overetching. Arsenic is then implanted for the source and drain of the device. At the dimple, electrical continuity is achieved when lateral diffusion from this implant joins with the outdiffusion of impurities from the poly layer on top of the buried contacts.

The SRAM cell that uses poly film resistor as a load is easier to implement than the six-transistor cell with polysilicon thin-film transistor as a load (see Fig. 26*b*), but

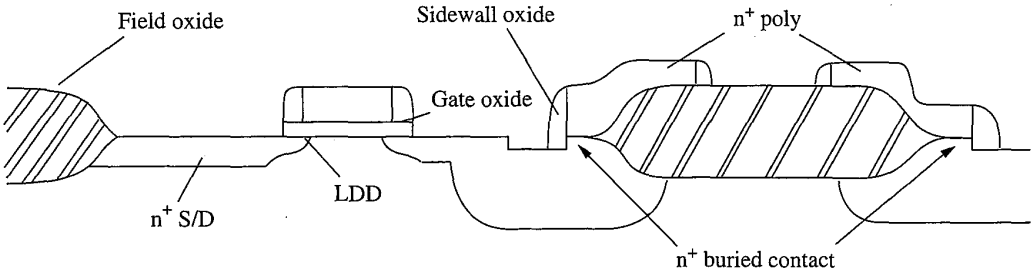


FIGURE 29
Limitations of buried contacts: (1) Phosphorus outdiffusion limits device isolation and buried-contact-to-gate design-rule scaling. (2) The use of n-type polysilicon limits buried contacts in CMOS to n-channel transistors only. (*After Tang et al., Ref. 28.* ©1987 IEEE.)

Micron Ex. 1029, p. 545
Micron v. YMTC
IPR2024-00911

it has a leakage current three orders of magnitude larger than that of a six-transistor cell. It is usually considered to be practical up to one-megabit memories.[68] Poly load resistors are fabricated from undoped polysilicon films implanted with arsenic in a dose from 1×10^{13} to 1×10^{15} cm^{-2}. The sheet resistivity can be controlled from 10^4 $\Omega/\square$ up to 10^{12} $\Omega/\square$. For a 1-Mb SRAM, a poly film with a sheet resistance of 26 $G\Omega/\square$ will be needed. This can be obtained with an arsenic implant dose of 3×10^{13} cm^{-2}.

At densities of SRAM above four megabits, polysilicon thin-film transistors must be used as loads. They are stacked above the n-channel devices (see Fig. 28). Such an arrangement has smaller cell area, smaller leakage current, and better alpha-particle soft-error immunity than a poly-resistor load.

9.6.3 Soft-Error Rate (SER) Consideration

Soft errors are single, nonrecurring read errors on a single bit of a memory array. It is not a permanent error in that the cause is not a process defect. Although soft errors can be caused by such circuit-related problems as supply-voltage noise, inadequate noise margins, and sense-amplifier imbalance, there is one specific physical failure mode that will cause soft errors. The cause of this failure mode was first identified as alpha particles originating from the decay of radioactive impurities in IC packages.[69]

The alpha particles, when incident on the memory chip, generate electron-hole pairs in the substrate and are collected by the depletion regions of the pn junction. Since a bit of information is stored on a DRAM or a SRAM cell by the presence or absence of charges in the potential well of the storage capacitor, additional carriers, because of the alpha particles collected by these junctions, can upset the original state of the junction and cause it to flip to the opposite state; thus an error results (see Fig. 30).

Soft errors can occur in both DRAM and SRAM, and the situation is aggravated as the memories are scaled down. One general way of reducing SER is to create a deep doping gradient that produces a potential-energy barrier to the diffusing electrons, preventing some of them from reaching the depleted regions of the storage nodes. The well-substrate junction acts as a reflective barrier for diffusing minority carriers created outside the well. Putting the memory on a heavily doped substrate with a lightly doped epitaxial layer on the surface is also beneficial. Since the minority carriers generated in the heavily doped substrate have a much shorter lifetime than those generated in the lightly doped regions, they are recombined before they can do harm.

9.6.4 Other Memory Technologies

Another family of semiconductor memory that uses submicrometer design rules is the nonvolatile memory, which includes electrically erasable programmable read-only memory (EEPROM), erasable programmable ROM (EPROM), and the flash EEPROM. However, because their combined production volume is eclipsed by that of DRAM and SRAM, these memories will not be discussed here. Among these ROM devices, flash EEPROM could emerge as a mainstream memory technology in the next decade.

Micron Ex. 1029, p. 546
Micron v. YMTC
IPR2024-00911

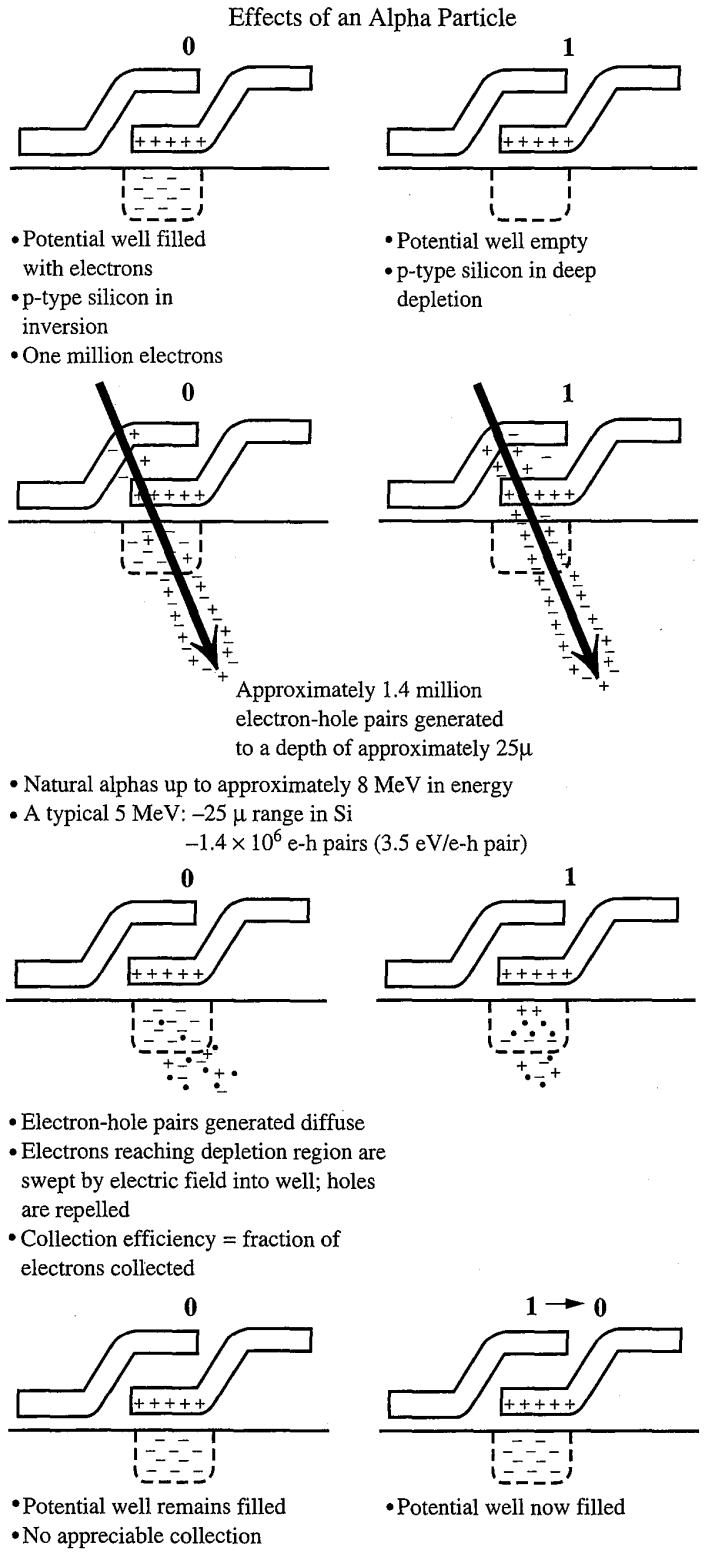


FIGURE 30
Stages of soft-error creation by alpha particles in dynamic memories. (*After May and Woods, Ref. 69.* ©1985, 1989 IEEE.)

Micron Ex. 1029, p. 547
Micron v. YMTC
IPR2024-00911

9.7 PROCESS INTEGRATION CONSIDERATIONS IN ULSI FABRICATION TECHNOLOGY

9.7.1 Introduction

With further shrinking of dimensions, the circuit processes become more stringent; old requirements are tightened, and new requirements have to be considered. In this section, we take a close look at some of the most important issues involved in the process integration of ULSI circuits.

9.7.2 Design Rules

A complete IC process usually involves ten to twenty masking steps according to the complexity of the process. Each mask introduces some changes in the surface features, whether they are in the silicon substrate or in the films deposited on it. Minimum-dimension rules, governed by processing or physical requirements, apply to these mask layers. When these minimum spacings are violated, havoc ensues in the circuit's operation. Circuit designers use the minimum design rules, or "design rules" for short, to construct the various layers, and they treat these rules as sacred. Many factors are involved in determining these rules. Some of the most important factors are described below:

Line-width tolerances

The line-width dimension on a finished wafer usually differs from the corresponding line-width dimension on the mask. This is because the photo and etching steps used to pattern this line generally shrink or expand the original line by a biased amount, Δ. Because this is a statistical process, this shrinkage or expansion introduces a variation to the uncertainty of this line width. This bias and its variation depend very much on the processing equipment and procedure.

Junction depth and lateral diffusion

After all the heating cycles are completed, each kind of junction has certain dimensions, both vertical and lateral, and each dimension has its own variation value. For instance, the n-well junction depth is about 2 μm, and the source/drain junctions are 0.1 to 0.3 μm deep. The lateral diffusion spread is sometimes estimated as 80% of the vertical junction depth.

Depletion width

When a voltage is applied to a pn junction, the depletion width change depends on the doping profile of the junction and the amount of voltage applied.[2,3]

Thickness tolerances

The thicknesses of the deposited films usually do not affect design rules; however, they will affect the rules if they give a lateral extent of a surface feature. For

Micron Ex. 1029, p. 548
Micron v. YMTC
IPR2024-00911

instance, an oxide spacer has a lateral dimension and its thickness will displace an implanted junction by that amount. The bird's beak at the edge of the field oxide is another example; the thickness of pad oxide usually affects the bird's beak length.

Masking tolerances

Masking tolerances result from two separate causes: Within the same mask, the dimensions may vary because of the mask-making process. Alignment tolerances also vary from one mask level to another. For a modern optical stepper tool, the former value is approximately 0.025 μm and the latter value is approximately 0.1 μm.

In determining the design rule of a particular feature, all the biased amounts of the factors mentioned above are added up algebraically, and their standard deviations are added in a root-mean-square manner. Consequently, the different tolerances are "nested" within each other. For example, if we want to find the minimum distance between two isolated n^+ junctions, we have the following calculations, where an uppercase letter represents the bias and a lowercase letter represents the deviation of each factor:

$$\text{Total bias} = 2D + 2B + 2A$$

$$\text{Total deviation} = 1.4 \times (d^2 + b^2 + e^2)^{1/2}$$

where D is the bird's beak, B is the lateral n^+ junction extent, and A is the process bias of the active area; and d and b are the deviations of D and B, respectively, and e is the mask run-out tolerance.

9.7.3 Compactness

IC density is of prime importance. Density can be achieved to some extent by the brute force approach, such as using better steppers or etchers. However, tighter design rules can be obtained by using more innovative process integration. One powerful method is self-alignment; the alignment tolerance between two mask levels can consume a great deal of chip area. A common example is the self-aligned source and drain implant with respect to the polysilicon gate. Another useful idea is the self-aligned contact. The object is to make a contact to the substrate between the poly gate and the field oxide or between two poly gates.[29] The main process flow, depicted in Fig. 9, is described as follows:

Usually, we need to create a clearance space between a gate and the contacts to its S/D in order to avoid a shorting problem; but the following method enables us to put a contact hole next to the edge of the gate making a connection to the n^+ junction. A polysilicon film is deposited with a thick CVD oxide on top, and then the composite film is etched at the gate stack. Another 200 nm of CVD oxide is then deposited, and the wafer is blanket-etched in plasma so that only oxide spacers are left at the edge of every poly step. At this stage, the poly gates are completely enclosed by oxide, both on the top and on the side, and the substrate areas where we want to make contacts are bare; so another poly film is deposited and patterned to form a contact landing pad. The advantage of this method is that no alignment borders are needed

Micron Ex. 1029, p. 549
Micron v. YMTC
IPR2024-00911

for these contacts. The sizes of these contacts on the mask and their alignment to the prior structure become noncritical; the actual contact dimensions on the substrate are limited by the extent of the spacers and by the bird's beak of the field oxide.

9.7.4 Step Heights

At the various processing stages, there are features on the chip that differ in height. The range of variation of this vertical profile at any mask level is termed *step height.* A major limitation in processing submicrometer feature-sized ICs is the very small depth of focus of the optical steppers used to pattern the circuit features. The focus margin (FM) or the usable depth of focus (DOF) is calculated by the following equation:

$$\text{FM} = \text{DOF} - \text{CDOF} - \text{FB} - H_s$$

where the focus budget, FB, equals 0.95 μm and is contributed by the imperfection of the optical system and mask; H_s is the step height for the specified layer; CDOF is the correction of depth of focus resulting from photoresist thickness variation; and DOF is a function of exposure geometry and is approximately 1.7 μm for 0.5-μm lines.

As an example, the focus margins for a hypothetical 16-Mb DRAM process of 0.5-μm geometry have been calculated and are shown in Table 1. Two of the layers, poly 4 and metal 1, have potential problems because their usable focus margins are negative; this is because their step height is excessive. These layers need to be planarized better so that their focus margins are within acceptable limits.

9.7.5 Process Window and Sensitivity

When a piece of equipment performs an operation, its performance is controlled by many parameters. A change in the magnitude of any of the process parameters

TABLE 1
Focus margins for DRAM process layers

Layer	H_S μm	Resolution μm	DOF μm	CDOF μm	FM μm
Active area	0	0.5	1.7	0.0	0.75
Poly 1	0.14	0.5	1.7	0.0	0.61
Cell contact	0	0.5	1.7	0.0	0.75
Poly 2	0.45	0.50	1.7	0.0	0.3
Poly 3	0.45	0.7	2.7	0.0	1.3
BL contact	0	0.5	1.7	0.0	0.75
Poly 4	1.00	0.5	1.7	0.0	−0.25
M 1 contact	0.506	0.6	2.25	0.0	0.79
Metal 1	1.3	0.6	2.25	0.91	−0.91
Via	0.5	0.6	2.25	0.0	0.8
Metal 2	0.5	0.7	2.7	0.91	0.34

Micron Ex. 1029, p. 550
Micron v. YMTC
IPR2024-00911

induces a corresponding response in its performance. The window of operation for this equipment is defined as the specification limits of the response induced by this change in parameter. For example, when exposing the photoresist of a masking layer using a stepper, the exposure dosage affects the line width of the resist. Higher dosage gives smaller line width. Therefore, in using this machine, there is a specific range of dosage used to obtain the acceptable line width. The incremental change of line width that occurs with a change of dosage is termed the *sensitivity* of the process. In designing a manufacturing process, it is desirable to have as large a process window and as small a sensitivity as possible, so that large variations in equipment performance can be tolerated. It may be helpful to plot the process variables versus the process response in a response contour surface-plot.

9.7.6 Process Parameter Monitoring

For ULSI, manufacturing technology is as important as processing technology. In fact, without a reliable way to fabricate ICs in quantity, it is doubtful that good-quality circuits can be made. An important cornerstone in IC manufacturing is statistical process control (SPC). Throughout the fabrication process, various processing parameters are collected, and their deviations from the targeted values, as well as their distribution variation over a period of time, are analyzed. Since these measurements are taken on-line, any abnormal deviation from prespecified control limits would immediately alert workers to find the underlying causes and correct the problem before a large amount of defective units are produced. Some examples of processing parameters are poly critical dimension (CD), gate oxide thickness, and poly sheet resistance. For the product, the parameters are the various wafer acceptance tests (WAT) of electrical measured data. For the equipment, the parameters are such items as the temperature control of a furnace or an implanter's dosage uniformity. For any given parameter, it is convenient to define the following quantities:[70]

$$C_p = |\text{USL} - \text{LSL}|/6s$$
$$K = (x - \text{TGT})/(|\text{USL} - \text{LSL}|)/2$$
$$C_{pk} = C_p(1 - |K|)$$

Where,

LSL (USL) = lower (upper) specification limits
TGT = target value, usually centered between LSL and USL
x = sample mean value
s = sample standard deviation

The spec of a given parameter usually has a TGT of $\pm 3s$. ULSI manufacturing demands very tight control; usually a parameter requires six-sigma values to be contained within its specification limits. Thus the quantities defined above must be limited to the following values:

$$C_p \geq 2$$

Micron Ex. 1029, p. 551
Micron v. YMTC
IPR2024-00911

$$K \leq 0.25$$

$$C_{pk} \geq 1.5$$

9.8 SUMMARY AND FUTURE TRENDS

In this chapter, we have presented the highlights and an overview of the most important developments in process integration for different technologies. A subject this broad requires volumes for a complete discussion of every detail. Although a chapter can only give a cursory treatment of the subject, we consider our task fulfilled if this whets the reader's appetite to seek other reading materials such as the references listed at the end of this chapter.

In the previous sections, we describe various current processing technologies. Silicon technology is treated exclusively because it has the widest application and is the only one adequate for ULSI. Bipolar devices have long been eclipsed by CMOS technology, although bipolar and CMOS technology have merged to form BiCMOS technology. Analog and high-voltage circuits are not covered in this chapter because they are involved at lower levels of integration. Thin-film transistor (TFT) display and silicon-on-insulator (SOI) systems are also omitted because they have not yet reached the ULSI level.

What future is in store for us? Prediction is always risky, often not because it may be too wild, but because it may be too conservative. Did Dr. William Shockley imagine his invention of the bipolar junction transistor would lead to submicrometer devices? Or did Dr. J. S. Kirby, while applying black wax for masking by hand, and Dr. Noyce, who used Al metallization for the first IC, foresee that their inventions would lead to a chip with one gigabit of memory? Nevertheless, we are safer to limit our predictions to around the year 2005, to see what new technology is waiting for us.

Semiconductor processing equipment is playing an increasingly dominant role in the advancement of IC technologies; it is the engine that drives progress. We can say with some certainty that optical lithography can reach 0.20 μm by utilizing wave-front engineering and using DUV (deep ultraviolet, wavelength less than about 200 nm) lithography. Pattern overlay accuracy with a total budget of 0.10 μm is feasible now for 0.5- to 0.35-μm technology, and may reach 0.02 μm. For patterning, highly anisotropic etching with zero etch bias is possible with the electron cyclotron resonance (ECR) type of plasma etcher, which offers low ion energies, low damage, high etch rates, and high selectivities. X-ray or electron-beam direct-write lithography may take over the job down to dimensions of 0.08 μm. The myriad problems that have plagued X-ray or electron beam direct-write lithography may be solved in time for the production of gigabit devices. Gate oxide annealed in N_2O ambient is found to yield devices with higher mobilities, better reliability and lifetime, less susceptibility to degradation, and a better barrier against impurity diffusion; thus a high-quality gate dielectric of 4 to 5 nm thickness is in sight. CVD-deposited metal film can give better step coverage and lower resistivity and is capable of filling via holes with large aspect ratios. The use of polyimide as inter metal dielectric (IMD)

Micron Ex. 1029, p. 552
Micron v. YMTC
IPR2024-00911

material coupled with pillar or reverse pillar technique can offer nearly perfect planarity and low dielectric constant to as many levels of interconnect as needed.[71]

For advanced device capabilities, it has been demonstrated by several groups that high-speed devices with 0.1 μm channel length were achieved with transconductance of 750 mS/mm for nMOSFETs, and 400 mS/mm for pMOSFETs.[72] Shallow-trench isolation coupled with retrograded wells should be capable of providing good isolation with 0.25-μm device spacings without the danger of latch-up.

If we consider the present generation to be 0.5-μm minimum geometry technology, we can project the future generation of technology to be 0.35 μm, 0.25 μm, 0.17 μm, and 0.12 μm, spaced evenly with the same ratio.[73] Assuming we maintain our present speed of progress of four years for every generation, before 2003 we could have one gigabit of DRAM memory cells per chip. Then another milestone will have been crossed; the era of giga-scaled integration (GSI) will have arrived. For product trends, fast synchronous DRAM memories with short data access time will be used as cache memory, and flash EPROM will be used as all programmable read-only memories. Highly integrated chips will be expected to incorporate as much as 10^8 transistors in the year 2000. On the basis of this capacity, a large-scale system-on-a-chip that integrates a microprocessor unit (MPU) core, an application-specific-standard-product (ASSP) core, Mb memory, and analog functions will become feasible.[74]

Because of various practical considerations, the drive for smaller minimum dimensions will probably be slowed after 0.10 μm; however, this by no means is the end of progress.[75] Better process module development can further increase the circuit density. A great deal of advancement can and will be made in semiconductor manufacturing technology.

Density and yield are the essence of semiconductor microelectronics and the key ingredients in the competitive market place. Price per circuit function is the driving force, the major components of which are performance, reliability, and manufacturing costs. Costs may only be reduced by factory automation and a reduction of packaging subassemblies through larger-scaled integration.

We predict that IC manufacturing is moving toward full-scale automation. This means complete computer control of processing equipment and data handling (such as real-time data collection and analysis, recipe download, and data upload); robotic material-handling systems to perform material transport and stocker control; and real-time lot scheduling and tracking, optimized batching, scheduling, and dispatching. Equipment tools will be mostly single-wafer processed, and cluster-oriented.

The wafer process environment will have to be much cleaner than it is now. As a rule, dust particles of one tenth the minimum geometry will be fatal defects, so we have to worry about 10-nm-sized particles. Our present practice of building cleanrooms of class one and beyond has long reached the point of diminished returns. It is common sense that contamination control is only as good as the weakest link; and most of our present sources of contamination are from the processing equipment, which so far we know little about. Many believe that, rather than making the whole building clean, it is more important to consider the immediate wafer environment. And if the wafer environment is sufficiently tight, workers should be able to work in

Micron Ex. 1029, p. 553
Micron v. YMTC
IPR2024-00911

shirtsleeves without any bad effects on the products. This would be a cleaner, much less expensive factory and more fitting for human workers.

As the IC industry reaches its maturity, we will see more consolidation of manufacturing facilities and standardization of processes, moving from a more exploratory age to a more routine one. But the electronic industry will not remain stagnant for long. As sure as the automobile followed the horse-and-buggy, somewhere a breathtaking breakthrough will be made. Some new material and device principles will be discovered, and progress will continue to new heights, which we cannot imagine now, in the twenty-first century and beyond.

REFERENCES

1. P. Chatterjee, "ULSI-Market Opportunities and Manufacturing Challenges," *IEDM Tech. Dig.,* 11 (1991).
2. S. M. Sze, *Semiconductor Devices: Physics and Technology,* Wiley, New York, 1985, *Physics of Semiconductor Devices,* 2nd ed., Wiley, New York, 1981.
3. M. Shur, *Physics of Semiconductor Devices,* Prentice Hall, Englewood Cliffs, NJ, 1990.
4. L. C. Parrillo, R. S. Payne, R. E. Davis, G. W. Reutlinger, and R. L. Field, "Twin-Tub CMOS—A Technology for VLSI Circuits," *IEDM Tech. Dig.,* 752 (1980).
5. C. Y. Lu, J. J. Sung, H. C. Kirsch, N. S. Tsai, R. Liu, A. S. Manocha, and S. J. Hillenius, "High-Performance Salicide Shallow-Junction CMOS Devices for Submicrometer VLSI Applications in Twin-Tub VI," *IEEE Trans. Electron Dev.,* **36(11)**, 2530 (1989).
6. R. D. Rung, C. J. Delloca, and L. C. Walker, "A Retrograde p-Well for Higher Density CMOS," *IEEE Trans. Electron Dev.,* **28(10)**, 1115 (1981); S. Odanaka, T. Yabu, N. Shimizu, H. Umimoto, and T. Ohzone, "A Self-Aligned Retrograde Twin-Well Structure with Buried p^+-Layer," *IEEE Electron Dev. Lett.,* **10(6)**, 280 (1989).
7. A. G. Lewis, R. A. Martin, T. Y. Huang, J. Y. Chen, and M. Koyanagi, "Latchup Performance of Retrograde and Conventional n-Well CMOS Technologies," *IEEE Trans. Electron Dev.,* **34(10)**, 2156 (1987).
8. E. Kooi, *The Invention of LOCOS,* IEEE, Inc., New York, 1991.
9. E. Kooi, J. G. van Lierop, and J. A. Appels, "Formation of Silicon Nitride at a Si/SiO_2 Interface Using Local Oxidation of Silicon and during Heat Treatment of Oxidized Silicon in NH_3 gas," *J. Electrochem. Soc.,* **123(7)**, 1117 (1976).
10. T. T. Sheng, C. Y. Lu, R. D. Chang, and S. T. Chiang, "From White Ribbon to Black Belt: A Direct Observation of the Kooi Effect Masking Film by TEM," *J. Electrochem. Soc.,* **140(11)**, L163 (1993).
11. R. H. Havemann and G. P. Pollack, U.S. Patent 4,541,167 (1985).
12. J. M. Sung, C. Y. Lu, L. B. Fritzinger, T. T. Sheng, and K. H. Lee, "Reverse L-Shape Sealed Poly-Buffer LOCOS Technology," *IEEE Electron Dev. Lett.,* **11(11)**, 549 (1990).
13. H. Mikoshiba, T. Homma, and K. Hamano, "A New Trench Isolation Technology as a Replacement of LOCOS," *IEDM Tech. Dig.,* 578 (1984).
14. S. H. Goodwin and J. D. Plummer, "Electrical Performance and Physics of Isolation Region Structures for VLSI," *IEEE Trans. Electron Dev.,* **31(7)**, 861 (1984).
15. S. Ogura, P. J. Tsang, W. W. Walker, D. L. Critchlow, and J. F. Shepard, "Design and Characteristics of the Lightly Doped Drain-Source (LDD) Insulated Gate Field-Effect Transistor," *IEEE Trans. Electron Dev.,* **27(8)**, 1359 (1980).

Micron Ex. 1029, p. 554
Micron v. YMTC
IPR2024-00911

16. T. Hori, Y. Odake, and T. Yasui, "Deep-Submicrometer Large-Angle-Tilt Implanted Drain (LATID) Technology," *IEEE Trans. Electron Dev.,* **39(10)**, 2312 (1992).
17. Y. Okumura, M. Shirahata, A. Hachisuka, T. Okudaira, H. Arima, and T. Matsukawa, "Source-to-Drain Nonuniformly Doped Channel (NUDC) MOSFET Structures for High Current Drivability and Threshold Voltage Controllability," *IEEE Trans. Electron Dev.,* **39(11)**, 2541 (1992).
18. C. Codella and S. Ogura, "Halo Doping Effect in Sub-Micron DI-LDD Device Design," *IEDM Tech. Dig.,* 230 (1985).
19. J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, "Generalized Guide to MOSFET Miniaturization," *IEEE Electron Dev. Lett.,* **1(1)**, 2 (1980).
20. A. Tanaka, T. Yamaji, A. Uchiyama, T. Hayashi, T. Iwabuchi, and S. Nishikawa, "Optimization of Amorphous Layer Thickness and Junction Depth on the Preamorphization Method for Forming Shallow-Junction in Silicon," *IEDM Tech. Dig.,* 785 (1989); S. N. Hong, G. A. Ruggles, J. J. Wortman, and M. C. Ozturk, "Material and Electrical Properties of Ultra-Shallow p^+-n Junctions Formed by Low-Energy Ion Implantation and Rapid Thermal Annealing," *IEEE Trans. Electron Dev.,* **38(3)**, 479 (1991).
21. M. C. Ozturk, J. J. Wortman, C. M. Osburn, A. Ajmera, G. A. Rozgonyi, E. Frey, W. K. Chu, and C. Lee, "Optimization of the Ge Preamorphization Conditions for Shallow-Junction Formation," *IEEE Trans. Electron Dev.,* **35(5)**, 659 (1988).
22. C. Carter, W. Maszara, D. K. Sadana, G. A. Rozgonyi, J. Liu, and J. Wortman, "Residual Defects Following Rapid Thermal Annealing of Shallow Boron and BF_2 Implants into Preamorphized Silicon," *Appl. Phys. Lett.,* **44(4)**, 459 (1984).
23. C. K. Lau, Y. C. See, D. B. Scott, J. M. Bridges, S. M. Parna, and R. D. Daires, "Titanium Disilicide Self-Aligned Source/Drain Plus Gate Technology," *IEDM Tech. Dig.,* 714 (1982).
24. C. Y. Lu, J. J. Sung, R. Liu, N. S. Tsai, R. Singh, S. J. Hillenius, and H. C. Kirsch, "Process Limitation and Device Design Trade-Offs of Self-Aligned $TiSi_2$ Junction Formation in Submicrometer CMOS Devices," *IEEE Trans. Electron Dev.,* **38(2)**, 246 (1991).
25. R. Liu, F. A. Baiocchi, L. A. Heimbrook, J. Kovalchick, D. L. Malm, D. S. Williams, and W. T. Lynch, "Formation of Shallow p^+/n and n^+/p Junction with $CoSi_2$," *Electrochem. Soc. Proceedings of the 1st International Symposium on ULSI Sciences and Technology,* **87–11** 1987, p. 446.
26. E. Nagasawa, H. Okabayahi, and M. Morimoto, "Mo and Ti-Silicided Low-Resistance Shallow Junction Formed Using the Ion Implantation through Metal Technique," *IEEE Trans. Electron Dev.,* **34(3)**, 581 (1987).
27. D. L. Kwong, Y. H. Ku, S. K. Lee, E. Louis, N. S. Alvi, and P. Chu, "Silicided Shallow Junction Formation by Ion Implantation of Impurity Ions into Silicide Layers and Subsequent Drive-In," *J. Appl. Phys.,* **61(11)**, 5084 (1987).
28. T. E. Tang, C. C. Wei, R. A. Haken, T. C. Holloway, L. R. Hite, and T. G. W. Blake, "Titanium Nitride Local Interconnect Technology for VLSI," *IEEE Trans. Electron Dev.,* **34(3)**, 628 (1987).
29. C. Y. Lu, D. S. Yaney, K. H. Lee, M. S. Twiford, N. S. Tsai, T. Kook, L. B. Fritzinger, M. L. Chen, and T. S. Yang, "A Folded Extended Window MOSFET for ULSI Applications," *IEEE Electron Dev. Lett.,* **9(8)**, 388 (1988).
30. W. T. Lynch, "Self-Aligned Contact Schemes for Source-Drains in Submicron Devices," *IEDM Tech. Dig.,* 354 (1987).
31. F. M. Wanlass and C. T. Sah, "Nanowatt Logic Using Field-Effect MOS Triodes," *ISSCC Tech. Dig.,* 32 (1963).
32. J. Y. Chen, *CMOS Devices and Technology for VLSI,* Prentice Hall, Englewood Cliffs, NJ, 1989.
33. S. J. Hillenius and L. C. Parrillo, U.S. Patent 4,554,726 (1985).

Micron Ex. 1029, p. 555
Micron v. YMTC
IPR2024-00911

34. C. Y. Lu, J. M. Sung, H. C. Kirsch, S. J. Hillenius, T. E. Smith, and L. Manchanda, "Anomalous C-V Characteristics of Implanted Poly MOS Structures in n^+/p^+ Dual-Gate CMOS Technology," *IEEE Electron Dev. Lett.*, **10(5)**, 192 (1989).
35. R. R. Troutman, *Latch-up in CMOS Technology: The Problem and Its Cure,* Kluwer Academic Publishers, Norwell, MA, 1986; D. B. Estreich and R. W. Dutton, "Modeling Latch-Up in CMOS Integrated Circuits," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Syst.*, **1(4)**, 157 (1982).
36. S. J. Hillenius, and W. T. Lynch, "Gate Material Work Function Considerations for 0.5 μm CMOS," *Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers, Port Chester, NY,* 1985, p. 147.
37. S. J. Hillenius, R. Liu, G. E. Georgiou, R. L. Field, D. S. Williams, A. Kornblit, D. M. Boulin, R. L. Johnston, and W. T. Lynch, "A Symmetric Submicron CMOS Technology," *IEDM Tech. Dig.*, 252 (1986).
38. M. Ghezzo and D. M. Brown, "Diffusivity Summary of B, Ga, P, As, and Sb in SiO_2," *J. Electrochem. Soc.*, **120(1)**, 146 (1973).
39. L. Manchanda, "Hot-Electron Trapping and Generic Reliability of p^+-Polysilicon/SiO_2/Si Structures for Fine-Line CMOS Technology," *Proceedings of the International Reliability Physics Symposium,* 1986, p. 183.
40. J. J. Sung and C. Y. Lu, "A Comprehensive Study on p^+ Polysilicon-Gate MOSFET's Instability with Fluorine Incorporation," *IEEE Trans. Electron Dev.*, **37(11)**, 2312 (1990).
41. J. R. Pfiester, F. K. Baker, T. C. Mele, H. H. Tseng, P. J. Tobin, J. D. Hayden, J. W. Miller, C. D. Gunderson, and L. C. Parrillo, "The Effects of Boron Penetration on p^+ Polysilicon Gated PMOS Devices," *IEEE Trans. Electron Dev.*, **37(8)**, 1842 (1990).
42. C. Y. Lu and J. M. Sung, "Negative Charge Induced Degradation of pMOSFETs with BF_2 Implanted p^+-Poly Gate," *Electron. Lett.*, **25,** 1685 (1989).
43. L. C. Parrillo, S. J. Hillenius, R. L. Field, E. L. Hu, W. Fichtner, and M. L. Chen, "A Fine-Line CMOS Technology That Used p^+ Poly/Silicide Gates for NMOS and PMOS Devices," *IEDM Tech. Dig.*, 418 (1984).
44. W. Lin, M. L. Chen, R. H. Doklan, and C. Y. Lu, "Dopant Diffusion in n^+/p^+ Poly Gate CMOS Process," *Solid-State Electronics,* **32(11)**, 965 (1989).
45. G. Wilson, "Advances in Bipolar VLSI," *Proc. IEEE,* **78(11)**, 1707 (1990).
46. W. C. Till and J. T. Luxton, *Integrated Circuits: Materials, Devices, and Fabrications,* Prentice Hall, Englewood Cliffs, NJ, 1982.
47. B. T. Murphy, V. J. Glinski, P. A. Gray, and R. A. Pederson, "Collector Diffusion Isolated Integrated Circuits," *Proc. IEEE,* **57(9)**, 1523 (1969).
48. J. Buie, "Improved Triple Diffusion Means Densest ICs Yet," *Electronics,* 101 (Aug., 1975).
49. A. Wieder, "Submicron Bipolar Technology—New Chances for High Speed Applications," *IEDM Tech. Dig.*, 8 (1986).
50. D. D. Tang, T. H. Ning, R. D. Isaac, G. C. Feth, S. K. Wiedmann, and H. N. Yu, "Subnanosecond Self-Aligned I^2L/MTL Circuits," *IEEE J. Solid-State Circuits* **15(4)**, 444 (1980).
51. G. P. Li, T. H. Ning, C. T. Chuang, M. B. Ketchen, D. D. Tang, and J. Mauer, "An Advanced High-Performance Trench-Isolated Self-Aligned Bipolar Technology," *IEEE Trans. Electron Dev.*, **34(11)**, 2246 (1987).
52. S. Wolf, *Silicon Processing for the VLSI Era, Vol. 2—Process Integration,* Lattice Press, Sunset Beach, CA, 1990, p. 510.
53. T. H. Ning and R. D. Isaac, "Effect of Emitter Contact on Current Gain of Silicon Bipolar Devices," *IEEE Trans. Electron Dev.*, **27(11)**, 2051 (1980).
54. S. M. Sze, *Semiconductor Devices: Physics and Technology,* chapter 4, "Bipolar Devices," Wiley, New York, 1985, p. 109.

Micron Ex. 1029, p. 556
Micron v. YMTC
IPR2024-00911

55. T. C. Chen, K. Y. Toh, J. D. Cressler, J. Warnock, P. F. Lu, D. D. Tang, G. P. Li, C. T. Chuang, and T. H. Ning, "A Submicrometer High-Performance Bipolar Technology," *IEEE Electron Dev. Lett.,* **10(8)**, 364 (1989).
56. H. C. De Graaff and J. G. De Groot, "The SIS Tunnel Emitter: A Theory for Emitters with Thin Interface Layers," *IEEE Trans. Electron Dev.,* **26(11)**, 1771 (1979).
57. P. Ashburn and B. Soerowirdjo, "Comparison of Experimental and Theoretical Results on Polysilicon Emitter Bipolar Transistors," *IEEE Trans. Electron Dev.,* **31(7)**, 853 (1984).
58. C. C. Ng and E. S. Yang, "A Thermionic-Diffusion Model of Polysilicon Emitter," *IEDM Tech. Dig.,* 32 (1986).
59. T. Yuzubira, T. Yamaguchi, and J. Lee, "Submicron Bipolar-CMOS Technology Using 16 GHz f_T Double Poly-Si Bipolar Devices," *IEDM Tech. Dig.,* 748 (1988).
60. A. R. Alvarez, "Introduction to BiCMOS," in A. R. Alvarez, Ed., *BiCMOS Technology and Applications,* Kluwer Academic Publishers, Narwell, MA, 1989, p. 1.
61. K. Miyata, "BiCMOS Technology Overview," *IEDM Short Course on BiCMOS Technology,* IEEE Inc., New York, 1987, p. 1.
62. R. A. Haken, R. H. Havemann, R. H. Eklund, and L. N. Hutter, "BiCMOS Process Technology," in A. R. Alvarez, Ed., *BiCMOS Technology and Applications,* Kluwer Academic Publishers, Narwell, MA, 1989.
63. R. Havemann, R. Eklund, R. Haken, D. Scott, H. Tran, P. Fung, T. Ham, D. Favreau, and R. Virkus, "An 0.8 μm 256K BiCMOS SRAM Technology," *IEDM Tech. Dig.,* 841 (1987).
64. J. Borland, M. Gangani, R. Wise, S. Fong, Y. Oka, and Y. Matsumoto, "Silicon Epitaxial Growth for Advanced Device Structures," *Sol. State Technol.,* 111 (Jan., 1988).
65. M. Graef and B. Leunissen, "Antimony, Arsenic, Phosphorus, and Boron Autodoping in Silicon Epitaxy," *J. Electrochem. Soc.,* **132(8)**, 1942 (1985).
66. W. P. Noble, A. Bryant, and S. H. Voldman, "Parasitic Leakage in DRAM Trench Storage Capacitor Vertical Gated Diodes," *IEDM Tech. Dig.,* 340 (1987).
67. M. Ando, T. Okazawa, H. Furuta, M. Ohkawa, J. Monden, N. Kodama, K. Abe, H. Ishihara, and I. Sasaki, "A 0.1-μA Standby Current, Ground-Bounce-Immune 1-Mbit CMOS SRAM," *IEEE J. Solid State Circuits,* **24(6)**, 1708 (1989).
68. N. C. C. Lu, L. Gerzberg, C. Y. Lu, and J. D. Meindl, "Modeling and Optimization of Polycrystalline-Silicon Resistors," *IEEE Trans. Electron Dev.,* **28(7)**, 818 (1981); N. C. C. Lu, L. Gerzberg, and J. D. Meindl, "Scaling Limitations of Monolithic Polycrystalline-Silicon Resistors in VLSI Static RAMs and Logic," *IEEE Trans. Electron Dev.,* **29(4)**, 682 (1982).
69. T. C. May and M. H. Woods, "Alpha-Particle-Induced Soft Errors in Dynamic Memories," *IEEE Trans. Electron Dev.,* **26(1)**, 2 (1989); B. Chappell, S. E. Schuster, and G. A. Sai-Halasz, "Stability and SER Analysis of Static RAM Cells," *IEEE Trans. Electron Dev.,* **32(2)**, 463 (1985).
70. P. E. Fieler, "Understanding Motorola's Six Sigma Program," *International Reliability Physics Symposium Tutorials, 2.1,* IEEE Inc., New York, 1990.
71. J. L. Yeh, G. W. Hills, and W. T. Cochran, "Reverse Pillar and Maskless Contact—Two Novel Recessed Metal Schemes and Their Comparisons to Conventional VLSI Metallization Schemes," *Proceedings of the VLSI-Multilevel Interconnection Conference, Sunnyvale, CA,* 1988, p. 95.
72. Y. Taur and Y. Mii, "0.1 μm CMOS and Beyond," *Proceedings of the 1993 International Symposium on VLSI Technology, Systems, and Applications,* Taipei, Taiwan, 1993, p. 1.
73. C. Hu, "Future CMOS Scaling and Reliability," *Proc. IEEE,* **81**, 682 (1993).
74. H. Komiya, "Future Technological and Economic Prospects for VLSI," *ISSCC Tech. Dig.,* 16 (1993).

Micron Ex. 1029, p. 557
Micron v. YMTC
IPR2024-00911

75. J. D. Meindl, V. K. De, and B. Agrawal, "Prospects for Gigascale Integration (GSI) Beyond 2003," *ISSCC Tech. Dig.*, 12 (1993).

PROBLEMS

1. The white ribbon, or the Kooi effect, in the LOCOS field oxidation process can cause thin spots in the gate oxide, resulting in low gate oxide yield. State how this effect can be remedied.
2. Compare the relative merits of the n-well, p-well, and the twin-well processes for the CMOS circuits.
3. Calculate the capacitance for the capacitor of a DRAM that has a total electrode surface area of 3.14 μm^2 and uses oxide-nitride-oxide as a dielectric, with consecutive layers 1 nm, 6 nm, and 1.2 nm thick.
4. The critical dimensions (CDs) of the poly gate were measured for five lots of a CMOS process:

Lot number	Poly gate CD (μm)
L233100	0.781
L233101	0.752
L233102	0.748
L233103	0.777
L233104	0.790

   The gate dimension on the mask is 0.70 μm. What are the average value and standard deviation for the poly gate dimensions?

5. From experiments, we know the following values:

Items	Bias	Deviation
Mask registration	0.00	0.028
Mask deviation	0.00	0.025
Misalignment between two mask layers	0.00	0.100
Contact layer process variations	0.04	0.100

   The processing spec. for the poly gate CD is 0.750 ± 0.100 μm. What is the minimium design rule of the poly gate in problem 4 for the contact spacing in a MOS transistor?

6. What are the C_p and C_{pk} values from the latest factory lots in problems 4 and 5?
7. Formulate a detailed process flow of the fabrication of a submicrometer n-channel MOS capacitor in order to monitor the capacitor's threshold voltage.
8. What are the essential requirements for device isolation?
9. Why is the "hot-electron" problem more severe for submicrometer devices, and what are the measures we use to solve this problem in processing?
10. Explain how alpha particles increase the soft-error rate (SER) for a memory array, and describe the process procedures for reducing SER.

Micron Ex. 1029, p. 558
Micron v. YMTC
IPR2024-00911

CHAPTER 10

Assembly and Packaging

T. Tachikawa

10.1 INTRODUCTION

The packaging of the ULSI chip (also called a die or bar) is a broad subject that ranges from preassembly wafer preparation to fabrication technologies for the packages. The purposes of the packaging are to provide electrical connection, expand the chip electrode pitch for the next level of packaging, protect the chip from mechanical and environmental stress, and provide a proper thermal path for the heat that the chip dissipates.

Packaging significantly affects and, in many cases, dominates the overall cost, performance, and reliability of the packaged chip and the system in which the package is applied. Enhanced performance and downsizing of electronic appliances result mainly from miniaturization of the geometry and higher integration of the chip itself, but they also have been aided by package improvements. Leading-edge ULSI devices require superior package performance. Memories have been the driving force in the advancement of leading-edge wafer fabrication technology; they lead to reduced package size and the resulting increased density. Logic and microprocessor devices, with high input/output (I/O) terminal numbers, higher speed, and high power dissipation, create additional challenges for packaging technology. Packaging designers need to achieve, simultaneously, larger I/O pin counts, a better thermal path, and excellent electrical performance in a smaller package.

Progress in the wafer fabrication process also leads to challenges in assembly techniques. For example, the introduction of new metallization material to reduce stress-induced migration requires a corresponding breakthrough in interconnection techniques. The ULSI package design must meet various demands from packaged chip testing, handling systems, and next-level packaging. Generally, a high-density ULSI package that contains a comparatively larger chip requires a smaller external terminal spacing. This jeopardizes the post-packaging processes described above.

Micron Ex. 1029, p. 559
Micron v. YMTC
IPR2024-00911

Mechanical, thermal, and electrical integrities in a single-chip package do not always guarantee integrities in next-level packaging. They sometimes impose burdens, such as printed wiring board (PWB) design restrictions, mounting and soldering difficulties, and solder joint reliability problems.

Packaging cost is important because it accounts for a considerable portion of the total cost of the packaged ULSI chip. The cost, typically 1 cent/pin in small-scale integration (SSI) and in medium-scale integration (MSI) packaging, has been increasing, often reaching as much as 10 cent/pin. Packaging cost is likely to be a more dominant cost than the cost of the chip itself. ULSI packaging is the technology that must bridge the gap between the chip and the next-level packaging at a reasonable cost and achieve the required performances through a variety of engineering trade-offs.

This chapter describes all of these packaging issues. It covers single-chip packages, assembly techniques, and mechanical, electrical, and thermal design considerations, paying special attention to relations with the next-level packaging, particularly at the PWB level.

10.2 PACKAGE TYPES

A wide variety of package types can be used for ULSI devices. However, they can be divided into two basic types: hermetic-ceramic and plastic packages. In a hermetic package, the chip resides in an environment decoupled from the external environment by a vacuum-tight enclosure. The package is usually ceramic-based and designed for high-performance applications that allow some cost penalties. In a plastic package, on the other hand, the chip is not perfectly decoupled from the external environment because it is encapsulated with resin materials, typically epoxy-based resin. The outside ambient affects the chip over a long time and gradually penetrates the plastic. Plastic packages have become more popular as their applications expand and they perform better. Because the production process is typically automated batch-handling, they are also cost competitive. This popularity persists because of improved plastics and other process developments. Design improvements that build performance features into the package also drive the trend.

Figure 1[1] illustrates a typical hermetic package. The chip resides in a cavity of the package. Package base material is a formed ceramic on which metallization wirings and external leads are placed. The chip and the package are interconnected by fine Al wire. Hermetic sealing is complete with the cap, usually ceramic or metal, lidded to the package. The package excludes environmental contaminants and has little mechanical or chemical effect on the chip, since the package components do not touch the chip surface. Al_2O_3 is the usual ceramic material; however, AlN would be a good choice where higher power dissipation is required.

Figure 2[2] shows a typical plastic package. The chip is attached to the paddle of the lead frame. The frame, made of etched or stamped thin metal (usually Fe-Ni or Cu alloys), serves as a skeleton around which the package is assembled, and it provides external leads in the completed package. Interconnections are fine gold

Micron Ex. 1029, p. 560
Micron v. YMTC
IPR2024-00911

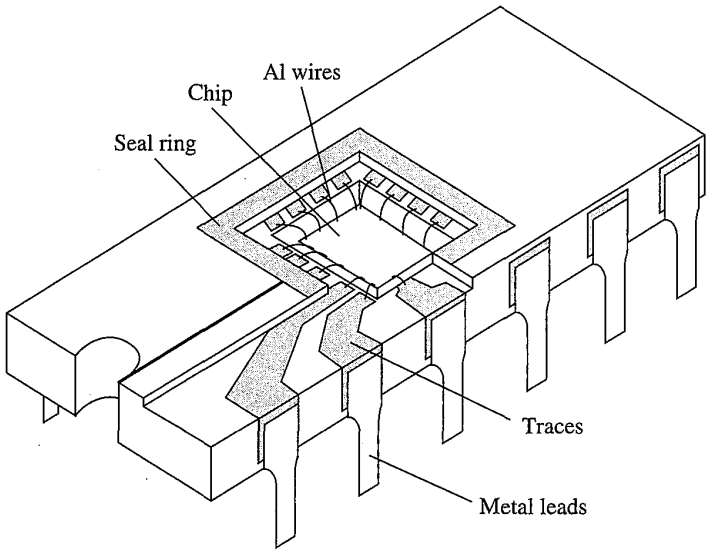


FIGURE 1
A typical hermetic package. A silicon chip is placed in the cavity of a ceramic-based package and wedge-bonded to make electrical connections to the terminals on the package. Metallized traces, usually of W, form electrical paths on the package. Lid sealing makes the cavity hermetic. (*After Prokop and Williams, Ref. 1.*)

wire. Encapsulation is carried out by the transfer molding method using epoxy resin. The resin covers the chip and forms the package's outer shape at the same time. External leads are formed into the final shape after molding. The package cannot thoroughly decouple the chip from the environment, so the mechanical and chemical effects from the plastic touching the chip require attention.

10.2.1 Through-Hole and Surface-Mount Packages

Next-level packaging, particularly at the PWB level, classifies single-chip packages into through-hole (TH) and surface-mount (SM) types. The TH types include the dual-in-line package (DIP) and the pin-grid-array (PGA) package. Both are available in hermetic-ceramic and plastic types. TH packages provide easier placement for soldering and stronger solder joints at the PWB level. However, they sacrifice PWB design flexibility, restrict mounting density because of the fine-pitch drilling process, and add drilling cost.

The prevalence of SM packages is remarkable. Over 50% of the packages used today are SM types, and this type has been accepted for ULSI packaging. Although SM packages include both hermetic-ceramic and plastic types, the lower-cost plastic package is preferred because package cost increases approximately in proportion to the increase in pin count. SM packages enhance PWB design flexibility because through-holes are unnecessary. They allow finer-pitch soldering than TH packages

Micron Ex. 1029, p. 561
Micron v. YMTC
IPR2024-00911

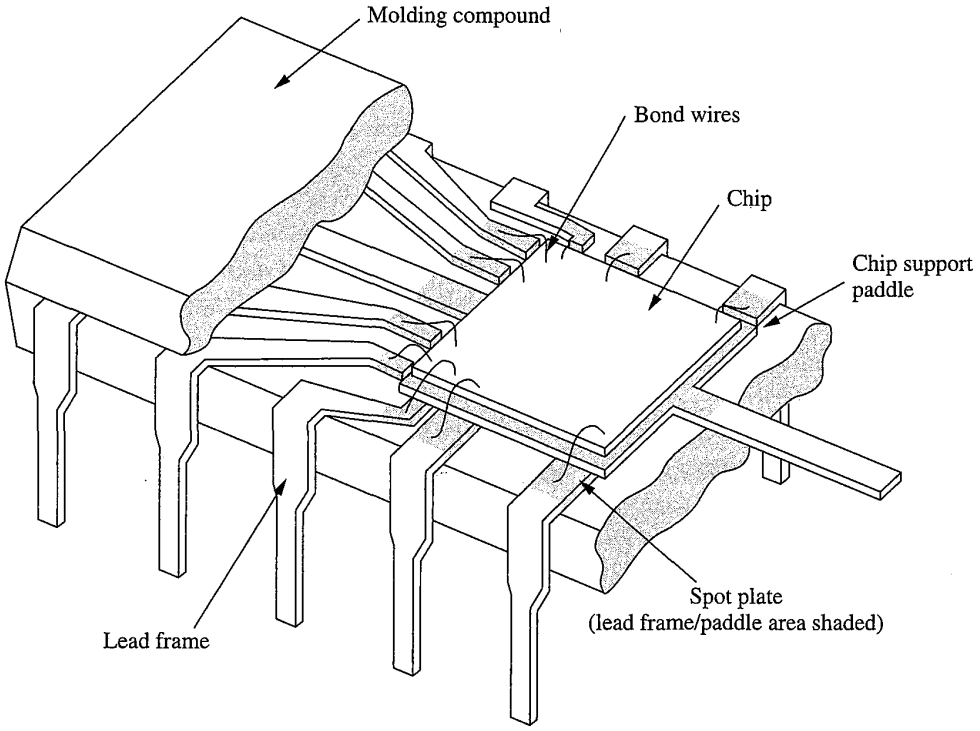


FIGURE 2
A typical plastic package. The package is a composite structure consisting of a silicon chip, a metal lead frame, and a plastic molding compound. The chip support paddle and inner tips of the lead frame are Ag-plated. External leads are solder-plated after molding. (*After Howell, Ref. 2.*)

do, because the pitch is determined by PWB (Cu foil) etching process capability, not by the drilling requirements. Generally, SM packages provide better area efficiency. Finer-pitch SM packages, especially higher-pin-count packages for ULSI devices, challenge both single-chip packaging and surface-mount techniques. The key issues are maintaining lead integrity (such as coplanarity and skew), uniform solder printing, precise mounting, and controlled reflow soldering. SM package leads come in two different geometric forms—*J-leads,* such as those shown in Fig. 6, and *gull-wing leads,* such as those shown in Fig. 25. The small-outline J-lead (SOJ) package and the quad flat J-lead (QFJ) package are examples of the former, and the small-outline package (SOP) and the quad flat package (QFP) are examples of the latter.

In memory ULSI applications, where comparatively smaller pin counts are required, the J-lead package is a frequent choice. The soldering portions of J-leads are located beneath the plastic body, so they do not need extra space for soldering as gull-wing leads do, thus decreasing the overall area the package occupies. However, in logic and microprocessor ULSI applications, where higher pin counts are needed, gull-wing leads predominate. Gull-wing leads are easier to form than J-leads, and the better geometric integrity of the leads that results is particularly suitable for higher-pin-count and finer-pitch areas.

Micron Ex. 1029, p. 562
Micron v. YMTC
IPR2024-00911

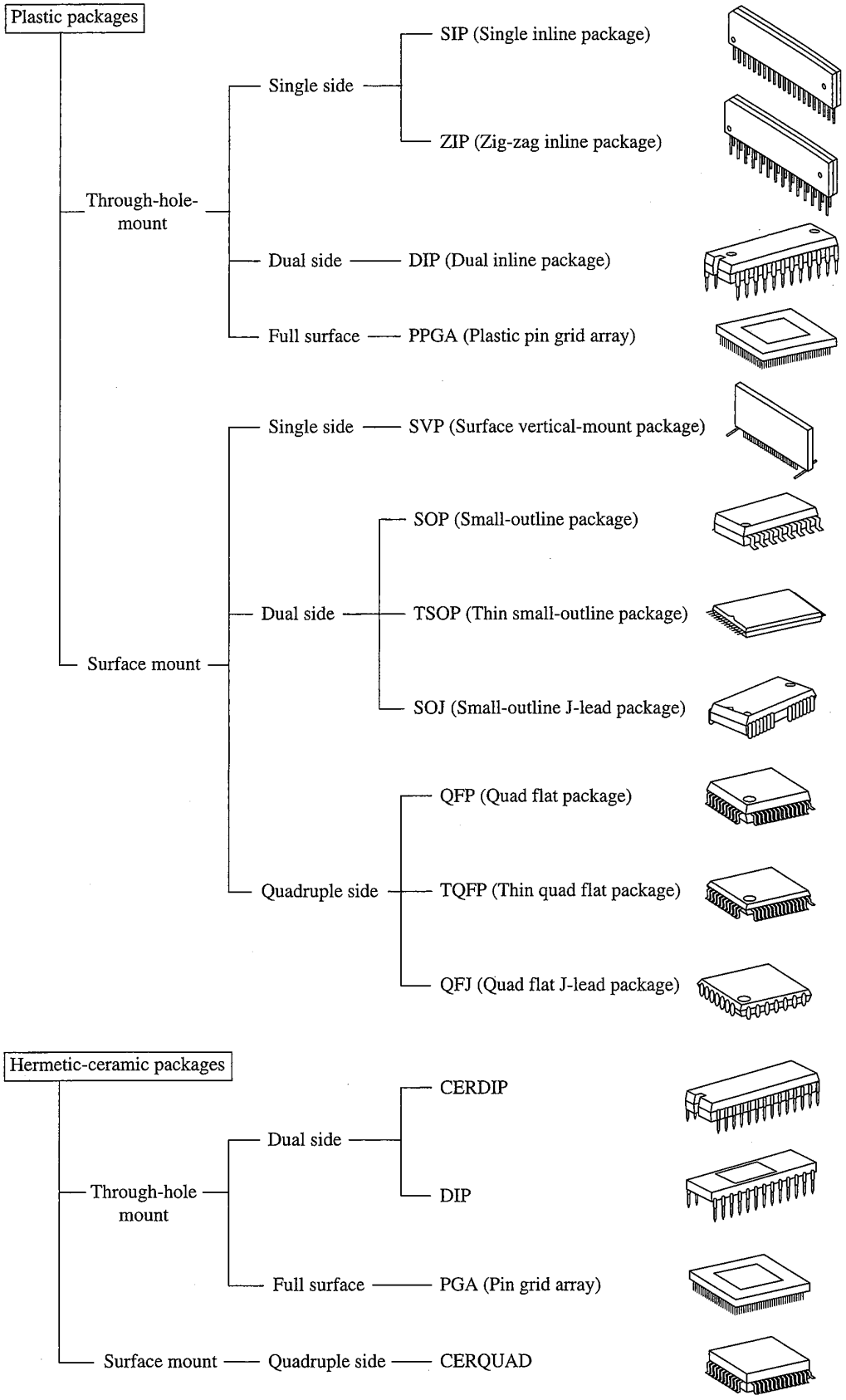


FIGURE 3
A variety of package designs. Not all ceramic package varieties are shown. (*Courtesy of Mitsubishi Electric Corp.*)

Micron Ex. 1029, p. 563
Micron v. YMTC
IPR2024-00911

Many package variations exist in the industry. Figure 3 illustrates some of these packages. (List of packages is based on standards and registrations of the Joint Electron Device Engineering Council (JEDEC).)

10.2.2 ULSI Package Trends

This section describes the recent trends for ULSI device packages. We shall consider device trends and their effects on packaging, newer packages for specific devices, and packaging level integrations.

Device trends and packaging

Memory chips have quadrupled in complexity, in terms of bits per chip, every three years. Today, 16-Mbit DRAM is in industrial use, and the subsequent generations, 64-Mbit DRAM and 256-Mbit DRAM, are at the research and development stage. The increase in complexity has accompanied a significant increase in chip size. Figure 4[3] shows the relationships among chip size, cell size, and DRAM density for several design generations. Each new generation increases the chip size approximately 1.5 times over that of the previous generation. Nevertheless, the trend for downsizing of electronic components requires that new chips be placed in the same

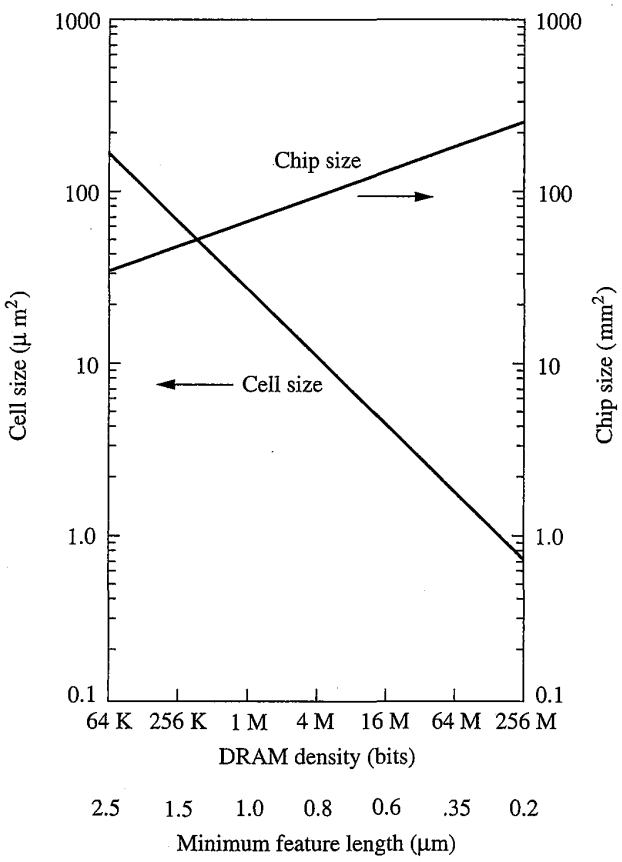


FIGURE 4
Relationship between chip size and DRAM generation. Bit density quadruples as the generation changes, every three years, because of the miniaturization of chip design features. (*After Prince, Ref. 3.*)

Micron Ex. 1029, p. 564
Micron v. YMTC
IPR2024-00911

size, or smaller, package as the previous generation. Memory packaging has so far succeeded in meeting the requirement by increasing the ratio of the chip-to-package area. The ratio is 70% in 16-Mbit DRAM.

Meanwhile, logic and microprocessor chips continuously increase in operating speeds, usually resulting in increased power consumption and higher I/O numbers in accord with the growth of circuit density. Figure 5[4] shows this trend in microprocessor devices, exploiting the SPECmark value that features operating speed efficiency. In CMOS logic devices, the relationship between operating speed and power consumption is expressed by[5]

$$P = \frac{(C \cdot V^2 \cdot \upsilon \cdot S_{\text{eff}})}{2} \tag{10.1}$$

where P is the power consumption, C is the load capacitance, V is the source voltage, υ is the switching cycle number, and S_{eff}, typically 0.4 through 0.6, is the circuit integration parameter.

Higher-speed operation means an increase in the υ value. Today, some state-of-the-art microprocessors consume as much as 30 watts. Consequently, package design, structure, and process techniques are affected by these increasing power requirements.

Rent's rule describes experimentally the relationship between I/O number and gate count in logic devices:[6]

$$\text{Number I/O} = \alpha(\text{Gate Count})^{\beta} \tag{10.2}$$

where α and β are constants determined by the device design. The rule has applied for logic devices that have partitioned roles in the system in which they are used.

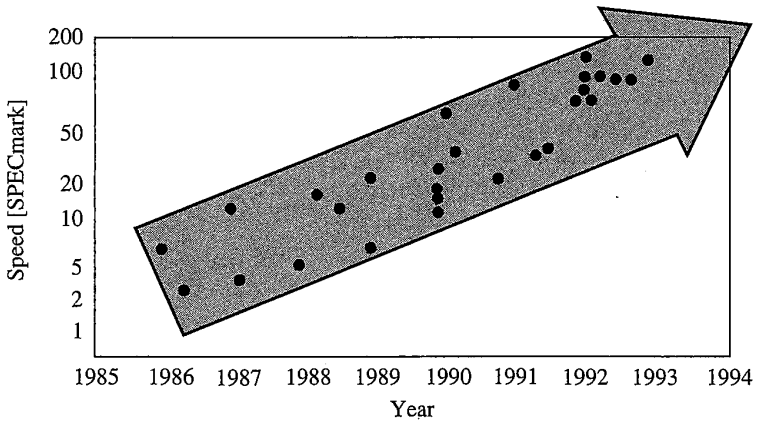


FIGURE 5
Trend of microprocessor devices. Each dot shows when a device of a given speed appeared in the industry. SPECmark is a typical benchmark value that expresses the operation speed of a microprocessor executing a standard program consisting of several calculations. (*After Nikkei Microdevices, Ref. 4.*)

Micron Ex. 1029, p. 565
Micron v. YMTC
IPR2024-00911

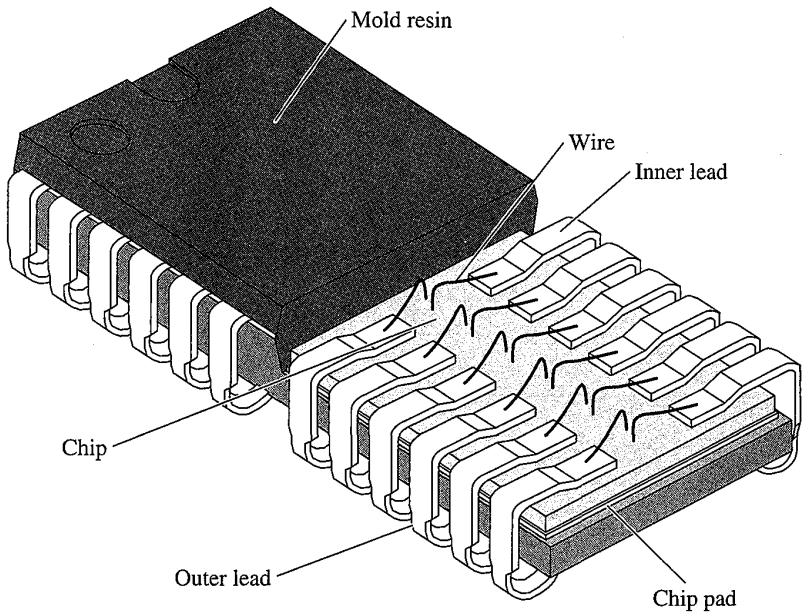


FIGURE 6
A typical LOC package structure. Tips of the lead frame extend above the chip center where the bonds are made. (*After Ueda, Ref. 9.*)

Gate counts, as large as 400,000 gates today and one million gates in the near future, require higher package pin counts in the range of 500 to 700 pins. These increases in package pin counts are caused not only by Rent's rule, but also by the increase in the number of power/ground pins. More power/ground lines are necessary for larger power consumptions and high-speed operations without electromagnetic noise. In leading-edge microprocessors, the bus-line-width has grown to 16 bits, 32 bits, and 64 bits. This growth also dictates the increase in package pin counts. To meet these requirements, leading-edge packages for ULSI devices that have more than 1000 pins are in the research stage.

Packages for memory ULSI devices

Memory devices usually must be low cost, since many of them are used in electronic appliances. The need for low-cost memory devices has been driving the trend for cheaper plastic packages. A variety of SM plastic packages have been developed and are in industrial use, such as SOJ, SOP, and thin SOP (TSOP). Except for TSOP, these have bodies typically 2 mm thick. TSOP packages have 1-mm-thick plastic bodies suitable for compact appliances. As discussed earlier, the chip occupancy continues to grow, and the stringent requirements this imposes have led to considerable changes in package structures. The lead-on-chip (LOC) structure,[7] in which wire interconnections within the package are made above the chip-circuitry surface, is notable.[8] In conventional packages for older-generation devices, such as that shown in Fig. 2, the interconnections were made only in the periphery and outside the chip area, exploiting an additional area for the interconnections. Figure 6[9] shows a typical LOC package. Here the tips of the lead frame extend over the chip

Micron Ex. 1029, p. 566
Micron v. YMTC
IPR2024-00911

surface, and Au wires are stitch-bonded to the lead frame tips to connect them with the chip bonding pads, which are located in the interior of the chip area. The LOC structure increases the chip occupancy to more than 70% of the package area. The structure provides chip design flexibility because it allows the pads to be located on the chip in almost any position. It also allows the placement of substitutes for additional inner lead portions (called bus-bars) that work as alternatives to power and ground Al metallizations on the chip circuitry. The bus-bars are numerous in some high-speed memory designs, since they can enhance the electrical performance of the devices without increasing the chip size. In current ULSI memories, LOC and chip-on-lead (COL) structures, in which the electrically insulated chip resides on the tips of the inner leads, are widely accepted. Some of these packages are vertical SM types, forming a category called surface vertical-mount packages (SVPs).[10]

Packages for logic and microprocessor ULSI devices

Figure 7 shows packages of choice for logic and microprocessor ULSI devices in terms of I/O pin count capability and corresponding external terminal spacing choices. PGA, a typical through-hole, high-pin count package, is widely used up to approximately 500 pins, because the package is easy to solder to the PWB and generally achieves a well-balanced performance among various requirements. Usual pin pitches are 2.54 mm (100 mil) and 1.78 mm (70 mil) for through-hole use and 1.27 mm (50 mil) for an SM package to prevent increasing the package size in higher pin count applications. The quad flat package (QFP) is the prevalent type used in logic and microprocessor packaging that requires up to approximately 300 pins. QFPs that vary in pin count, geometric dimensions, and package internal structure are now in industrial use, and their external lead pitches are shrinking continuously. The 1.27 mm pitch is now rather obsolete, because pitches are decreasing to 0.8 mm, 0.65 mm, and 0.5 mm. Today, 0.4 mm and 0.3 mm are the finest pitches in use. QFPs that can accomodate thinner bodies suitable for smaller electronic components are being standardized as low profile QFP (LQFP) and thin QFP (TQFP), segregated by their overall heights. The quad tape carrier package (QTP) enables more than 500 pins and introduces finer pitches of, typically, 0.25 mm or 0.2 mm. QTP is one of the choices to meet the highest pin count requirements. However, the pitches challenge both single-chip and next-level packagings stringently.

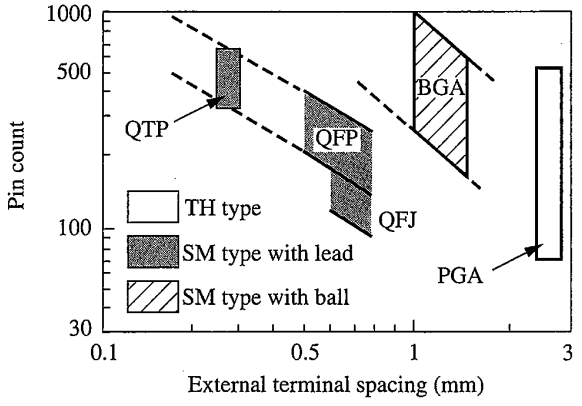


FIGURE 7
Packages of choice for logic and microprocessor devices in terms of I/O pin count and external terminal spacing.

Micron Ex. 1029, p. 567
Micron v. YMTC
IPR2024-00911

Newer packages, such as the ball-grid array (BGA) package,[11] which has external terminals on one surface of the package and not on the sides, have been proposed as candidates to solve the problem. These packages provide larger external terminal pitches of 1.27 mm or 1.0 mm, for example, since they are located in a matrix. We discuss the details in Section 10.7.

Packaging-level integrations

As circuit integration proceeds on the chip, it also proceeds in the packaging through interactions among several levels of packaging. Generally, packaging exclusive of the final system construction is classified into three levels, as shown in Fig. 8. Final system requirements determine a specific selection of the packaging method, or how to combine the levels.[12, 13, 14] Types 1 through 4 show the major methods that have been used in the industry. Type 1 is the most common choice. In type 1, the chip is first packaged as a single chip and then packaged at the third level, typically at the PWB level. Types 2 and 3, usually called multi-chip-module (MCM) technologies, are used in high-performance systems, typically in mainframes. In type 2, the chip is single-chip-packaged as in type 1, and the packaged chips are then packaged at the second level onto a smaller substrate; this forms a functionally larger and geometrically smaller unit and utilizes the finer multilayer wiring of the substrate. The substrate is attached to a larger mother board in the third-level packaging. Type 3 is similar to type 2 in that it uses a smaller substrate as an intermediary stage. However, bare chips are attached directly to the substrate, usually providing a superior electrical and geometrical performance but with some drawbacks, such as more difficult testability, increased cost, and lower yield. Type 4 is the simplest way of packaging, where bare chips are attached directly to the system board.

ULSI devices with complex functions will affect packaging significantly; that is, type 1 has penetrated into higher-performance applications, such as the engineering work station trend, and type 4, which originated in consumer products, is now seen in some personal computers. The single-chip packaging technologies discussed in the following sections are in transition, but they continue to be fundamental techniques.

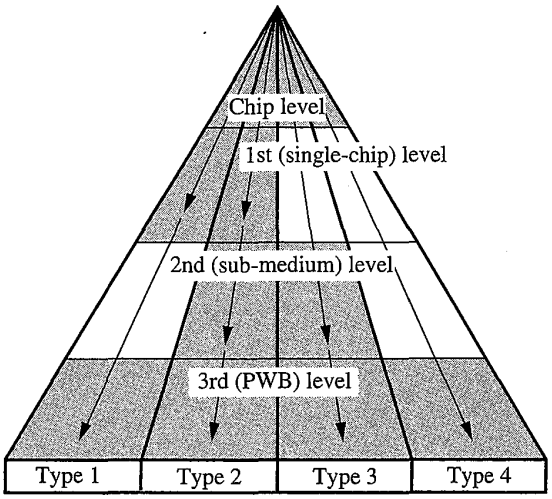


FIGURE 8
Circuit-level integrations for final system construction.

Micron Ex. 1029, p. 568
Micron v. YMTC
IPR2024-00911

10.3 ULSI ASSEMBLY TECHNOLOGIES

This section covers the basic assembly operations in use today for ULSI devices. We consider the steps from wafer preparation through wire bonding, including some design considerations that are closely related to assembly process technologies. Figure 9 shows a generic assembly flowchart that applies to both plastic and ceramic packages.

10.3.1 Wafer Preparation

Wafer preparation starts with the back-grinding process at the end of the wafer fabrication process. Large-diameter, and hence thick, wafers need to be thinned before assembly begins. For example, 6-inch-diameter wafers are thinned from 650 μm to approximately 400 μm thickness. Thicker wafers place more demands on the dicing process. Moreover, they place other demands on package design and assembly processes,[15, 16] such as chip interconnection, encapsulation, and final packaged chip reliability. In thinner SM ULSI packages that need higher area or volume densities, thinner wafers are especially beneficial to reduce the thermal stresses resulting from mismatching of the thermal coefficient of expansion (TCE) between the larger ULSI chip and the plastic molding materials described later. Back-grinding is done by fully

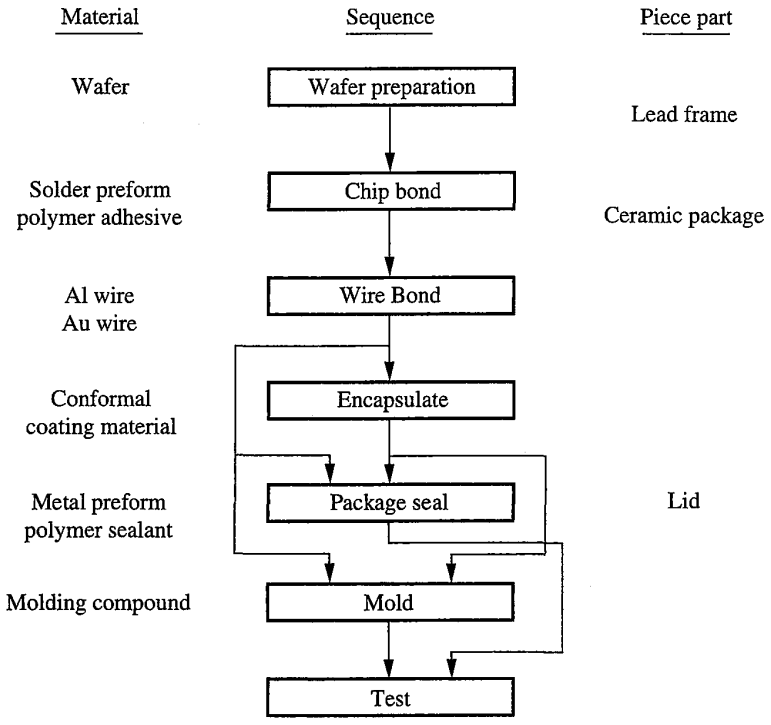


FIGURE 9
Generic assembly sequence for plastic and ceramic packages.

Micron Ex. 1029, p. 569
Micron v. YMTC
IPR2024-00911

automated machines. This operation must be done carefully. Poor operation degrades the mechanical strength of the chips and may lead to chip-crackings, which begin in the back after the chips are packaged. If required, back-side metallization is done after back-grinding. Au, multilayer metallization of Au-Ni-Ag, and Ti-Ni-Au, in order from the silicon side, are the usual choices for metallization. Proper combinations of the back-side metallization materials and chip-bonding operations (described later) provide superior adhesion strength and electrical connections.

A diamond-blade dicing saw separates wafers into individual chips. Wafers are typically adhesive-mounted to a tape that has been preassembled to a frame before dicing. An advanced diamond blade, typically 25 μm thick rotating at a speed of 20,000 rpm, cuts the wafer from 90 to 100% saw-through, allowing dicing streets as narrow as 60 μm with dissimilar materials such as passivation, metallization, and gold usually serving as test patterns. Established methods such as scribe-and-break or 50% saw-through dicing-and-break are not usually acceptable for ULSI devices and have been replaced by 90 through 100% saw-through dicing, which leads to better-quality cuts without damage to the chip surface. Wafer-dicing machines are completely automatic and contain features such as alignment systems, an integral wafer-cleaning station, a drying oven, and quality monitors. The sawed-apart wafers, still mounted in the tape-frame fixtures, are then loaded into the chip-bonding operation, where an automatic bonder sorts good chips either by recognizing the bad-chip marks or by reading the positional mapping data from the wafer test operations.

10.3.2 Chip Interconnection

Chip interconnection consists of two steps. In the first step, chip bonding, the back of the chip is mechanically attached to an appropriate medium, such as a ceramic substrate, a multilayer ceramic-package piece part, or a paddle of a metal lead frame. The chip bonding operation is necessary as preparation for the next interconnect operation, wire bonding. Chip bonding also provides a thermal path from the chip to the medium and sometimes allows electrical connections to be made to the back of the chip. In the second step, the bond pads on the circuit side of the chip are electrically interconnected to the package by wire bonding. Fine metal wires (Au or Al) are used for wire bonding for SSI up to USLI packaging. Alternatives also used today for ULSI devices are tape-automated bonding (TAB) and flip-chip solder bonding, described in Section 10.7.

10.3.3 Chip Bonding

Chip bonding is the first operation in which the chip meets the package. The bonding technology is determined by selecting among a variety of trade-offs, such as thermal stresses caused by TCE mismatch between the chip and the package materials, power dissipation, electrical capabilities, resulting reliabilities, and cost. This selection has become more difficult and sophisticated in ULSI, because the growing chip size and more stringent demands on the performance of the packaged chip allow little room to make engineering trade-offs. Some features that were negligible in SSI,

Micron Ex. 1029, p. 570
Micron v. YMTC
IPR2024-00911

MSI, and even in VLSI have led to degradations of performance and reliability in ULSI. Chip-bonding technology has become more difficult; hence, before describing the chip-bonding process itself, we discuss the major structural problems of ULSI packaging.

The first problem that increased ULSI chip size presents is the TCE mismatch between the chip and the package medium to which the chip is attached. The TCE mismatch causes various thermal stresses in several operations that the chip undergoes. The first stress occurs when the chip is cooled to room temperature after chip bonding. Then wire bonding and encapsulation temperatures create major additional stresses during assembly operations. Final test operations, including burn-in, soldering at the next packaging level, and power cycling in actual use mixed with cooling conditions, place different and integrated stresses on the devices. Sometimes these stresses cause chip cracking or chip delamination from the package medium, resulting in degraded thermal performance and reliability. In plastic packages, the stress, combined with plastic moisture absorption, might lead to package cracking rather than chip cracking; this also results in degraded reliability.

Figure 10[17] shows TCEs of materials for semiconductor packaging. Materials that have TCEs close to that of a silicon crystal are preferable in package construction. Choosing only the packaging materials that have TCEs the same as that of silicon would be ideal, reducing the thermal stresses to zero. Realistically, however, no material in industrial use allows such a choice. In ceramic packages, Al_2O_3 substrates that provide good TCE matching have been preferred for many years. The established chip-bonding process for the combination of a silicon chip and Al_2O_3 is the Au-Sn eutectic method. Although this method, called hard soldering, provides

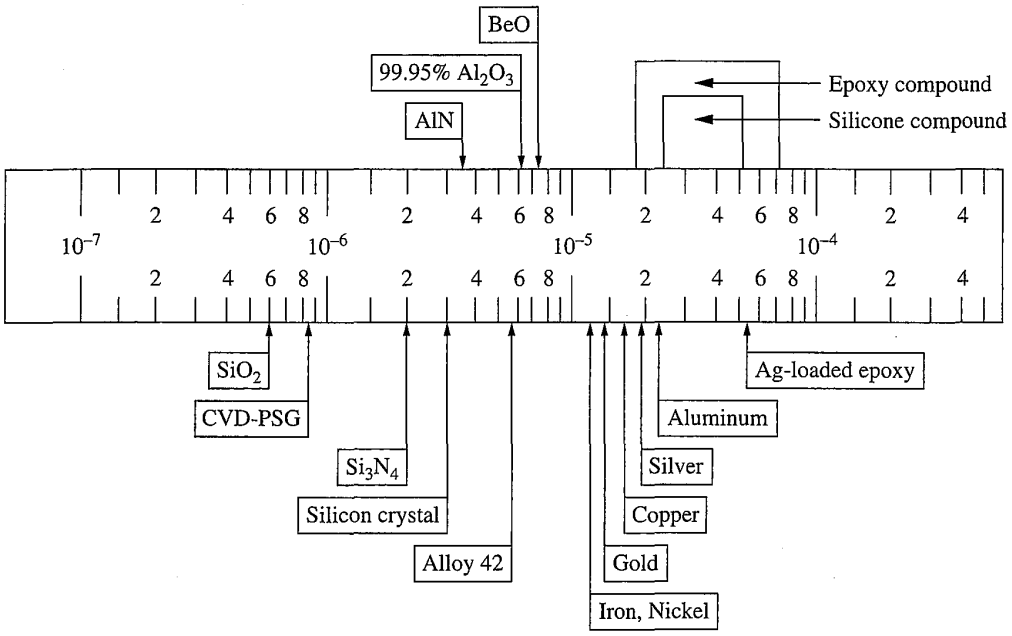


FIGURE 10
Thermal coefficients of expansion (TCE) of materials used in semiconductor devices, in $°C^{-1}$. (*After Kakei, Ref. 17.*)

Micron Ex. 1029, p. 571
Micron v. YMTC
IPR2024-00911

little stress buffer effect, it has an effect even in ULSI chip bonding where the TCE mismatch is smaller and the established bonding technique is appreciated. In plastic packages, the Alloy 42 (42% Ni–58% Fe alloy) lead frame has been widely used for the same reason. Chip-bonding techniques began with Au-Si hard soldering for SSI; later, resin chip-attach materials and Pb-Sn solders (soft solders) were used to bond MSI chips up to approximately 5 mm square because they were better stress buffers. However, in ULSI packaging, where enhanced thermal and electrical capabilities are sometimes more important than the higher reliabilities that come with better TCE matching, Alloy 42 no longer can be the single candidate. Most of the heat that a chip dissipates usually flows to the package medium through the chip attachment. The medium also determines the allowable transient heat-peak value. In ULSI plastic packages, especially in logic and microprocessor applications, a significant number of packages now use Cu alloys that provide approximately ten times the thermal conductivity of Alloy 42. Cu alloys that have smaller permeability are valued highly for their electrical superiority as well. The adoption of a Cu alloy challenges the chip-bonding technique to solve the TCE mismatch problem. Today, improved, epoxy resins with Ag flakes, which have smaller elastic moduli (E), are commonly used along with improved process control techniques. Controlled chip-attach material thickness (usually as thick as 10 to 20 μm) and a smaller E value compensate TCE mismatch to some extent.

The second problem is the TCE mismatch between the chip and the molding compound in plastic packages. This mismatch does not exist in ceramic packages, because no packaging materials come into contact with the chip surface. Although the α_1 value (that is, the TCE below the glass transition temperature, T_g) of the improved molding compound has been reduced to 10 to 20 ppm/°C, the mismatch is still significant because the ULSI chips are large. The resulting thermal stresses on the chip start as compressive stresses after the molding operations, where the chip is cooled from approximately 175°C to room temperature. Of course, the stress is complicated, since the package at this stage has a composite structure consisting of a chip, a metal lead frame, and plastic molding compound. Reduction of the TCE of molding compounds, typically to 7 ppm/°C, provides better matching with the chip, but such an excessive reduction induces a serious side effect—an intolerable TCE mismatch between the molding compound and the Cu alloy lead frame. In ULSI plastic package constructions, Cu alloy lead frame materials, which usually have a TCE of 17 ppm/°C, are becoming more popular, but they require a larger TCE of the molding compound, from 10 to 15 ppm/°C. Countermeasures are required, such as increasing adhesion strengths of contacting package components, optimizing geometric dimensions, and modifying structure.

The third and usually the last TCE mismatch problem is between the completed package and the board to which the package is attached. This problem is discussed in Section 10.5. Table 1 summarizes the mechanical characteristics of packaging materials that influence TCE mismatches.

In the chip-bonding operation, prepared chips are attached to the package media, typically paddles of a lead frame or ceramic package cavities, using an automatic chip bonder. The chip bonder usually has features such as a good-chip sorting function, a pick-up and place unit, a chip-attach material supply station, and a lead frame loader and unloader.

Micron Ex. 1029, p. 572
Micron v. YMTC
IPR2024-00911

TABLE 1
Mechanical characteristics of packaging materials

Material	Thermal coefficients of expansion ($\times 10^{-6}$/K)	Elastic modulus (GPa)	Thermal conductivity (W/m-k)
Silicon	2.6–3.6	10.2–16.3	150
SiO_2	0.6–0.9	≈ 7.1	1.4
SiN	2.8–3.2	≈ 32.6	3.0
Polyimide	20–50	3–4	0.12–0.2
Au	1.4	78	318
Al	23	70.5	229
Cu	16.7	112.7	394
Cu alloys	16.3–17.3	117–132	151–351
Ni-Fe alloys (Alloy 42)	4.14	147	15
Kovar	4.4	143	16.7
Mo	3.7–5.3	343	130
37Pb-63Sn	24.7	32.1	49
95Pb-5Sn	29.8	7.4	63
80Au-20Sn	12.3	59.2	24
Al_2O_3	6.7–7.1	262–378	19–25
AlN	4.4–4.9	326–357	70–170
Glass epoxy	16	0.26	0.36
CuW	6.0–7.0	347	210
Epoxy	12–22	13.2–17.3	0.67
Ag epoxy	33	1.96–6.86	1.26

The major choices for a chip-attach process include hard (or eutectic) solder, polymers, and Ag-filled glasses. A comparison of the chip-attach processes for hermetic-ceramic and plastic packages shows two emerging technologies for ULSI—eutectic for hermetic-ceramic packages and polymers for plastic packages. The eutectic process is essentially contamination-free, has excellent shear strength, and ensures low-moisture packages. The major disadvantages are that it is difficult to automate and it places high thermal stresses on the chip because of the high process temperature. Typical polymer chip-attach materials for ULSI in plastic packages are epoxies, with and without Ag flakes to enhance electrical and thermal conduction. Epoxies with Ag flakes are more popular. Formerly, epoxies were high in hydrolyzable ions, which could directly affect reliability. Epoxy suppliers have improved their materials, controlling the residual sodium and chloride ions to less than 10 ppm each. Epoxies are widely used today for ULSI.

Although Ag-filled glasses and polyimides, with and without Ag, are choices in some ULSI packages, the following sections describe only the eutectic and epoxy chip-attach processes, because they are the most widely used today.

Eutectic chip bonding

Figure 11 illustrates the fundamental aspects of a chip bond. Eutectic chip bonding metallurgically attaches the chip to a substrate material, typically a metal lead frame made of Alloy 42, or to a ceramic substrate, usually 90 to 99.5% Al_2O_3. Metallization is often required on the back of the chip to make it wettable by the

Micron Ex. 1029, p. 573
Micron v. YMTC
IPR2024-00911

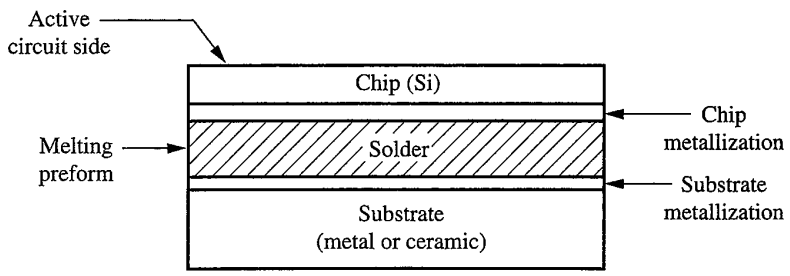


FIGURE 11
The basic structure of a silicon device chip-bonded with a metal preform.

chip-bonding preform. The preform is a thin sheet, less than 0.05 mm thick, of the appropriate solder-bonding alloy. Table 2[18] lists the compositions and melting points for preform materials. The melting point must be higher than the temperature of the next operation, usually wire bonding, otherwise remolten preform makes the operation difficult. The substrate material is usually metallized with plated Ag (lead frame) or Au (lead frames or ceramic). Solder chip bonding to ceramic packages, which are to be hermetically sealed, is usually performed with a Au or Au–2% Si preform. The TCE mismatch between silicon and the ceramic substrate is relatively small, so thermal stress caused by a TCE mismatch is not a critical issue. Still, the bonding must be done carefully to create uniform wetting; partial wetting induces undesirable stress concentrations on the bond and may lead to chip cracking. In the presence of some mechanical scrubbing and a temperature of about 370°C, the preform reacts to dissolve the silicon. The Au-Si eutectic composition is reached and then exceeded, because the bonding temperature is set to more than 400°C. As the composition of the composite structure becomes more silicon-rich, it freezes and the chip bond is completed. In other applications that require lower temperatures or more ductile chip-bonding solders, Pb-Sn-based solders are widely used.

Epoxy chip bonding

Silver-filled epoxy adhesives are major choices as polymer-based, chip-attach materials. The silver filler, usually flakes, makes the epoxy both electrically conduc-

TABLE 2
Compositions and melting points for chip-attach preforms[18]

Composition	Temperature (°C) Liquids	Solids
80% Au, 20% Sn	280	280
92.5% Pb, 2.5% Ag, 5% In	300	—
97.5% Pb, 1.5% Ag, 1% Sn	309	309
95% Pb, 5% Sn	314	310
88% Au, 12% Ge	356	356
98% Au, 2% Si	800	370
100% Au	1063	1063

Micron Ex. 1029, p. 574
Micron v. YMTC
IPR2024-00911

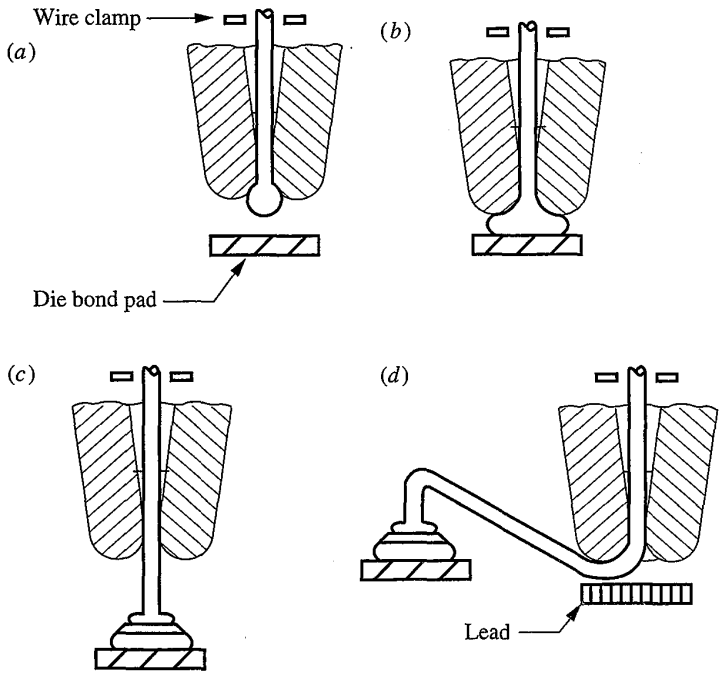


FIGURE 12
Tailless ball-and-wedge bonding cycle. (*a*) The capillary targets the chip's bond pad and is positioned above the chip, with the ball formed on the end of the wire and pressed against the face of the capillary. (*b*) The capillary descends, bringing the ball in contact with the chip. The inside cone, or radius, grips the ball in forming the bond. In a thermosonic system, ultrasonic vibration is then applied. (*c*) After the ball is bonded to the chip, the capillary rises to the loop-height position. The clamps are open and wire is free to feed out of the end of the capillary. (*d*) The lead of the devices is positioned under the capillary, which is then lowered to the lead. Wire is fed out the end of the capillary, forming a loop.

tive, to provide lower resistance between the chip and the substrate, and thermally conductive, to allow a good thermal path between the chip and the rest of the package. Silver-filled epoxies are less expensive than the high-gold-content hard solders, are more flexible in absorbing the thermal stress between the chip and the substrate, and make it easy to automate the attach processes. In ULSI manufacturing, epoxy is fed onto the substrate material through a multinozzle or a well-designed, single-nozzle dispenser to ensure the required bond line thickness is created without voids. The back of the chip often does not require metallization because epoxy provides better adhesion to the natural silicon oxides on the back, whereas the substrate is usually metallized because the chip-attach portion is plated when other necessary substrate portions (such as stitches for wire bonding) are plated. The automatic chip-attach operation proceeds at a speed of 1 to 2 sec per chip. Epoxy chip bonds do not need mechanical scrubbing, unlike eutectic chip bondings; hence, there is a lower probability of damage to the chip edges in an epoxy chip bond. Because epoxies are thermosetting polymers, they must be cured at elevated temperatures to complete the chip bond. Typical cure conditions range from 125 to 175°C and require 1 to 2 hours.

Micron Ex. 1029, p. 575
Micron v. YMTC
IPR2024-00911

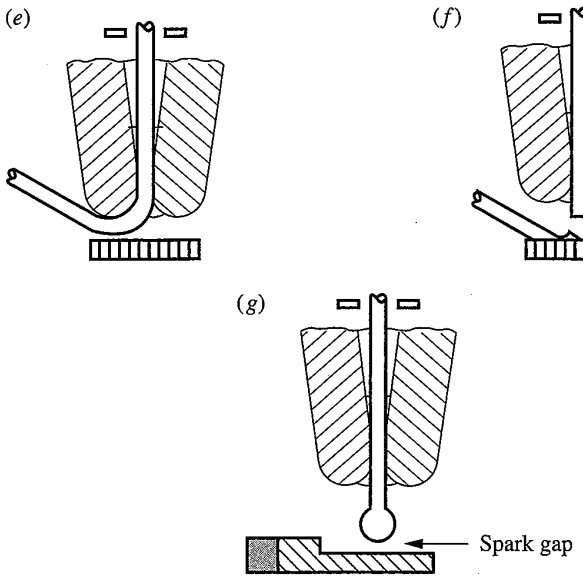


FIGURE 12 (*continued*)
(*e*) The capillary deforms the wire by pressing against the lead, producing a wedge-shaped bond that has a gradual transition into the wire. In a thermosonic machine, ultrasonic vibration is then applied. (*f*) The capillary rises off the lead, leaving the stitch bond. At a preset height, the clamps are closed while the capillary is still rising with the bonding head. This prevents the wire from feeding out of the capillary, and pulls at the bond. The wire breaks at the thinnest cross section of the bond. (*g*) A new ball is formed on the tail of the wire that extends from the end of the capillary. A hydrogen flame or an electronic spark may be used to form the ball. The cycle is completed and ready for the next ball bond. (*After Kulicke and Soffa Industries, Ref. 19.*)

10.3.4 Wire Bonding

Wire bonding is the most common method for connecting the pads on the chip to those on the package. It is performed on chips after they have been chip-bonded to the appropriate piece part. Au or Al is usually chosen as the wiring material because they bond well to pads on the chip and to the metallized portions on the package. In addition, these wires form metallurgically stable connections as long as the bonding conditions are properly controlled, and the manufacturing processes of fine (typically 25 to 30 μm in diameter) wires are mature and well-prepared. Typically, gold wire is ball-and-wedge bonded by thermosonics or thermocompression; that is, it is ball-bonded to the chip bond pad (Al) and wedge-bonded to the package substrate (Au or Ag), as shown in Figure 12.[19] One advantage of ball-and-wedge bonding is that the wedge bond can be performed anywhere on a 360° arc around the ball bond. There are two reasons for this. First, the symmetrical geometry of the capillary tip allows the form action of a symmetrical first bond from which the wire can dress in any direction. Second, the basic metallurgical connection of the ball-bond is performed through thermocompression, not ultrasonic vibration. In the ultrasonic bonding of Al wire, the wire direction needs to be the same as that of the vibrating horn (the machine head that holds the capillary tip or wedge), whereas thermocompression

Micron Ex. 1029, p. 576
Micron v. YMTC
IPR2024-00911

allows the wire to dress in any direction regardless of the horn direction. As the horn is placed in a designated direction, the versatility of ball-and-wedge bonding allows the wire to dress in any direction around the ball. Au wire ball-and-wedge bonding is widely used in ULSI packaging, especially in plastic packaging operations. The state-of-the-art, fully automatic ball-and-wedge bonders run at speeds over 6 wires per sec. Usually, ball bonds are Au-Al and wedge bonds are Au-Au or Au-Ag metallurgical diffusions performed at bonding temperatures ranging from 300 to 350°C in thermocompression bonding and from 150 to 250°C in thermosonic bonding.

The major issue in establishing high-quality, reliable wire bonding is process control of variables such as bonding temperature, loads, and ultrasonic vibration magnitudes. Visual inspections performed according to MIL-STD-883 METHOD 2010 can detect defects such as chip cracks, missing wires, and poor wire shapes that might cause short circuits. In addition, two types of process tests are used today—the wire-pull test and the ball-shear test. Wire-bond pull tests have long been used to evaluate both ball and wedge bonds. The test is done by engaging the wire with a hook near the center of the wire span and pulling to destruction. Failure modes in the test are shown in Fig. 13.

The ball-shear test (Fig. 14) has also been around a long time. It evaluates the quality of ball bonds by pushing the ball to destruction; this characterizes the Au-Al intermetallic formation, which usually cannot be characterized by the wire-pull test.

Metallurgical diffusion, in its earlier stage, follows the equations

$$X^2 = Dt \tag{10.3}$$

$$D = D_o \exp(-Q/RT) \tag{10.4}$$

where X is the diffusion thickness, D is the diffusion constant, t is the storage time, D_o is the frequency factor, Q is the activation energy, R is the gas constant, and T is the storage absolute temperature. D_o and Q are constants determined by the combination of the two materials.

Gold and aluminum form a variety of intermetallics; of these, Au-Al is formed first and gradually changes to Au-Al_4, degrading the bond strength. Newer chip

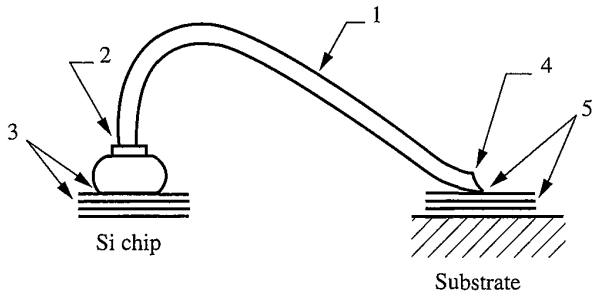


FIGURE 13
Failure modes in wire-bond pull tests. (1) Break in wire. (2) Wire break above the ball. (3) The ball-bond failure; it should be determined whether the location is in the Au-to-Al or in the Si chip buried layers. (4) Break in the wire at the wedge-bond heel. (5) Wedge-bond interface failure; it should be determined whether the location is in the Au-to-Ag or in the metallized layers.

Micron Ex. 1029, p. 577
Micron v. YMTC
IPR2024-00911

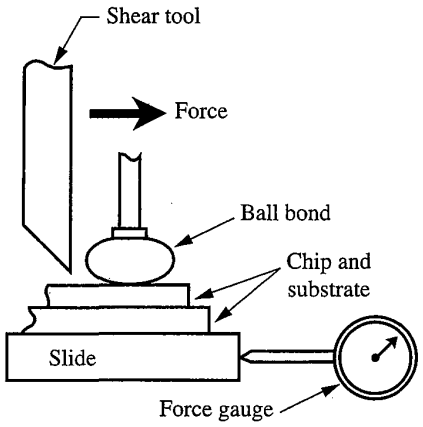


FIGURE 14
Setup of the shear test. Force is applied by the shear tool and the force needed to shear is recorded on the force gauge.

metallization materials, such as Al-Si-Cu, Al-Cu, and Cu, that are either being used in the industry or still being studied, must be accompanied by improvements in wire bonding techniques; otherwise, they invite poor interconnections, leading to degraded reliability.

Al wire wedge bonding (Fig. 15) is another major choice in ULSI wire bonding techniques. It is widely accepted in hermetic-ceramic packages.

ULSI design rules

The establishment of, and adherence to, good chip design rules is absolutely essential to achieve high yields in ULSI package assembly. Rules must be generated for the particular package type used and must be compatible with the assembly equipment. These decisions should be made early, preferably before the chip layout is started and definitely before the layout is completed. Bonding-pad design rules are especially important. In memory ULSI, pads are often placed only on the two facing sides of the chip; or they are located on the chip-center in LOC packages,

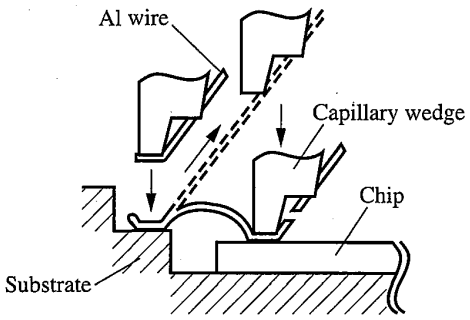


FIGURE 15
Al wire wedge bonding. A capillary wedge holding the wire descends and presses the wire against the pad on the substrate. The first bond is made by applying ultrasonic vibration. Then the capillary wedge rises to the loop-height position and moves to target the pad on chip, forming a loop. The capillary wedge is lowered to and pressed against the pad on chip. After the second bond is produced, wire clamps (not shown in the figure) pull and break the wire, and the bonding cycle is completed.

Micron Ex. 1029, p. 578
Micron v. YMTC
IPR2024-00911

to increase the chip occupancy or the area efficiency of the PWB where the packages are placed. In logic and microprocessor ULSIs, the growing I/O count and the shrinking active device size present stringent constraints on chip design. The space required for interconnection not only represents a large fraction of the chip area, but sometimes determines the chip-size, and considerable extra space is needed around the active area for pad placement. To avoid this problem, the effective bonding-pad sizes and spacings (also called pitch) should be reduced. Today, possible spacings range from approximately 100 to 150 μm. Figure 16 shows the relationships between chip size and the allowable pin count for several spacings. Typical design rules for wire-bonding pads are illustrated in Fig. 17; these cover recommended pad sizes and pitch, clearances between pads and the edge of the chip, and clearances to internal adjacent metal conductors of other critical design features.

The chip designer must understand another essential restriction that affects package selection and bonding-pad layout. Bond pads on packages, metallized pads in ceramic packages, or lead frame tips in plastic packages need particular spacings that are determined by the capabilities of the manufacturing process. Lead frames are usually chemically etched or mechanically stamped using a progressive die to produce designated patterns from 100- to 150-μm-thick flat metal. For lead frames, the aspect ratio in the etching process or the strength of the die parts determines the spacings.[20] For ceramic packages, screen printing of the metallized-conductor is the constraint. Today, the finest spacing in industrial use is 220 μm, both for lead frames

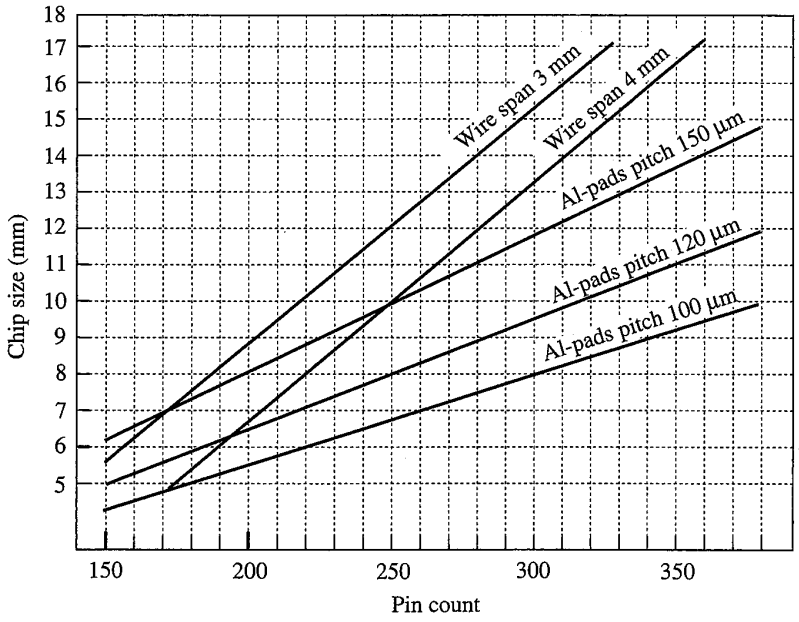


FIGURE 16
Pin count and required minimum chip size for several wire spans and Al-pad pitches. Where the lines of the two constraints, wire span and Al-pad pitch, cross, the upper line represents the bottleneck restriction.

Micron Ex. 1029, p. 579
Micron v. YMTC
IPR2024-00911

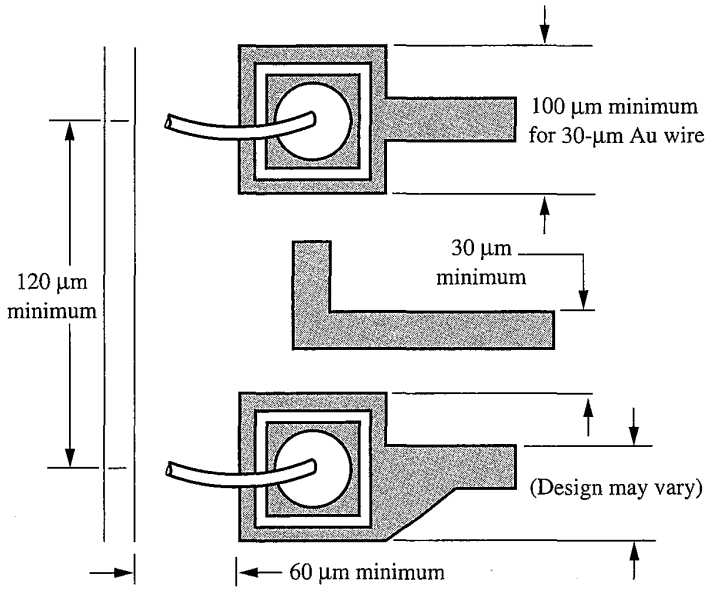


FIGURE 17
Typical bonding pad sizes and spacing. (*Courtesy of Mitsubishi Electric Corp.*)

and metallized conductors. The maximum allowable I/O number or package bond-pad number that can be placed around a chip that has specific geometric dimensions is determined by the package bond-pad spacings and the possible wire spans that connect the chip and the package. Wire span (from the pad on the chip to the substrate or to the lead frame bond pad) is usually about 1 to 4 mm, and the resulting pin count limitations are shown in Fig. 16 for several wire spans. Longer wire could lead to potential edge shorts to the chip and shorts to adjacent wires, or to the conductor on the lead frame or metallized conductors of the package, since longer wire is more likely to droop or deform. In plastic packages, longer wire can also be deformed by the resin flow during the molding operation. This wire span is often stringently restrictive in high-I/O chip design; hence, terraced packages, often combined with staggered pads on the chip, are sometimes used as a solution (Fig. 18).

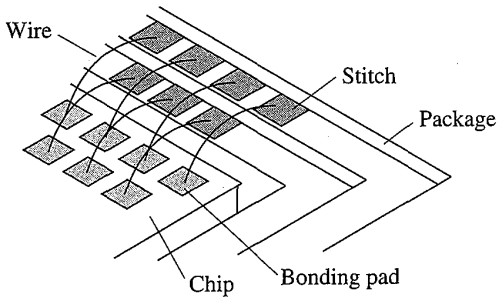


FIGURE 18
A terraced package and staggered pads on the chip—an arrangement that results in a lower-pitch bonding.

Micron Ex. 1029, p. 580
Micron v. YMTC
IPR2024-00911

These design rules can be integrated into chip-design CAD systems, so that the chip designer can choose a proper lead frame (or a substrate); the location of all lead frame wedge-bond targets and the optimal location for the ball bond on the chip can appear as design templates in the system.

10.4 PACKAGE FABRICATION TECHNOLOGIES

Single-chip packages are usually based either on refractory-ceramic technology or molded plastics. Ceramic packages are usually used for state-of-the art devices where maximum reliability or high-power treatment is required. For more mature products where low cost is critical but a hermetic seal is still required, ceramic dip (CERDIP) technology is employed. This technology uses a combination of lead frames, dry-pressed refractory ceramic parts, and glass sealing. Plastic packaging usually is used for more mature products where cost is important and a hermetic seal is not required. However, rapid and continuous improvements in plastic technology, such as highly reliable plastic resins, proper molding process controls, and progress in relevant techniques, have been broadening its applications to leading-edge ULSI devices. This section discusses processing technologies that generally apply to a variety of package types.

10.4.1 Ceramic-Package Technology

Ceramic-package fabrication consists of the operations that are done before the chip assembly operations. Figure 19[21] illustrates multilayer ceramic technology. A dispersion, or slurry, of ceramic powder and liquid vehicle (solvent and plasticized resin binder) is first prepared and then cast into thin sheets by passing a leveling blade (doctor blade) over the slurry. After they dry, the sheets are cut to size, and via holes (holes through the dielectric layers, through which interconnections are made) and cavities are mechanically punched into the sheet. Then custom wiring paths, usually a slurry of tungsten powder, are screened onto the surface and the via holes are filled with metal. Several of these sheets are press-laminated together in a precisely aligned fixture, and the entire structure is fired at 1600°C to form a monolithic sintered body. The refactory ceramic technology is a complex process requiring careful process control.

After the laminate is sintered, it is ready for the finishing operations of lead attachment and metallization plating. Nickel is plated over the tungsten in preparation for lead brazing. The lead material is either an Fe-Ni-Co alloy called Kovar, or an Fe-Ni alloy, usually Alloy 42; the brazing material is a silver-copper eutectic alloy. All exposed metal surfaces are electroplated or chemically plated (usually gold over nickel) for bondability and environmental protection. Multilayer ceramic packages can be made of sizes up to 100×100 mm in lateral dimensions to a tolerance of $\pm 0.5\%$, and they can consist of up to 30 layers in the most advanced processes.

Ceramic-package technology is very effective for constructing complex packages with many signal, ground, power, bonding, and sealing layers. It does have,

Micron Ex. 1029, p. 581
Micron v. YMTC
IPR2024-00911

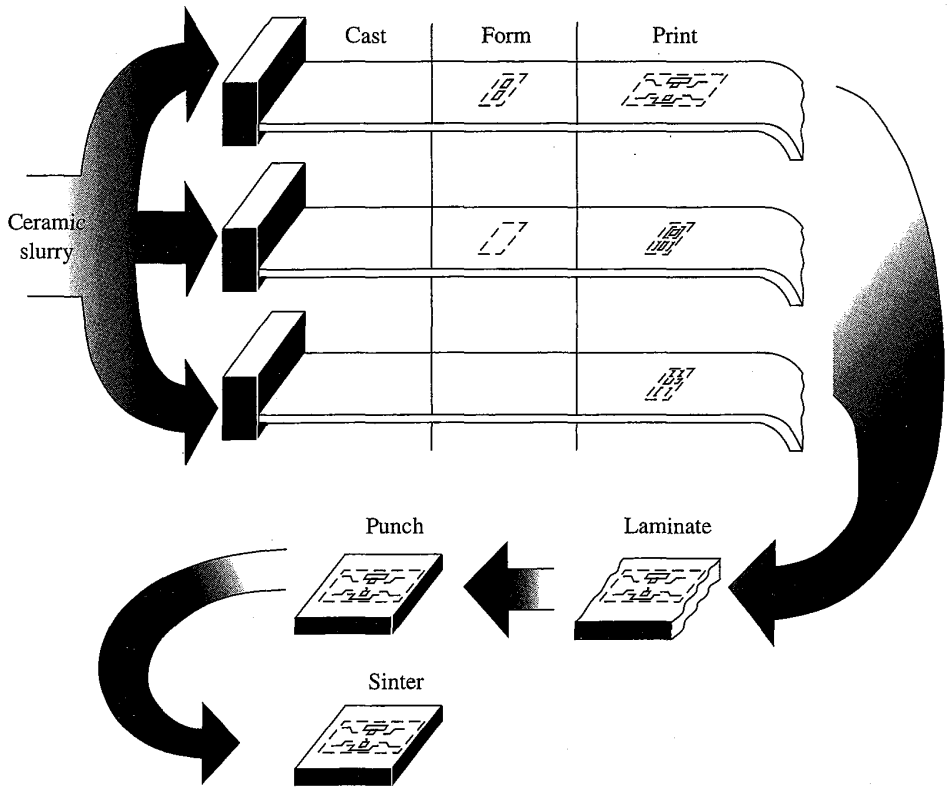


FIGURE 19
Process sequence to create a laminated refractory-ceramic product starting from a ceramic slurry. (*After Gardner and Nufer, Ref. 21.*)

however, three drawbacks: hard-to-control tolerances caused by high shrinkage during processing, high dielectric constant ($\epsilon_i/\epsilon_0 = 9.5$), and the modest conductivity of alumina. The tolerance problem makes it difficult to use edges as accurate references, and the high dielectric constant affects signal-line capacitive loading. Although they are very expensive, beryllia (BeO) or aluminum nitride (AlN) modules could be used instead of Al_2O_3 modules to produce greatly superior thermal performance and a significantly lower dielectric constant.

Ceramic packages can have brazed pins or leads and can have edge or array pinouts. The package designer has the option of locating the chip cavity either on the side with the pins or on the side away from the pins. Devices requiring good thermal-dissipation characteristics are packaged with the cavity on the leaded side. This provides a direct thermal path to the surface (facing away from the PWB), which is exposed to the cooling air flow. In this arrangement, the heat transfer can be enhanced by attaching a heat fin to the package. The disadvantage of this arrangement is increased package size, since the pins cannot be placed at the cavity area. In contrast, when thermal dissipation is comparatively small, the cavity can be located on the opposite side, providing a smaller outer package size. Figure 20 shows an example of refractory multilayer-ceramic packages for ULSI use. After chip interconnections are formed, hermeticity is completed by lidding operations, in which a metal or ceramic cap is attached with glasses or Au-Sn eutectic materials.

Micron Ex. 1029, p. 582
Micron v. YMTC
IPR2024-00911

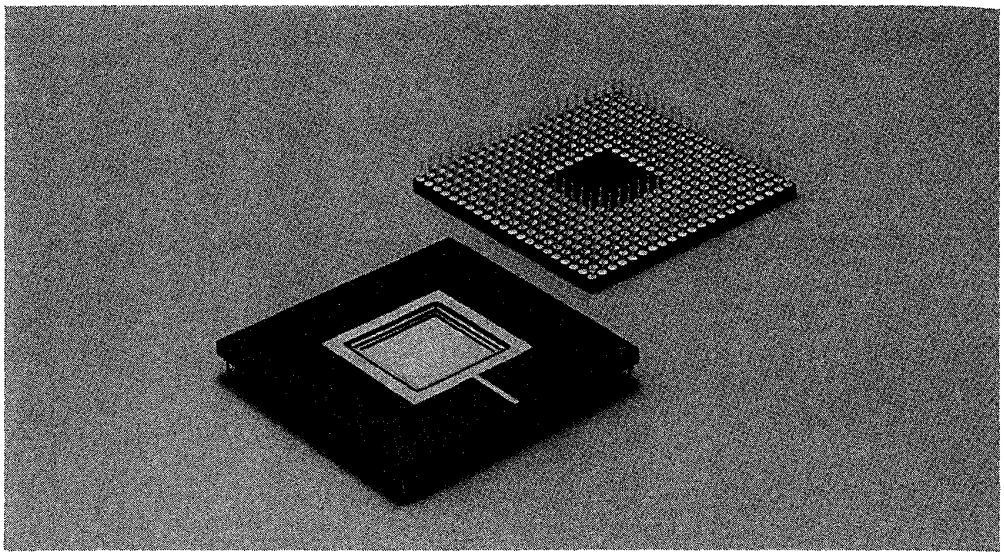


FIGURE 20
Refractory multilayer-ceramic pin-grid-array (PGA) package. (*Courtesy of Mitsubishi Electric Corp.*)

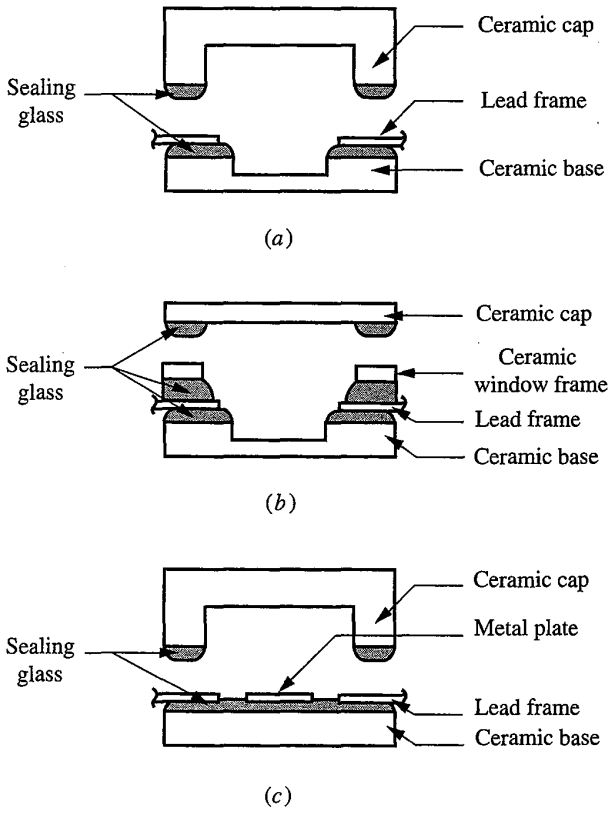


FIGURE 21
Structure of CERDIP and CERQUAD (or "quad CERPAC"). (*a*) Construction details for the standard CERDIP structure consisting of a base, a cap, a lead frame, and a glass-frit seal. (*b*) Construction details for a package consisting of a base, lead frame, window frame subassembly, and a ceramic cap using polymer or soldering sealing. (*c*) Construction details for a CERQUAD consisting of a base, a planar lead frame with an integral chip pad, a domed ceramic cap, and a glass-frit seal. This arrangement provides a good TCE match to the chip and a simpler means of providing an electrical connection from the chip pad to an external I/O lead. (*Courtesy of Kyocera International, Inc.*)

Micron Ex. 1029, p. 583
Micron v. YMTC
IPR2024-00911

10.4.2 Glass-Sealed Refractory Technology

Figure 21 illustrates a low-cost ceramic technology. This technology relies on glass-sealing a lead frame (after chip interconnections are formed) between two pressed ceramic units using low-temperature glass. The glasses used for glass-sealing are PbO-ZnO-B_2O_3 types. Both crystallizing and noncrystallizing glasses are used. Sealing is usually performed above 400°C in an oxidizing ambient to avoid de-oxidizing the metallic components of the glass and, consequently, degrading the electrical insulations. Since Au-Al intermetallic growth would be severe at these temperatures, Al bonding wire is used. The lead frames, typically Alloy 42, usually have a strip of Al deposited or clad on their lead tips so that an all-aluminum bonding system is possible. The hermetically glass-sealed package has the potential of being automated, and it competes with plastic technology for low-cost packaging. But in ULSI packaging, where delicate submicrometer chips are concerned, this glass-sealing technology must be used carefully. The high sealing temperature, above 400°C, would cause additional thermal diffusions at the transistor junctions, so that the electrical characteristics of the chips might shift slightly.

Both TH types (CERDIPs) and SM types (CERQUADs) are available. The laminated ceramic technology described in Section 10.4.1 can provide DIP and quad packages, but the terms CERDIP and CERQUAD specifically apply to glass-sealed refractory packaging.

10.4.3 Plastic Molding Technology

Plastic encapsulation involves a variety of techniques. In glob-top-coating, for example, the post-bonding chip is coated with liquid plastic resin and the plastics are cured for cross linking.

In ULSI plastic packaging, a premolding technique is sometimes used. This technology is the plastic equivalent of the refractory ceramic-cavity package. First the package is molded together with a lead frame, forming a plastic body and a cavity equivalent; then the chip and interconnects are added. The premold package has considerable future potential, but in ULSI packaging, postmold technology is still dominant. Only the postmold technology is discussed here. The postmolding technology is a transfer molding method using thermosetting (cross-linking) epoxy resins to mold around the lead frame-chip assembly, forming the package body. This process is relatively harsh, because the chip and its wire bonds are exposed to viscous molding material. However, improvements in molding materials, in molding process control, and in overall integrated automation of the assembly process, especially the automation of the chip-bonding, wire-bonding, and molding steps, have established the reliability to meet ULSI device objectives at a lower cost.

Molding material

The polymers most commonly used for IC packaging are epoxies. The original epoxy resin used to mold ICs was made by condensing epichlorohydrin with bisphenol-A to produce a material called Epoxy-A. An excess of epichlorohydrin

Micron Ex. 1029, p. 584
Micron v. YMTC
IPR2024-00911

was used to leave epoxy groups on each end of the low-molecular-weight polymer. Today, NOVOLAC epoxies are generally preferred; their higher functionality makes them heat resistant because each repeating group contains an epoxy group. These resins, called Epoxy-B, are made by reacting epichlorohydrin with NOVOLAC phenolic resin and a base. The synthesis of the resins produces NaCl as a by-product. Both Na and Cl ions reduce device reliability, so these by-products must be carefully washed from the resins before they are mixed into molding compounds. For ULSI plastic packages, lower-molecular-weight epoxy resins like biphenyl are used in addition to NOVOLAC epoxies, because they reduce moisture absorption from the ambient. Molding compounds are mixtures of epoxy resin, fillers, small amounts of pigments, mold-release agents, antioxidants, plasticizers, and flame retarders, usually formed into tablet shape. The selections and the ratios of these materials have significantly changed over the past few years[22, 23] to meet the challenges of ULSI devices. These changes greatly affect the rheological, chemical, and thermophysical properties of the molding compounds, and they improve the moldability, the in-process yield, and the reliability of packaged chips.

Thermal coefficient of expansion (TCE) is the most important feature of the molding compound. Fillers, usually amorphous or crystalline SiO_2, added to the resin decrease the resultant TCE of the compound. In ULSI packaging, higher filler loading, close to 80% by weight, reduces α_1 (TCE below T_g) of the compound to as low as 10 to 17 ppm/°C, providing better TCE matching to the lead frame (α = 17ppm/°C in Cu alloys and 5 ppm/°C in Alloy 42) and to the chip (α = 4 ppm/°C). Highly loaded fillers improve the mechanical strength and the thermal conductivity of the resin; however, they also increase the elastic modulus (E) and cause poor moldability as side effects. Because a larger E results in a larger stress in the package, the TCE is reduced by modifying the epoxy resin in two ways—adding a flexible material and changing the resin chemistry. Sometimes a hardener is added. Silicone rubber mixed with epoxy resin forms a "sea-island" structure; this has been widely used, as well as new, lower-molecular-weight and more flexible chemistry, such as biphenyl. Reducing the moisture-absorption ratio of the molding compound is essential in SM packages; this subject is detailed in Section 10.5.

ULSI memory, logic, and microprocessor devices challenge designers to formulate newer molding compounds that reduce TCE and E, reduce moisture absorption, improve adhesion strength to the chip and the lead frame, improve moldability, and offer various physical and chemical features from the increasing variations in plastic packages.

Molding process

Thermoset molding materials, that is, epoxy compounds, are usually transfer-molded in large multicavity molds. After it enters the pot, the molding compound is usually preheated, melts under pressure and heat, and flows to fill the mold cavities containing lead frame strips with their attached chips. The lead frame often has long, fragile fingers, and the chip is interconnected to these thin leads with 25- to 30-μm diameter Au wire. To avoid damaging this fragile structure, the viscosity and velocity of the molding compound must fall within certain ranges. Commercial molding compounds are designed to meet these requirements when molded at approximately

Micron Ex. 1029, p. 585
Micron v. YMTC
IPR2024-00911

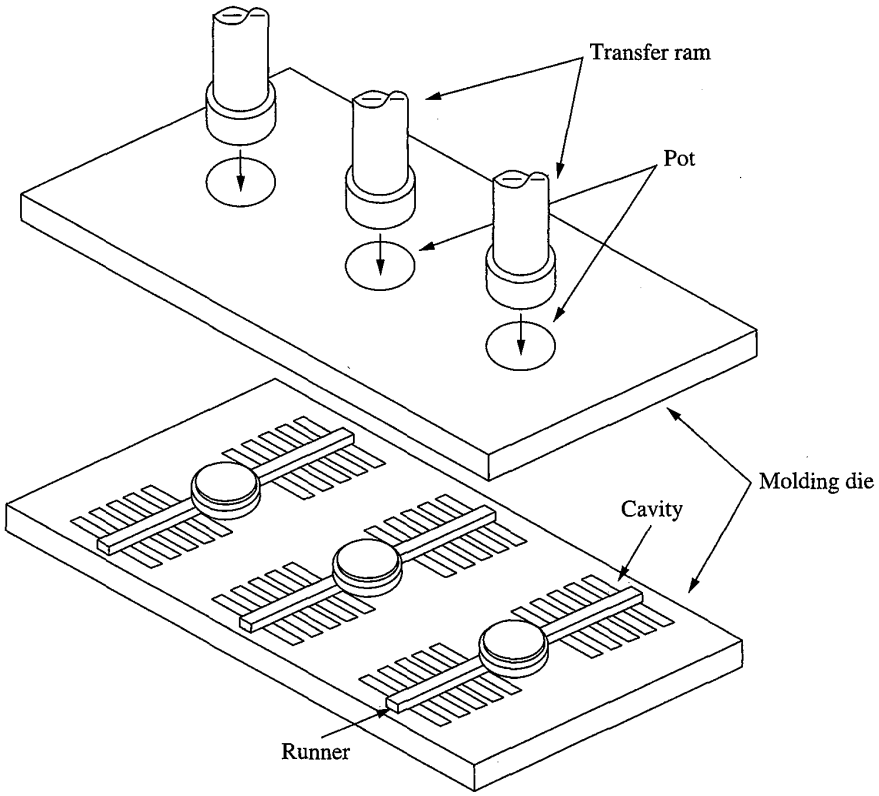


FIGURE 22
Schematic of a multipot transfer-mold system featuring plural ram, pot, and runner sets. Small mold-compound tablets, each large enough to fill a few cavities, are used in this arrangement; these provide better process control compared to a single-pot system.

175°C and at pressures of about 6 MPa, while the upper and lower molding dies are pressed at pressures of about 18 MPa to avoid resin flashes. To control the velocity of the molten molding compound, each device cavity has a gate to slow the material flow. The gate also influences the viscosity of the fluid; the mechanics of this fluid are relatively complex because the materials are non-Newtonian. In addition, partial cross-linking can occur during the molding process, which affects the material viscosity.

For molding ULSI devices, which are usually more difficult because of the larger package sizes and longer Au wire lengths, the multipot approach (see Fig. 22) is preferred because it provides more stable process control and operates at cycles ranging from 1 to 5 minutes. Products usually have insufficient cross-links after the molding operation and require post-cure storage at approximately 175°C for several hours.

10.5 PACKAGE DESIGN CONSIDERATIONS

In this section, we discuss some basic package design features. The package designer must simultaneously realize the electrical, thermal, and mechanical performances

Micron Ex. 1029, p. 586
Micron v. YMTC
IPR2024-00911

required in one specific package, make engineering trade-offs, and consider the manufacturing process capabilities described in Sections 10.3 and 10.4.

10.5.1 Electrical Design Considerations

As the speed of ULSI devices increases and their noise margins decrease, electrical design for packages must be carefully considered. A poorly designed package may spoil all other efforts, such as enhancing drivability of MOS transistors, increasing switching speed by reducing gate and line capacitances, lowering voltage interface at the chip level, and processing pipeline or parallel signals at the system level. Packages with a design geometry much larger than that of silicon can significantly affect the electrical performance of the packaged chips. Several electrical performance criteria are important: minimum signal delay, signal-reflection control, and noise reduction, including simultaneous switching noise and cross talk. These criteria, often discussed in PWB design, also apply to the package. They are mutually dependent and require trade-offs, since they are related through simple package-geometry variables.

Signal delay

Signal delay time, t_d, is given by

$$t_d = \frac{l}{v} = \frac{l}{c/\sqrt{\epsilon_r}} \tag{10.5}$$

where l (m) is the signal line length, v (m/sec) is the velocity of signal, c (m/sec) is the velocity of light, and ϵ_r is the dielectric constant of the surrounding material.

High-speed operation requires smaller t_d. The ratio of t_d to the cycle time usually dominates the system performance. In package construction, a short signal line (bonding wire length plus lead length) in small dielectric material, typically polyimide resin, is preferable. Table 3 lists the dielectric constants of common packaging materials. An excessively small dielectric-constant value of the surrounding

TABLE 3
Dielectric constants of packaging materials

Material	Dielectric constant (at 1 MHz)
Al_2O_3	9.6–10.2
AlN	8.7
Mold compounds	3.9–4.3
Polyimide	3.5
Si	11.7
GaAs	12.9
SiO_2	3.8
Glass ceramic	3.9–7.8
Glass epoxy	4.2

Micron Ex. 1029, p. 587
Micron v. YMTC
IPR2024-00911

material, however, induces signal reflections that degrade operating speed; hence, an optimum dielectric value exists.

Signal reflection

A mismatched impedance causes signal reflections when a signal is transmitted from a driver to a receiver through a transmission line. In CMOS ULSI devices, multiple reflections occur at the driver and receiver ends when the output impedance of the output buffer is smaller than that of the transmission line. These reflections cause a ringing phenomenon that may slow operation or cause the circuit to malfunction.[24] These reflections cannot be treated lightly when the following relationships exist.

$$l > \frac{c}{\sqrt{\epsilon_r} \cdot \upsilon_0} \tag{10.6}$$

$$\upsilon_0 = \frac{0.35}{t_r} \tag{10.7}$$

where l (m) is the signal line length, c (m/sec) is the velocity of the light, ϵ_r is the dielectric constant of the surrounding material, υ_0 (1/sec) is the critical frequency, and t_r (sec) is the signal rise (or fall) time.

The equations show that a shorter signal line is beneficial to high-speed operation requiring a smaller package size. Larger package constructions that have longer signal lines can no longer be dealt with as lumped-element circuits but must be considered distributed-element circuits. In particular, longer wires, longer via-hole connects or both, which have larger impedance-mismatch potentials, should be avoided if possible, or matched-impedance designs should be used instead. Leading-edge packages, multilead frame plastic packages, and some other multilayered packages that contain strip, microstrip, or coplanar constructions provide better impedance matching. Figure 23 illustrates the strip, microstrip, and coplanar structures. The characteristic impedances, $Z_0(\Omega)$, of strip and microstrip structures are expressed by[25]

$$Z_0 = \frac{60}{\sqrt{\epsilon_r}} \left[\ln \left(\frac{4b}{0.67\pi W(0.8 + t/W)} \right) \right] \quad \text{(strip line)} \tag{10.8}$$

$$Z_0 = \frac{87}{\sqrt{\epsilon_r + 1.41}} \left[\ln \left(\frac{5.98h}{0.8W + t} \right) \right] \quad \text{(microstrip line)} \tag{10.9}$$

where ϵ_r is the dielectric constant of the dielectric material, t (m) and W (m) are the thickness and width of the conductor, and h (m) and b (m) are the thicknesses of the dielectric material beneath and surrounding the conductor, respectively.

For a typical coplanar structure such as a symmetrical double-strip, coplanar waveguide, the characteristic impedance, $Z_0(\Omega)$, is expressed by[26]

$$Z_0 = \frac{\eta_0}{\pi \sqrt{(\epsilon_r + 1)/2}} \left[\ln \left(2 \cdot \frac{1 + \sqrt{S/(2W + S)}}{1 - \sqrt{S/(2W + S)}} \right) \right] \quad \text{(coplanar line)} \tag{10.10}$$

where η_0 (equal to $120\pi\ \Omega$) is the characteristic impedance of free space, S (m) is the gap between the two lines, and W (m) is the width of the line. Assumptions are

Micron Ex. 1029, p. 588
Micron v. YMTC
IPR2024-00911

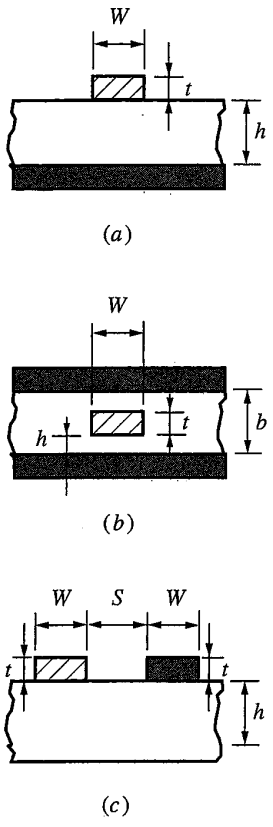


FIGURE 23
Cross-sectional sketches of several structures for impedance control: (*a*) microstrip line, (*b*) strip line (*after Kaupp, Ref. 25*), and (*c*) coplanar structure.

that the conductor lines are infinitely thin ($t = 0$), the dielectric material is infinitely thick ($h = \infty$), and the gap widths range is $0.173 \leq S/(2W + S) < 1$.

Noise

Two typical types of noise, cross-talk noise and simultaneous switching noise (ΔI noise), are discussed here.

Cross-talk noise (see Fig. 24*a*) occurs when a line is undesirably affected by another line that is placed very close to it because of the electromagnetic coupling between the two lines. The noise, coupled by C_m, L_m, or both (m: mutual) between the two lines, increases in proportion to the signal-voltage or current gradient and the AC-coupling strength. Cross-talk noise is a more serious problem in ULSI packaging intended to handle higher speeds, larger signal counts, and the resulting narrow signal-line spacings. Major countermeasures in package design are shorter parallel-signal runs, closer ground (or power) planes, and lower dielectric-constant materials.

Simultaneous switching noise, one of the most practical electrical-design problems, particularly in CMOS ASIC devices, occurs when many output buffers switch simultaneously. Figure 24*b* illustrates the mechanism. When an output buffer switches from high to low, transition current (i) flows from the power line (V_{cc}) into the load capacitance (C_l), inducing the noise voltage given by

$$V_n = L_g \frac{di}{dt} \tag{10.11}$$

Micron Ex. 1029, p. 589
Micron v. YMTC
IPR2024-00911

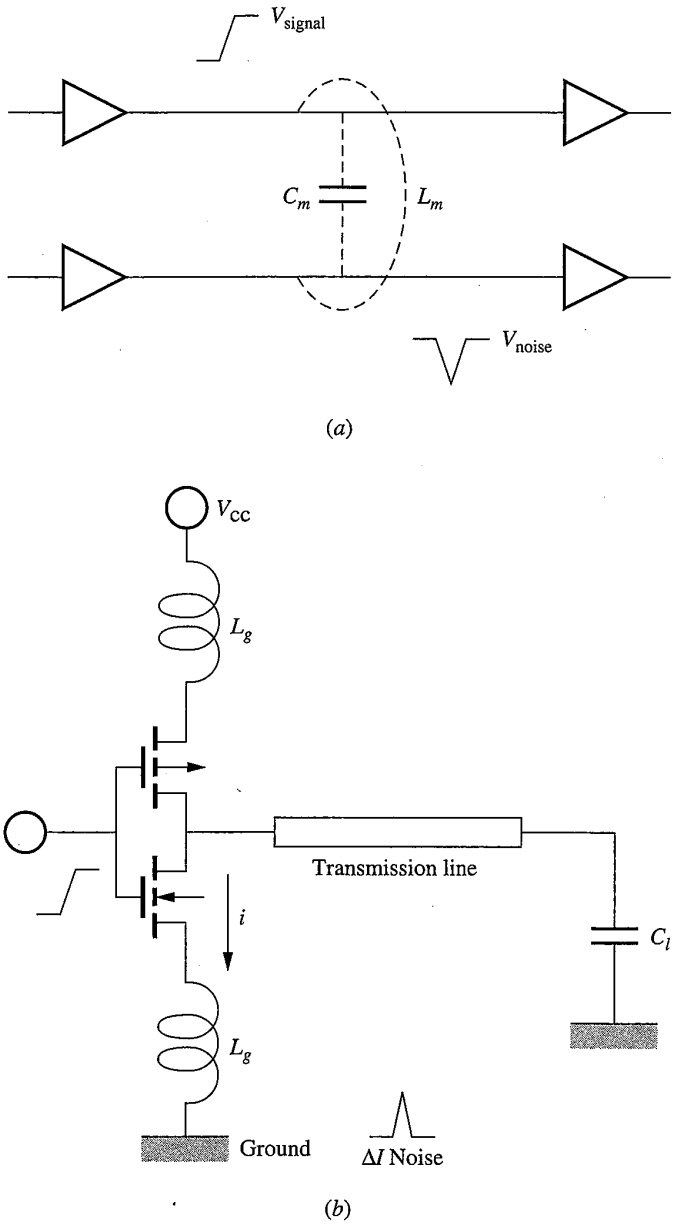


FIGURE 24
(*a*) Cross-talk noise that appears in one of the two adjacent lines, and (*b*) simultaneous switching (ΔI) noise that occurs when the output buffers switch simultaneously.

where V_n (V) is the induced voltage, L_g (H) is the inductance of the power lead, and di/dt is the derivative of current with respect to time.

In addition, when a line switches from low to high, an electric charge stored in the load capacitance flows into the ground line through the transmission line, inducing the same noise voltage as shown in Eq. (10.11). If j lines are switching simultaneously, then V_n is given by

Micron Ex. 1029, p. 590
Micron v. YMTC
IPR2024-00911

$$V_n = \sum_j L_{g_j} \frac{di_j}{dt} \tag{10.12}$$

This noise voltage, V_n, leads to misoperation of other signal lines because it causes a bounce in the ground level. To reduce V_n, the L_g of the package must be reduced.[27] In actual package design, mutual inductance, that is, interaction between a line and the adjacent line, plays a significant role in reducing the L_g. But here we discuss only self-inductance, L, for simplification.

The Biot–Savart law derived from Maxwell's fundamental equation gives the lead inductance value L for a simplified columnar conductor model, assuming $l \gg a$:

$$L = \frac{l}{2\pi}\left(\frac{\mu}{4} + \mu_0\left[\ln\left(\frac{2l}{a}\right) - 1\right]\right) \tag{10.13}$$

$$= L_i + L_e \tag{10.14}$$

where l (m) is the length, a (m) is the radius, μ (kg-m/C^2) is the magnetic permeability of the conductor, and μ_0 (kg-m/C^2) is the permeability of a vaccum. L is the sum of L_i, the internal inductance, which depends on the internal magnetic flux, and L_e, the external inductance, which depends on the external magnetic flux. Either L_i or L_e, or both, if possible, must be smaller to suppress ΔI noise. Cu alloy reduces L_i considerably, because it has a smaller μ value, whereas shorter and wider geometric dimensions of the lead reduce L_e. Multiple grounds control the noise effectively, since, if m ground leads are used, the total inductance becomes approximately L_g/m. However, multiple grounds have a large impact on packaging density. Inductance can be reduced significantly by using large-area power and ground planes within the package. Figure 25 illustrates an example of a multilayer lead frame[28, 29, 30] used to provide the planes.

The electromagnetic-field simulator has become a powerful analytical tool for predicting the impact of package design on the various kinds of noise described above. Through the simulation, electrical parameters such as inductances, capacitances, resistances, and impedances of the package portions are calculated from the geometric dimensions of the package portions and the electrical properties of the components. These parameters allow a description of the package as an equivalent circuit and are used to perform overall packaged-chip circuit simulations, typically using Simulation Program with Integrated Circuit Emphasis (SPICE).

10.5.2 Thermal Design Considerations

The objective of thermal design is to keep the operating junction temperature of a silicon chip low enough to minimize the failure rate caused by temperature-activated failure mechanisms. The design must prevent the chip from exceeding the acceptable thermal limit for a particular application.[31] Only in the simplest possible applications can this be done by considering the packaged silicon device alone. Usually the packaged-device environment must be established for most of the following variables[32]: PWB temperature, total power dissipation on the board, local-neighbor power dissipation, degree of forced-air cooling, lateral and vertical space

Micron Ex. 1029, p. 591
Micron v. YMTC
IPR2024-00911

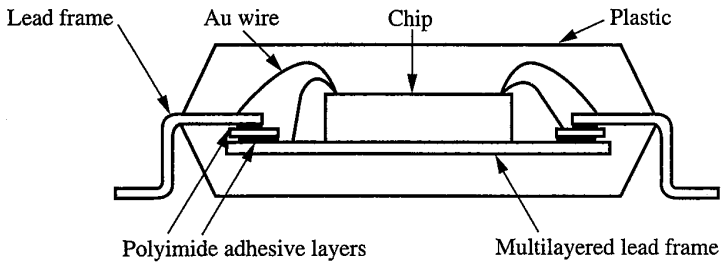


FIGURE 25
Multilayer lead frame package that provides large-area power and ground planes. The ground plane, which also serves as the chip support paddle, and the power plane are attached to the lead frame with electrically insulating adhesives. Many bonds are made from pads on the chip to these planes.

between boards that can be used by the package, conductivity of the PWB, and ideal performance of the isolated package. The thermal modeling of a single chip in a package is discussed below.

We start from the definition of thermal resistance. When heat P_{XY} (watts) flows from point X to point Y, the thermal resistance R_{XY} (°C/watt) between the two points is given by

$$R_{XY} = \frac{(T_X - T_Y)}{P_{XY}} \tag{10.15}$$

where T_X, and T_Y, in °C, are the temperatures at points X and Y and $T_X > T_Y$. Similarly, thermal resistance of a package (see Fig. 26) is expressed as

$$\theta_{ja} = \frac{(T_j - T_a)}{P} \tag{10.16}$$

where θ_{ja} (°C/W) is the junction-to-ambient thermal resistance, T_j (°C) is the average chip or junction temperature, T_a (°C) is the ambient temperature, and P (watts) is the power.

Figure 26 is a simplified heat-transfer model of a packaged chip, where heat is transferred from the chip to the surface of the package by conduction and from the

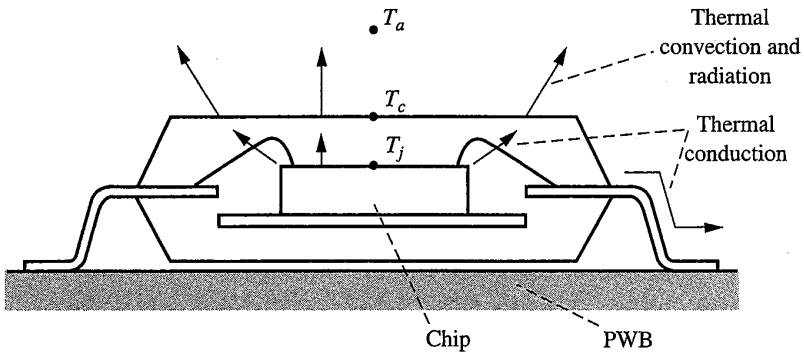


FIGURE 26
A simplified heat-transfer model of a packaged chip.

Micron Ex. 1029, p. 592
Micron v. YMTC
IPR2024-00911

package surface to the ambient by convection and radiation. In most applications, the temperature difference between the case (or the package surface) and the ambient is small, and radiation can be neglected. Conduction heat transfer through the package terminals (see Fig. 26) can be significant, particularly in high-I/O ULSI packages. However, if we neglect it for simplification, the overall thermal resistance in this model can be considered as the sum of two thermal components, θ_{jc} and θ_{ca}, defined as

$$\begin{aligned}\theta_{ja} &= \theta_{jc} + \theta_{ca} \\ &= \frac{(T_j - T_c)}{P} + \frac{(T_c - T_a)}{P}\end{aligned} \tag{10.17}$$

where θ_{jc} (°C/watt) is the junction-to-case thermal resistance, θ_{ca} (°C/watt) is the case-to-ambient thermal resistance, and T_c(°C) is the average case temperature.

θ_{jc} is relatively insensitive to the ambient and is mainly a function of package materials and geometry. θ_{ca} depends on the package geometry, the package orientation in the application, and the conditions of the ambient in the operating environment (whether heat transfer is free or by forced-convection). Heat transfer is classified into three categories—conduction, convection, and radiation. How these work in ULSI packages is summarized below.

Conduction

Conduction dominates the heat transfer from the chip to the package surface. Fourier's equation at one dimension and at a stationary state is given by

$$Q = (T_1 - T_2) \cdot \kappa \cdot \frac{S}{L} \tag{10.18}$$

where Q (watts) is the amount of heat flow, T_1 and T_2 (°C) are the temperatures at the two ends of the thermal path, κ (watts/m-K) is the thermal conductivity, S (m^2) is the cross-sectional area, and L (m) is the length of the insulated heat flow path.

In actual package, Eq. (10.18) can be written as

$$P = (T_j - T_c) \cdot \kappa \cdot \frac{S}{L} \tag{10.19}$$

So, we have

$$\theta_{jc} = \frac{(T_j - T_c)}{P} = \frac{L}{(\kappa \cdot S)} \tag{10.20}$$

ULSI packages have a high packaging density (i.e., smaller S), so high-thermal-conductivity components such as Cu alloy lead frames, AlN substrates, and thermo-conductive molding compounds are particularly important because they increase the overall package κ value. Thinner package constructions, which reduce the L value, are also important.

Convection

Heat transfer from the package surface to the ambient results mostly from convection, given by Newton's cooling law,

$$Q = h \cdot A \cdot (T_c - T_a) \tag{10.21}$$

Micron Ex. 1029, p. 593
Micron v. YMTC
IPR2024-00911

where Q (watts) is the amount of heat flow, h (W/m^2-K) is the heat transfer coefficient, A (m^2) is the surface area, and T_c and T_a (°C) are the temperatures of the surface and the ambient.

From Eqs. (10.17) and (10.21), we have

$$\theta_{ca} = \frac{(T_c - T_a)}{Q} = \frac{1}{(h \cdot A)} \tag{10.22}$$

θ_{ca} is reduced through increased conduction[33] and larger package surface area in ULSI packaging. The application system constructions are forced-air convections, liquid coolants in place of air coolings, and additional heat sinks attached to the package surface.

Radiation

Radiation helps transfer some heat from the package surface to the ambient, but usually the contribution is small. Blackbody radiation follows the Stefan–Boltzmann law

$$E_b = \sigma \cdot T^4 \tag{10.23}$$

where E_b (W/m^2) is radiation energy flux density, σ is Stefan–Boltzmann's constant (5.67×10^{-8} W/m^2-K^4), and T (K) is the absolute temperature.

Since actual materials are not perfectly black, actual radiation E_1 becomes

$$E_1 = \epsilon \cdot E_b = \sigma \cdot \epsilon \cdot T^4 \tag{10.24}$$

where ϵ is the emissivity.

Heat transfer caused by radiation between two plates, that is, between a package of temperature T_1 and emissivity ϵ_1, and an ambient of temperature T_2 and emissivity ϵ_2, is expressed as

$$Q = \sigma \cdot f \cdot A(T_1^4 - T_2^4) \tag{10.25}$$

where Q (watts) is the amount of heat flow, A (m^2) is the package surface area, and f is expressed by

$$f = \frac{1}{1/\epsilon_1 + 1/\epsilon_2 - 1} \tag{10.26}$$

When $T_1 - T_2 \ll T_1$ and $T_1 - T_2 \ll T_2$, Q could be expressed as follows, by placing $(T_1 + T_2)/2 = T_m$

$$\begin{aligned} Q &= 4\sigma \cdot f \cdot T_m^3 \cdot A(T_1 - T_2) \\ &= hr \cdot A(T_1 - T_2) \end{aligned} \tag{10.27}$$

where $hr = 4\sigma \cdot f \cdot T_m^3$ and is the radiation transfer coefficient.

Therefore, the thermal resistance of the radiation is

$$\theta_{\text{rad}} = \frac{(T_1 - T_2)}{Q} = \frac{1}{hr \cdot A} \tag{10.28}$$

In actual applications, black-dyed packages and external heat sinks are sometimes preferred, since they increase hr values.

Micron Ex. 1029, p. 594
Micron v. YMTC
IPR2024-00911

10.5.3 Mechanical Design Considerations

In this section, we discuss mechanical design issues. We focus on plastic packages because TCE mismatches result in more complex problems in packages where the molding compound comes into contact with the chip and the lead frame surfaces.

Thermal stress of plastic mold

Figure 27 shows a simplified plastic package model in which the silicon chip and the lead frame are assumed to be structurally homogeneous, having the same TCE. When the construction is held at a temperature T (°C), the compressive stresses (σ_c) in the plastic resin and in the chip (including the lead frame) are respectively given by

$$\begin{aligned} \sigma_{cr} &= E_r \frac{\lambda - \alpha_r \cdot \Delta T \cdot L}{L} \\ \sigma_{cc} &= E_c \frac{\lambda - \alpha_c \cdot \Delta T \cdot L}{L} \end{aligned} \tag{10.29}$$

where E (GPa) is the elastic modulus, λ (m) is the elongation of the construction, ΔT is the temperature difference between the temperature T and that of the molding environment, L (m) is the package length, and the subscripts r and c correspond to resin and chip.

From Eq. (10.29), we see that

$$\frac{\sigma_{cr}}{E_r} + \alpha_r \cdot \Delta T = \frac{\sigma_{cc}}{E_c} + \sigma_{cc} \cdot \Delta T \tag{10.30}$$

Also, from equilibrium of the forces, we find that

$$\sigma_{cr} \cdot A_r + \sigma_{cc} \cdot A_c = 0 \tag{10.31}$$

where A (m^2) is the lateral cross-sectional area of the package.

By combining Eq. (10.30) and Eq. (10.31), stresses can be given by

$$\begin{aligned} \sigma_{cr} &= A_c \cdot E_r \cdot E_c \cdot (\alpha_r - \alpha_c) \cdot \frac{\Delta T}{(A_r \cdot E_r + A_c \cdot E_c)} \\ \sigma_{cc} &= A_r \cdot E_r \cdot E_c \cdot (\alpha_r - \alpha_c) \cdot \frac{\Delta T}{(A_r \cdot E_r + A_c \cdot E_c)} \end{aligned} \tag{10.32}$$

E_r and α_r are actually temperature-dependent functions; hence, Eq. (10.32) becomes

$$\sigma_{cc} = \int_{T_m}^{T} E_c \cdot \frac{\alpha_r(T) - \alpha_c}{1 + A_c E_c / A_r E_r(T)} \tag{10.33}$$

Since $\alpha_r(T) > \alpha_c$, it is apparent that the chip receives a compressive stress when the package is cooled from T_m, and it receives a tensile stress when heated. Stress reduction is essential because stress induces various kinds of defects such as device-parameter shift, chip cracking, Al-metallization slide, and Au wire fatigue. The latter results from relative dislocation between the wire and the plastic and might lead to wire breakage after temperature cycling. The countermeasures are to control

Micron Ex. 1029, p. 595
Micron v. YMTC
IPR2024-00911

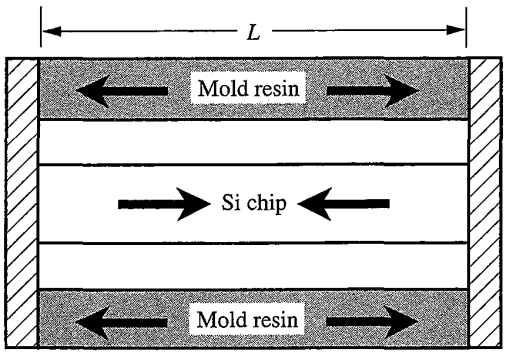


FIGURE 27
A simplified plastic package model, assuming that the silicon chip and the lead frame are structurally homogeneous. Arrows show the directions of stresses.

package-geometry dimensions and to carefully select package materials and process variables; choosing molding compounds that provide a smaller α_r and a reduced E_r is fundamental.

Moisture-induced cracking

Moisture-induced cracking resistivity[34] is another important issue, especially in thinner plastic package design. Moisture-induced cracking occurs when a moisture-saturated package is exposed to a reflow soldering temperature of approximately 215°C to 260°C.[35] Figure 28 illustrates the mechanism.[36] Absorbed water concentrates on the boundary surface between the chip pad and molding resin because the adhesion there is usually poor, and it flashes to steam when the package is reflow soldered. The steam pressure is sufficient to rupture thinner packages, because they have thin sections of plastic over and under the chip/lead frame surfaces. Recent studies[37, 38] on this phenomenon express the maximum stress of resin (σ_{max}) that occurs at the center in terms of the vapor pressure P by[39]

$$\sigma_{max} = k \cdot P \cdot a^{x_1}/h^{x_2} \tag{10.34}$$

where k is the proportional constant, a is the length of the shorter side of the chip pad, h is the resin thickness under the chip pad, and x_1 and x_2 are indices that depend on the package type.

Equation (10.34) shows that a smaller chip pad and thicker resin under the chip pad improve antimoisture capability. These measures, however, often jeopardize the ability to put larger ULSI chips in thinner packages. Therefore, additional measures are being implemented: improving molding resin (lower E and TCE), improving adhesion between the molding resin and the chip pad (dimples or holes in the chip pad), baking the package and packing it in a moisture-free container for shipment by the supplier and/or prebaking it before mounting by the user (reduced P), and lowering soldering temperatures at the next level of assembly operations (reduced P).

TCE mismatch between package and PCB[40]

When a ceramic package (typically TCE = 6 ppm/°C) is surface-mounted onto a plastic board (typically TCE = 15 ppm/°C), the TCE mismatch sometimes causes a large thermal stress on the lead-solder joints that could lead to joint breakage caused by solder fatigue from temperature cycling operations. A similar problem occurs in ULSI plastic packages, typically in a thinner small-outline package (SOP), with a chip occupancy as large as approximately 80% and an overall

Micron Ex. 1029, p. 596
Micron v. YMTC
IPR2024-00911

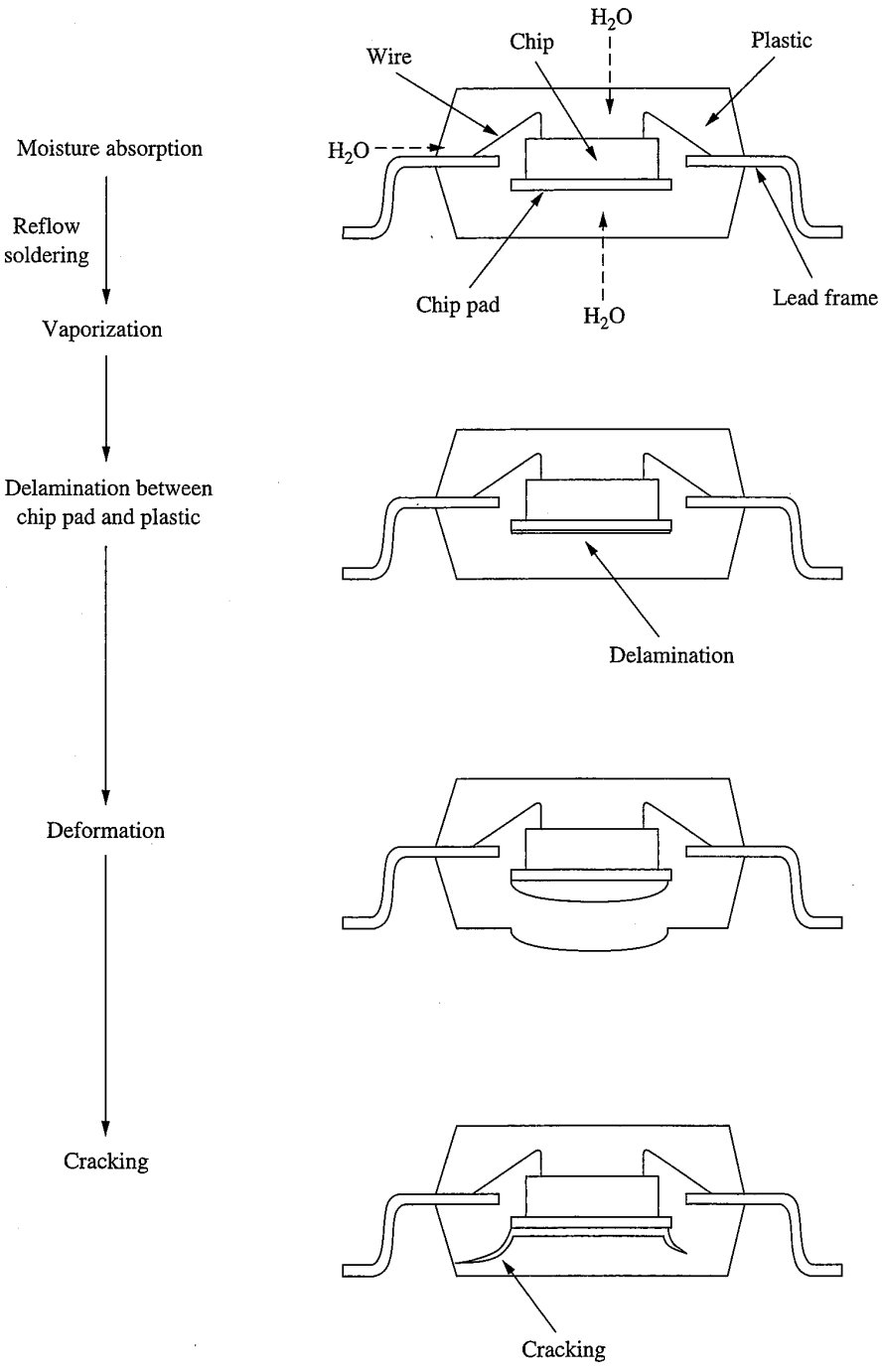


FIGURE 28
Mechanism of moisture-induced cracking. Plastic package absorbs moisture from the room ambient. The moisture flashes into steam during reflow soldering of the package to the PWB. Cracking resulting from the vapor pressure is apt to occur at the edge of chip support paddle.

Micron Ex. 1029, p. 597
Micron v. YMTC
IPR2024-00911

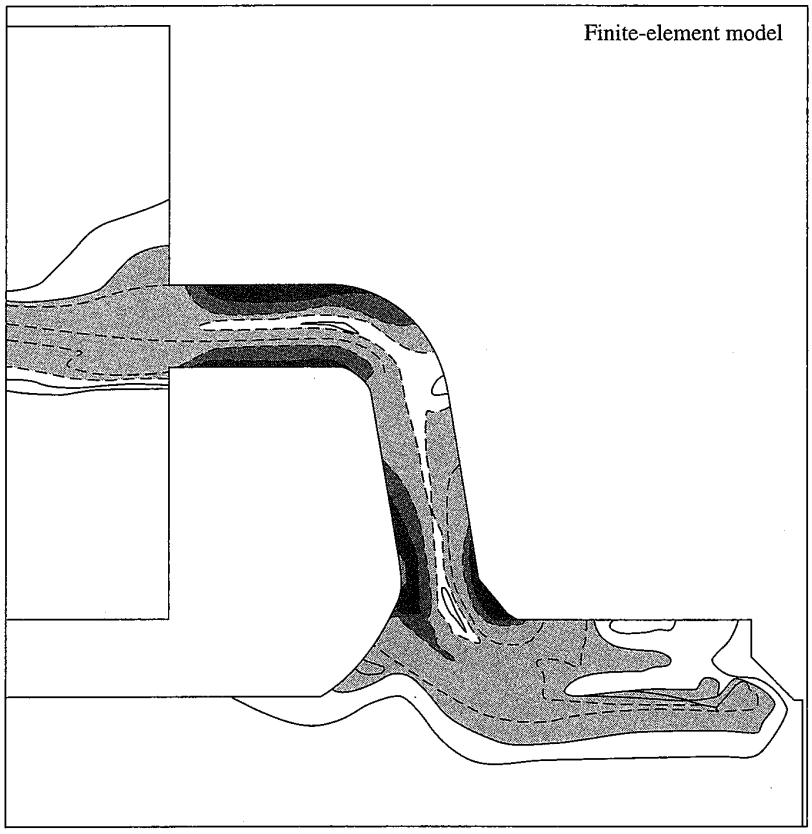


FIGURE 29
Finite element analysis (FEA) showing contours of constant stress in the plastic package lead and solder joint. The stress is greater where the shading is darker. (*Courtesy of Mitsubishi Electric Corp.*)

package TCE that is much closer to silicon's TCE (4 ppm/°C), than to that of the PCB. Solder fatigue[41, 42, 43] starts in the plastic deformation of the solder material and is followed by a tiny crack at the stress concentration point that develops along the crystal boundary into a breakage through accumulated temperature cycling operation. Finite-element analysis has become a powerful tool to characterize the thermal stresses on the solder joint, predicting[44] the stress concentration point and the location where the crack is likely to occur (Fig. 29).

Although higher and thinner leads using low-E materials have been successful[45, 46, 47] in settling the problem to some extent, improvements in both the single-chip and the next-level packagings are still required for a final solution.

10.6 SPECIAL PACKAGE CONSIDERATIONS

Various kinds of packaged-chip failures originate in the packaging. Figure 30 illustrates major package-induced failure modes for hermetic-ceramic and plastic packages. Table 4 lists the usual reliability tests performed to detect these failures.

Micron Ex. 1029, p. 598
Micron v. YMTC
IPR2024-00911

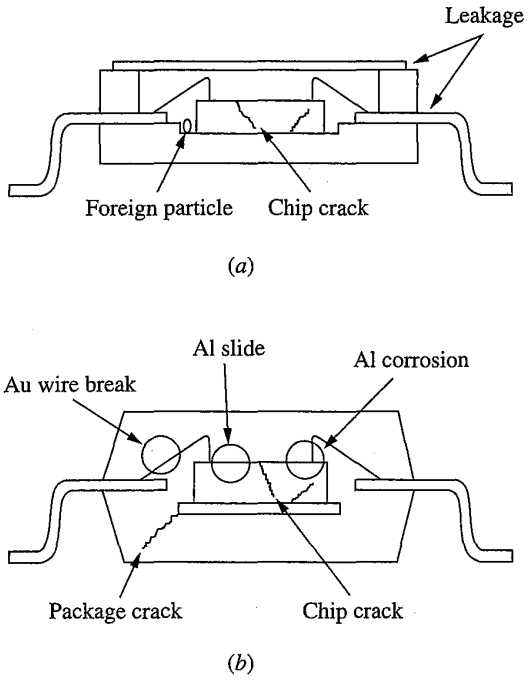


FIGURE 30
Major package-induced failure modes for (*a*) hermetic-ceramic and (*b*) plastic packages. Chip cracking from the back sometimes results from poor back grinding operation for both packages.

This section examines other aspects of packaging that must be addressed for a complete package-assembly methodology. Among them are the cleaning and assembly environment, package or lid sealing, corrosion, and α-particle protection.

10.6.1 Cleaning and Assembly Environment

The most critical cleaning step the chip undergoes is performed before bonding, encapsulation, or final lid sealing. This cleaning process must be chemically compatible with the chip metallurgy. Aluminum has a very narrow range of pH values in which its oxide protects it from corrosion in aqueous solution. Corrosion reactions, with metal dissolution, can occur in both basic and acidic solutions. There are two objectives in cleaning. One is to remove organic species that can affect bondability; an organic solvent is usually required for this. The other is to remove ionic species that can cause corrosion during the life of the device or, in an unusual instance, contribute to surface-charge accumulation. Water is a good solvent for ionic species.

Because environmental considerations have curtailed the use of powerful organic solvents, organic removal in the assembly process has become difficult. Today, new alternative chemicals are being used to some extent in place of chlorofluorocarbons. This difficulty and the ever-decreasing feature sizes on the ULSI chip, down to the submicrometer range, make environmental control in the assembly room a new issue that must be addressed. Typical assembly operations for ULSI devices are done in class 1000 to 10,000 cleanrooms. Otherwise, particles in the ambient that are not completely removed before molding (in plastic packages) or before lid seal-

Micron Ex. 1029, p. 599
Micron v. YMTC
IPR2024-00911

TABLE 4
Reliability tests performed to detect package-induced failures

Test type	Typical conditions	Failure mode	Major causes	Operation to note
Temperature cycling	−65°C–150°C	Chip crack	TCE mismatch of components	Chip bond
High-temperature storage	150°C, 175°C	Wire lift-off	Undesirable intermetallics	Wire bond
Temperature cycling	−65°C–150°C		Poor bonding	
			Bonding pad contamination	
Temperature cycling	−65°C–150°C	Wire break	Poor bonding	
Vibration			Stress from the molding resin	
High-temperature, high-humidity storage	85° C, 85%	Al corrosion	Chip contamination	
High-temperature, high-humidity storage with bias	85°C, 85%, 7 V		Seal leakage	Encapsulation
Gross leak			Impurities from the molding resin	(mold, seal)
Pressure-cooker test (PCT)	130°C, 85%			
Pressure-cooker test with bias	130°C, 85%, 7 V			
Operating life	125°C, 7 V	Malfunction	TCE mismatch of components (stress from the molding resin)	Chip bond through encapsulation

Micron Ex. 1029, p. 600
Micron v. YMTC
IPR2024-00911

ing (in ceramic packages) might degrade the reliability of the packaged device, most likely by hard failures caused by metal corrosion from particles with high levels of contaminants.

10.6.2 Package or Lid Sealing and Corrosion

The major objective of package sealing is to protect the device from external contaminants during its lifetime. Further, any contaminants present before sealing must be removed to an acceptable level before or during sealing. Ceramic packages may be lidded with polymers. However, since moisture usually penetrates organic sealants within a short period of time, almost all high-reliability applications have a hermetic seal made with glass or metal. Glass sealing for CERDIP is discussed in Section 10.4, and the process is essentially the same for lid sealing. Many of the metal alloys used for chip bonding are suitable for lid sealing. A leak-tight seal that excludes the external environment can be made without difficulty. The real difficulty has been freeing the package of contaminants, especially water, before sealing.[48, 49]

Problems can arise in a device in a low operating temperature, where water condenses inside the package. The following reactions can then lead to corrosion in the presence of a small halide contamination:

$$6HCl + 2Al \rightarrow 2AlCl_3 + 3H_2\uparrow$$

$$AlCl_3 + 3HOH \rightarrow Al(OH)_3 + 3HCl$$

$$2Al(OH)_3 + \text{Aging} \rightarrow Al_2O_3 + 3H_2O$$

Similar aluminum corrosion can occur in plastic packages, when water that penetrates the package condenses on the chip after a long period of time. In the presence of halide solubles from the molding compound, corrosion occurs particularly between the two adjacent and electrically biased lines:

$$Al + 4Cl^- \rightarrow AlCl_4^- + 3e^-$$

$$AlCl_4^- + 3H_2O \rightarrow Al(OH)_3 + 3H^+ + 4Cl^- \qquad \text{(Cathode line)}$$

$$H_2O + e^- \rightarrow \tfrac{1}{2}OH^- + H_2\uparrow$$

or

$$Na^+ + e^- \rightarrow Na, \qquad Na + H_2O \rightarrow Na^+ + OH^- + \tfrac{1}{2}H_2\uparrow$$

Then

$$Al + 3OH^- \rightarrow Al(OH)_3 + 3e^- \qquad \text{(Anode line)}$$

P_2O_5 in the chip passivation layer can accelerate the corrosion by forming H_3PO_5 in the presense of water[50]:

$$2Al + 6H^+ \rightarrow 2Al^{3+} + 3H_2\uparrow$$

$$2Al^{3+} + 3H_2O \rightarrow 2Al(OH)_3 + 6H^+$$

Micron Ex. 1029, p. 601
Micron v. YMTC
IPR2024-00911

10.6.3 α-Particle Protection

Soft errors in memory circuits caused by α-particles emanating from packaging materials were first reported[51] in 1978; there have been many papers on the subject since then. The α-particles are emitted by the decay of uranium and thorium atoms contained as impurities in the packaging materials. Decreasing device design rules make circuits sensitive to this problem. Packaging materials, particularly molding compounds, have been pushed below the level of 0.001 to 0.01 α-particle/cm^2-h. Because α-particles have low penetrating power in solids, low-α materials, typically polyimides, have been used as α-absorbing coatings on silicon chips.[52, 53]

Although great progress has been made by reducing α-particle emission in current packaging materials, error correction is still required to control soft errors caused by α-particles.

10.7 OTHER ULSI PACKAGES

Sections 10.1 through 10.6 discuss standard assembly and package fabrication technologies, and they examine some design features that are mainly on standard single-chip packages. Other packages and their assembly techniques are described in this section.

10.7.1 Tape Carrier Package

Tape carrier package (TCP) is a generic term for packages in which a tape-automated bonding (TAB) technique is used for the pads-on-chip interconnects. The TAB technique, which started in gang-bonding, was originally used to improve the productivity of interconnects. TAB uses finely patterned thin metal, usually Cu foil plated with Au or Sn, in place of wires and connects the metal tips metallurgically to corresponding bumps (plated Au) formed on the Al pads on the chip. TAB has become a choice in finer-pitch interconnects for high-I/O ULSI devices because it enables finer pitch and longer span bondings than those achievable by wire bonding. TCPs are usually plastic-based.

Tape carrier

Figure 31[54] illustrates typical tape carriers used in TCP fabrication. The one-layer tape carrier is a thin Cu foil. It is cost effective but not preferred today; electrical testing of the packaged chip is difficult since all the Cu leads are mechanically connected as one body.

Two-layer tape that has a Cu layer, typically 25 μm thick, on a polyimide base tape, typically 75 μm thick, is manufactured in two ways—either Cu plating on the polyimide or polyimide casting on the Cu foil. A window in which to place the chip is formed, and along with other necessary holes, by etching away the polyimide tape, and the Cu layer is patterned into the designated shape by plating or etching

Micron Ex. 1029, p. 602
Micron v. YMTC
IPR2024-00911

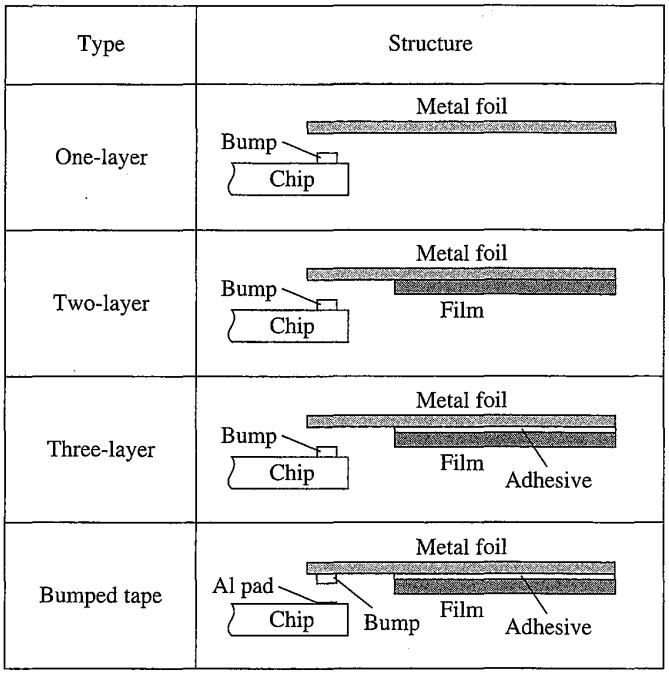


FIGURE 31
Cross section of typical tape carriers used in TCP fabrication. (*After Clain, Ref. 54.*)

operations. Three-layer tape is similar to two-layer tape but has an additional adhesive layer between the Cu and the base plastic layers, enabling a choice of base plastics, such as glass-epoxy, BT resin, polyester, and the most prevailing polyimide resin.

Two-layer and three-layer tapes, particularly those with (0.5 μm-thick plating) Cu foils that are chemically Sn-plated are widely used in ULSI packaging. Bumped tape, which has Au bumps on the Cu foil innerlead tips, is also accepted to some extent, since it does not require wafer-bumping operations. Figure 32[55] illustrates the transfer-bumping operation used as a typical bumped-tape manufacturing method. Au bumps are formed by first performing plating operations on a substrate (typically glass-based). The bumps are then transferred to the aligned tips of the Cu foil (Sn-plated) through a bonding operation to formulate Au-Sn eutectic alloys. Bumped tape eliminates bumped chip availability problems.

Tape-automated bonding

Tape-automated bonding (also called innerlead bonding) is divided into two methods: bonding to the bumps-on-chip using one- to three-layer tapes, and bonding to Al pads-on-chip using a bumped tape. The former method is more popular in the industry today, therefore only this method is detailed here. In this method, bumps (25-μm-thick Au) are formed on the Al pads-on-chip and are attached to Sn-plated or Au-plated Cu foil tips.

Micron Ex. 1029, p. 603
Micron v. YMTC
IPR2024-00911

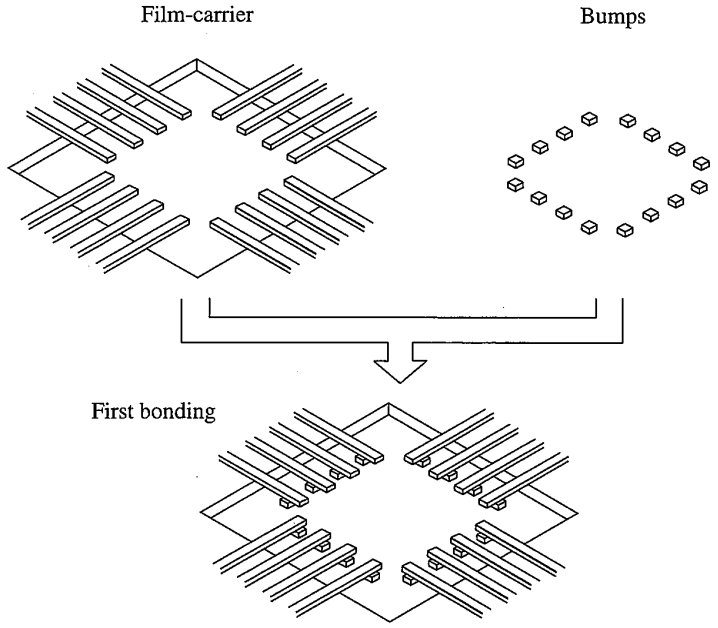


FIGURE 32
Illustration of the transfer-bumping operation. (*After Hatada, Ref. 55.*)

Figure 33[56] illustrates an example of the Au-bump-formation process that is performed at the wafer stage. First, a wafer covered with passivation glass except for the bond-pad areas is full-face sputtered by a combination of barrier metals such as Cr/Cu/Au, Ti/Ni/Au, or TiW/Au. Next, photolithography, in which 25- to 30-μm-thick photoresist materials are used, forms open windows at the bond pads where 20- to 25-μm-thick Au electroplating is performed, forming the Au bumps. Photoresist removal operations, and the following barrier-metal-etching operations, in which the formed Au bumps serve as masks to protect the barrier metals beneath the bumps, complete the bump formation.

Tape-automated bonding is usually a gang operation that connects all the bumps and corresponding foil tips at one time. Typical process conditions are thermocompression using heated (approximately 500°C) flat tools made of diamond and applying loads (approximately 10 kg/mm^2) for periods of time that range from 1 to 2 seconds. Figure 34 shows an example of a bonded chip at 110-μm pad spacing. Improved TAB techniques have been successful in reducing the spacing to less than 100 μm.

Tape-carrier package

Varieties of TCPs are used in industry today for ULSI devices. The package fabrication is usually done by plastic technologies—glob-top coating, or transfer molding that encapsulates the bonded chip with plastic resin materials. Figure 35 shows an example of a package in which the packaged chip can be electrically tested

Micron Ex. 1029, p. 604
Micron v. YMTC
IPR2024-00911

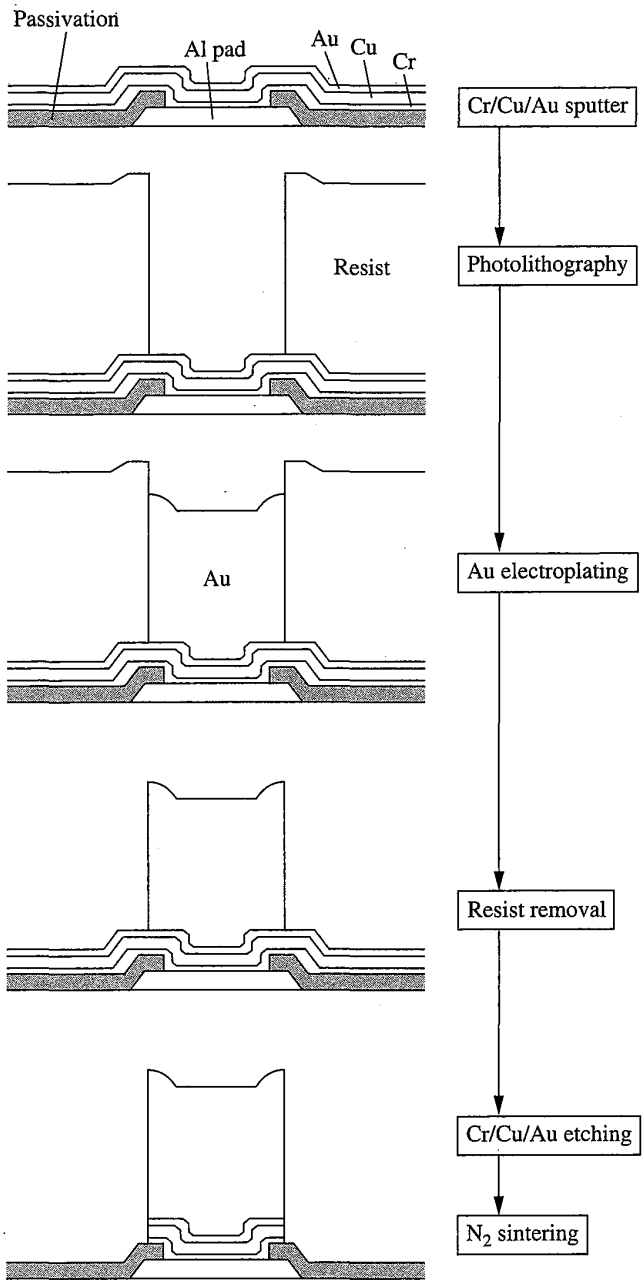


FIGURE 33
Example of the Au-bump-formation process that is done at the wafer stage. (*After Kohara, Ref. 56.*)

(including burn-in operations) using the larger test pads placed on the base polyimide tape that forms the opposite end of the bonded innerleads.[57] Posttesting and presoldering to the next-level packaging operation are the external lead-trim and form operations through which the package is formed into its final outer shape.

TCP has the potential to become a major choice in higher-I/O ULSI packaging since it enables one of the finest-pitch interconnects to be made. Obstacles to be

Micron Ex. 1029, p. 605
Micron v. YMTC
IPR2024-00911

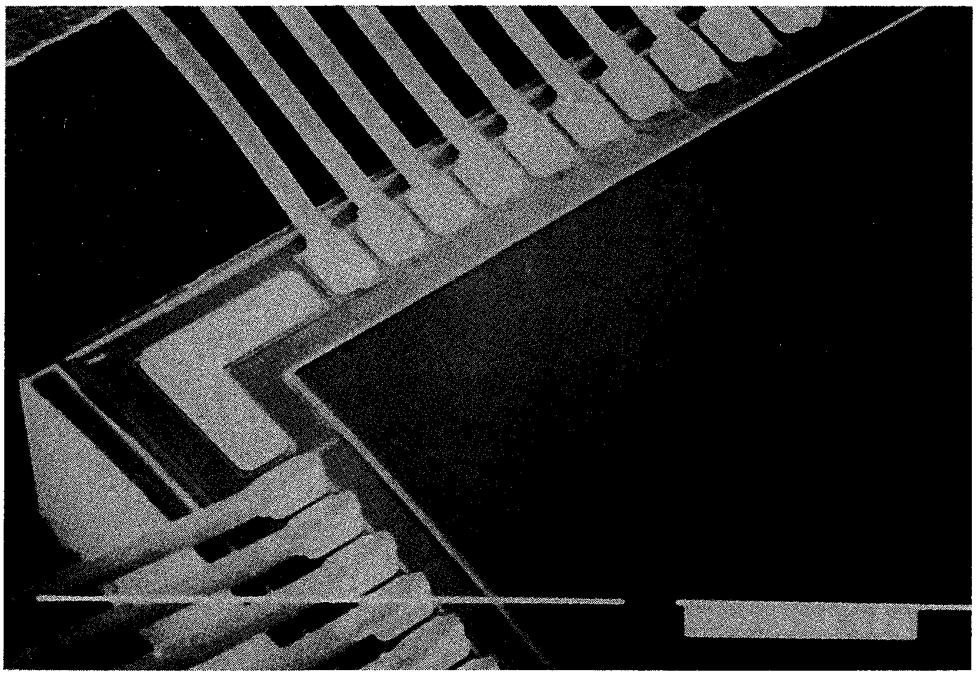


FIGURE 34
Scanning electron microscope photograph of an innerlead-bonded chip with 110-μm pad spacing. (*Courtesy of Mitsubishi Electric Corp.*)

surmounted are the relatively higher cost, the modest infrastructures involving test fixtures, and the next-level packaging technologies.

10.7.2 Flip-Chip Package

Flip-chip technologies fabricate bumps (typically Pb/Sn solders) on Al pads-on-chip and interconnect the bumps directly to the package media, which are usually ceramic- or plastic-based.

As Fig. 36[58] illustrates, in flip-chip technologies, the chip is face-down bonded to a package medium through the shortest path. These technologies can be applied not only to single-chip packaging, but also to higher or integrated levels of packaging in which the package media are larger and more sophisticated substrates that accomodate several chips to form larger functional units. Consequently, encapsulating flip-chip packages involves many variables, such as transfer molding, glob-top coating, and lid sealing, depending on the packaging level. Figure 36 shows a cross section in which the chip is attached to a ceramic substrate, yielding 528 pin interconnects. The flip-chip technique, using an area array, has the advantages of achieving the highest density of interconnection to the device and a very-low-inductance interconnection to the package. However, pretestability, postbonding visual inspection, and TCE matching to avoid solder bump fatigue are still challenges.

Micron Ex. 1029, p. 606
Micron v. YMTC
IPR2024-00911

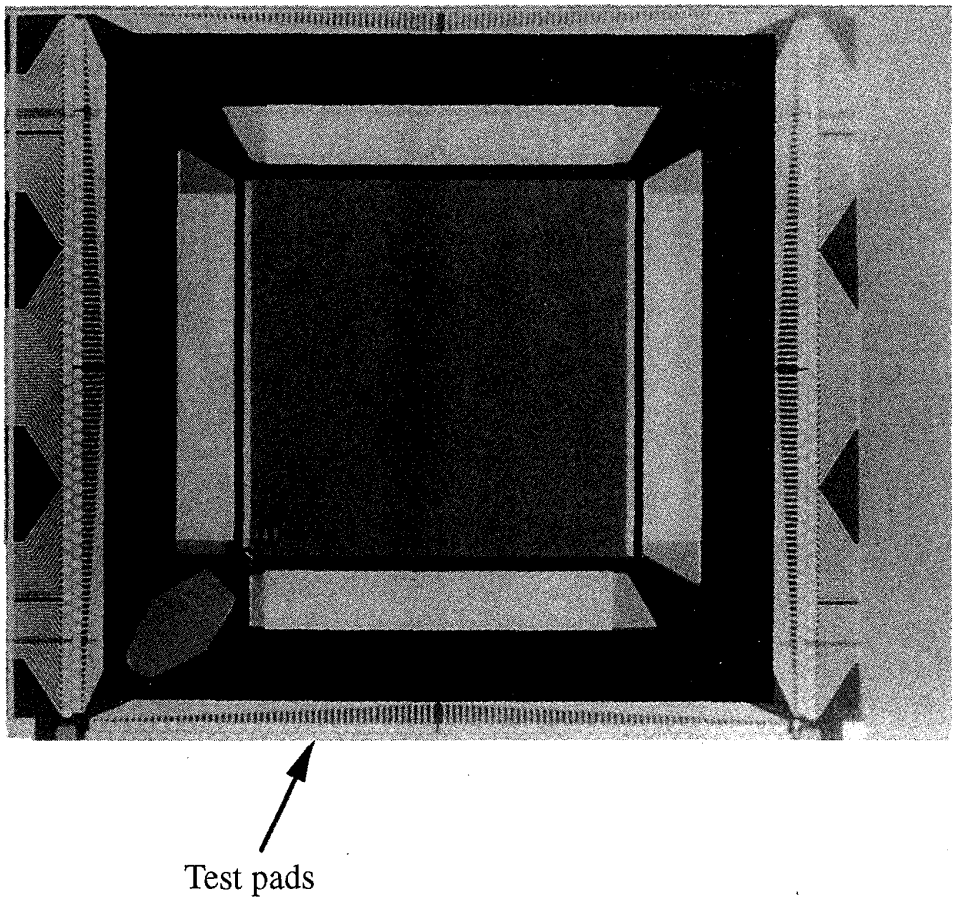


FIGURE 35
Example of a plastic TCP, in which the packaged chip can be electrically tested using the larger test pads placed on the base polyimide tape. (*Courtesy of Mitsubishi Electric Corp.*)

The solid, curved line in Fig. 37[59] shows that the original solder bump area is usually larger than that of barrier metals placed beneath. This area becomes closer in size to that of the barrier metals during the reflow-soldering operations. Thus, proper combinations of bump area, bump thickness, and barrier metal area must be selected to guarantee the interconnect integrities.

10.7.3 Ball-Grid Array Package

The recently developed ball-grid array (BGA) package is of interest in ULSI packaging. The BGA package is ceramic- or plastic-based and involves a variety of internal package structures. This package is now being assessed, standardized, and registered. This type of package features area-array external electrodes, usually Pb/Sn-

Micron Ex. 1029, p. 607
Micron v. YMTC
IPR2024-00911

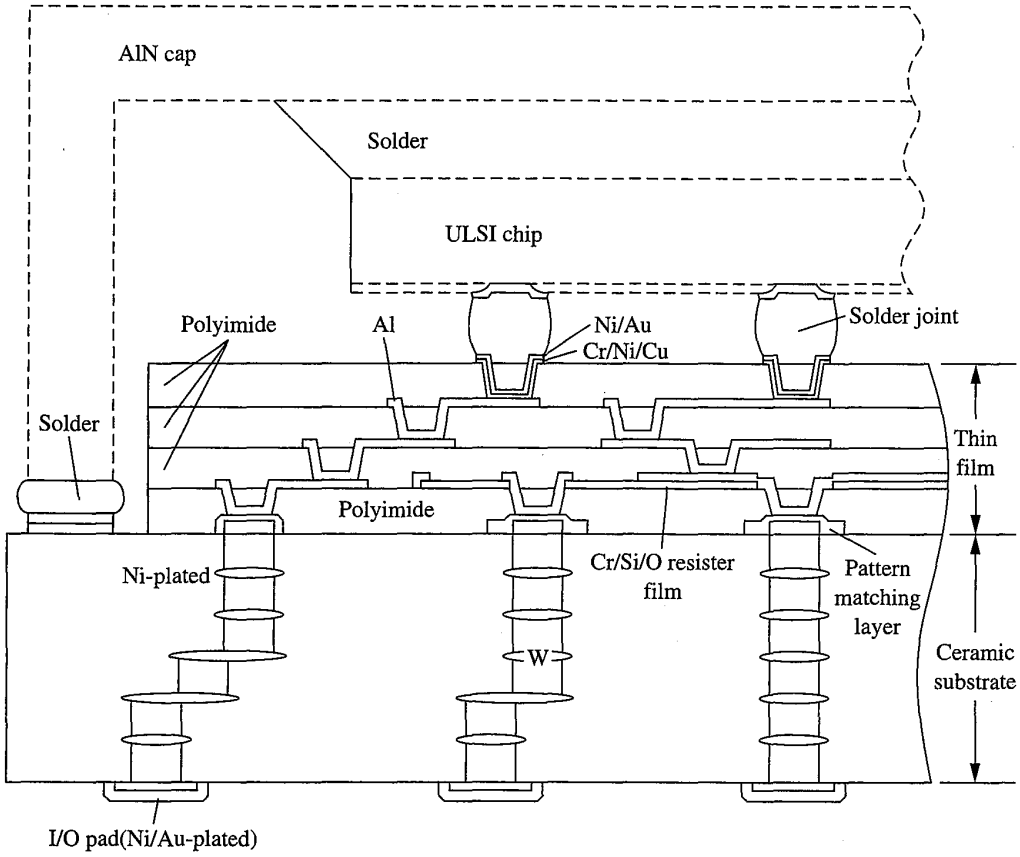


FIGURE 36
Cross section of a flip-chip package in which the chip is attached to a ceramic substrate and hermetically sealed. (*After Inoue, Ref. 58.*)

based bumps, placed on the package back surface at spaces ranging from 1 to 2 mm. The potential advantages are better external pin-count density, easier soldering to the PWB, lower inductance to the PWB, larger thermal paths to the package surroundings, and improved pretestability.

Figure 38[60] illustrates a simple BGA construction. An interconnected chip on a plastic resin substrate (typically BT resin) is transfer-molded, and solder balls are attached to the package back in post-molding operations. The figure shows thermal via holes placed in the substrate that provide shorter and significantly larger thermal paths from the chip to the mother PWB through the holes and the solder balls. More sophisticated BGA package structures have been proposed; these would contain multilayer substrates that have broader power planes, broader ground planes, or both for low-inductance and larger thermal-path connections.

Two popular choices for solder-bump formation are solder-ball attachment and solder-paste screen printing. Both methods would be followed by reflow operations to complete the metallurgical connections.

Micron Ex. 1029, p. 608
Micron v. YMTC
IPR2024-00911

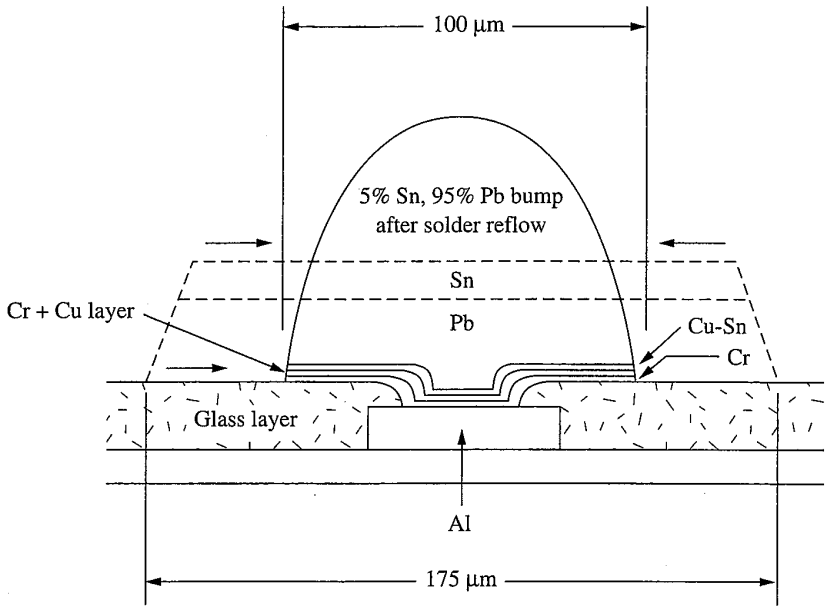


FIGURE 37
Cross section of a solder bump used in flip-chip technologies and the deformation caused by reflow-soldering operations. (*After Totta and Sopher, Ref. 59.*)

10.8 SUMMARY AND FUTURE TRENDS

Package pin count will undoubtedly continue to increase with increasing IC complexity. Alumina ceramic packages will continue to dominate a significant portion of the high-performance ULSI packaging technology until factors such as their high

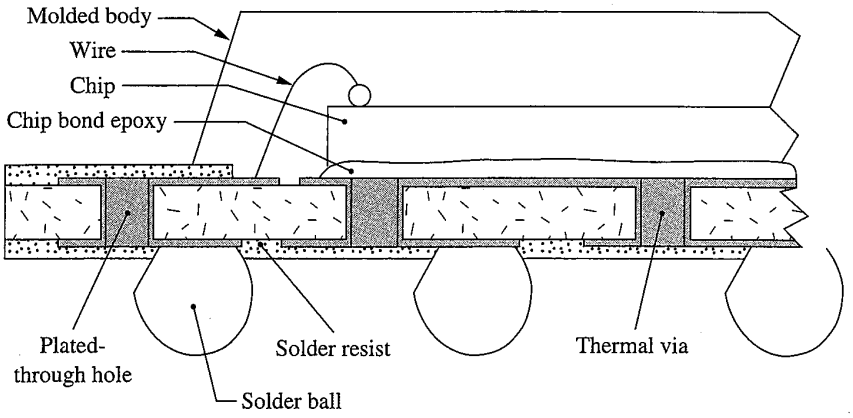


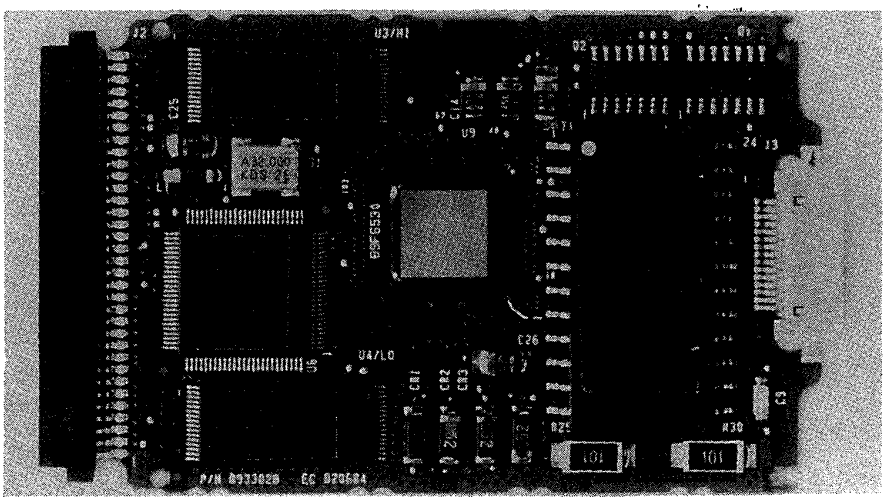
FIGURE 38
Cross section of a simple BGA package. The package is plastic-based and features area-array solder balls. (*After Sloan, Ref. 60.*)

Micron Ex. 1029, p. 609
Micron v. YMTC
IPR2024-00911

dielectric constant, modest thermal conductivity, or cost force a change to other contending materials. Plastic packages, especially posttransfer-mold plastic packages, will dominate most of the ULSI packaging technology. Higher packaging density on the PWB level will continue to drive package design gradually but steadily toward smaller lead pitches, to approximately 0.3-mm spacings.

If chip integration keeps pace with decreasing feature size, today's interconnection technologies, especially wire bonding, will be challenged.[61, 62] So far, wire bonding has successfully maintained the relative proportions of chip size and interconnection area generally up to an I/O of approximately 300 and often up to an I/O of approximately 500 interconnects on a chip with the penalties of cost, yield, or some chip design restrictions. TAB and flip-chip techniques will be the candidates to replace wire bonding. However, it probably will be some time before appropriate infrastructures are available to replace wire bonding technologies. Interconnections of high I/O (> 500 per chip) will have to be the choice among the three techniques.

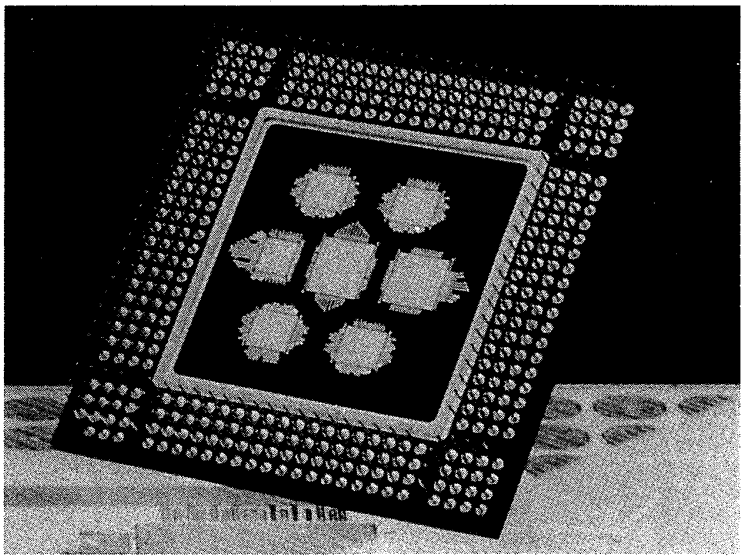
System design will increasingly depend on systematically optimizing the entire interconnection scheme to achieve the potential benefits of improved silicon capability. This optimization may lead to completely new requirements for assembly and packaging, but more than likely it will lead to sorting out the various existing assembly technologies and discovering the combinations that lend themselves to packaging-level integrations. The approach for various kinds of multichip modules (MCMs), shown in Fig. 39, verifies this trend.[63]



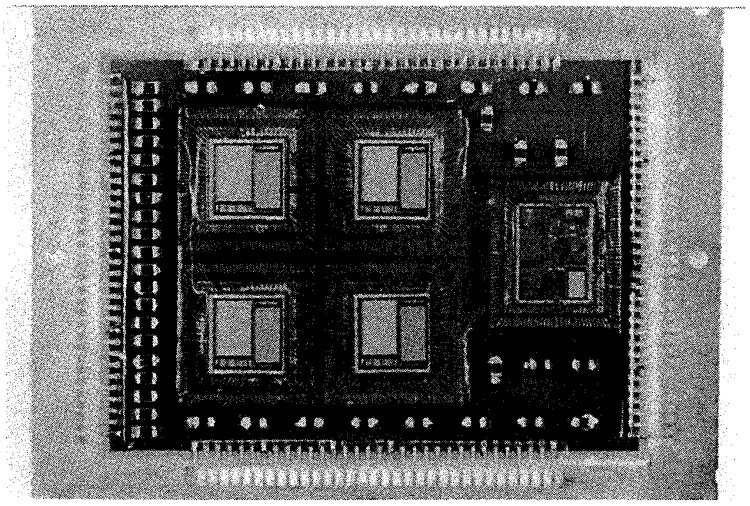
(*a*)

FIGURE 39
Three kinds of multichip modules. (*a*) MCM-Laminate (L) is PCB-based. Single-chip-packaged ULSI devices and other components, such as capacitors and resistors, are soldered onto the module. Sometimes bare chips can be used in lieu of packaged chips. Although the PWB is usually finely patterned, established surface-mount technology favors MCM-L growth. (*Courtesy of IBM Japan, LTD.*) (*continued*)

Micron Ex. 1029, p. 610
Micron v. YMTC
IPR2024-00911



(*b*)



(*c*)

FIGURE 39 (*continued*)
(*b*) MCM-Ceramic (C) is co-fired ceramic-based. Although packaged devices can be placed on the substrate, MCM-C is more likely to have bare chips, since the ceramic provides excellent TCE matching and thermal conductivity. (*Courtesy of Kyocera Corp.*) (*c*) MCM-Deposition (D) is usually ceramic- or silicon-based. Wiring patterns of the substrate are formed by deposition and achieve finer-features than MCM-L or MCM-C. Today, the use of this technology is limited, since the manufacturing cost is still high. (*Courtesy of Mitsubishi Electric Corp.*)

Micron Ex. 1029, p. 611
Micron v. YMTC
IPR2024-00911

REFERENCES

1. J. S. Prokop and D. W. Williams, "Chip Carriers as a Means for High-Density Packaging," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **3,** 297 (Sept., 1978).
2. J. R. Howell, "Reliability Study of Plastic Encapsulated Copper Lead Frame Epoxy Die Attach Packaging System," *Proceedings of the International Reliability Physics Symposium,* 1981, p. 104.
3. B. Prince, *Semiconductor Memories,* 2nd ed., Wiley, New York, 1991, p. 209.
4. N. Miyazaki and Y. Hayashi, *Nikkei Microdevices,* 37 (Mar., 1992).
5. K. Ohtsuka, "Trend of ASIC Package Technology," *Semic. World,* 114, (Jun., 1991).
6. C. J. Bartlett, "Advanced Packaging for VLSI," *Sol. State Technol.,* 119 (Jun., 1986).
7. I. Anjoh, "Revolutions in Leadframe for Memory, LOC Structures for 16MEGDRAM," *Nikkei Microdevices,* 89 (Feb., 1991).
8. W. C. Ward, "Volume Production of Unique Plastic Surface-Mount Modules for the IBM 80-ns 1-Mbit DRAM Chip by Area Wire Bond Techniques," *Proceedings of the 38th Elect. Components and Technol. Conference,* 1988, p. 552.
9. N. Ueda, "Studies on a New LOC Structure Package," *Proceedings of the VLSI Packaging Workshop,* III-4, 1993.
10. T. Chiu, "High-Density Memory Packaging with Vertical Surface Mount Package (VPAK)," *Proceedings of the 1st VLSI Packaging Workshop,* 1992, p. 37.
11. B. Frayman, "Overmolded Plastic Pad Array Carriers (OMPAC): A Low-Cost, High-Interconnect-Density IC Packaging Solution for Consumer and Industrial Electronics," *Proceedings of the 41st Elect. Components and Technol. Conference,* May, 1991, p. 176.
12. R. R. Tummala, "Electronic Packaging in the 1990s—A Perspective from America," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **14(2),** 262 (Jun., 1991).
13. H. Wessely, "Electronic Packaging in the 1990s—A Perspective from Europe," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **14(2),** 272 (Jun., 1991).
14. T. Ohsaki, "Electronic Packaging in the 1990s—A Perspective from Asia," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **14(2),** 254 (Jun., 1991).
15. R. Iscoff, "Thin Outline Packages, Handle With Care!," *Semic. Intl.,* 78, (1992).
16. K. Fujita, "Development of Ultra-Thin Surface Mounting IC Package (0.8mm Thick)," *Proceedings of the 40th Elect. Components and Technol. Conference,* May, 1990.
17. M. Kakei, "Low Stress Molding Compounds for VLSI Devices," *Nikkei Microdevices,* 82 (1984).
18. C. E. T. White and J. Slatery, "An Update on Preforms," *Circuit Manuf.,* 78 (Mar., 1978).
19. *Bonding Tools and Production Accessories, Bonding Handbook and General Catalog,* Kulicke and Soffa Industries (1980).
20. S. Uchida, "Technology Trend for Leadframe Manufacturing," *Proceedings of the SEMI Tech. Symposium,* November, 1988, p. 423.
21. R. A. Garner and R. W. Nufer, "Properties of Multilayer Ceramic Green Sheets," *Sol. State Technol.,* 38 (May, 1974).
22. N. Mogi and H. Yasuda, "Development of High-Reliability Epoxy Molding Compounds for Surface-Mount Devices," *Proceedings of the 42nd Elect. Components and Technol. Conf.,* May, 1992, p. 1023.
23. K. Tomiyoshi, "The Technical Trend of Epoxy Molding Compounds for Thinner Package of Semiconductor Devices," *Proceedings of the SEMI Tech. Symposium,* Oct., 1990, p. 399.
24. T. Sudo, "Considerations on Package Design for High Speed and High Pin Count CMOS Devices," *Proceedings of the 39th Elect. Components and Technol. Conference,* 1989, p. 531.

Micron Ex. 1029, p. 612
Micron v. YMTC
IPR2024-00911

25. H. R. Kaupp, "Characteristics of Microstrip Transmission Line," *IEEE Trans. on Computers,* **EC-16,** 185 (Apr., 1967).
26. R. K. Hoffmann, *Handbook of Microwave Integrated Circuits,* Artech House, Inc., Norwood, MA, 1987.
27. B. K. Bhattacharyya, "Ground Plane Design Parameters for CMOS VLSI Multilayer Packages," *Proceedings of the Technical Conference, Ninth Annual International Electronics Packaging Conference,* International Electronic Packaging Society, 1989, p. 659.
28. D. Mallik and B. K. Bhattacharyya, "High-Performance PQFP," *Proceedings of the 39th Elect. Components and Technol. Conference,* **VIII-1,** 1989, p. 493.
29. D. Mallik, "Multi-Layer Molded Plastic Package," *Proceedings of the 1989 Japan Intl. Elect. Manuf. Technol. Symp.,* **B4-2,** (Apr., 1989).
30. M. Aghazadeh and D. Mallik, "Thermal Characteristics of Single and Multilayer High Performance PQFP Package," *Proceedings of the 6th IEEE SEMI-THERM Symp.,* 33 (1990).
31. K. Manchester and D. Bird, "Thermal Resistance; A Reliability Consideration," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **31(4),** 550 (Dec., 1980).
32. P. G. Gabuzda, "Air Management System Yields High-Performance Cooling," *Electronic Packaging & Production,* 64 (May, 1988).
33. R. C. Chu, "Thermal Consideration and Techniques for Electronic Digital Computers," *Proceedings of the 9th IECP Symposium,* **5-3,** 1989, p. 3.
34. S. Kawai, "Structure Design of Plastic IC Packages," *Proceedings of the SEMI Tech. Symposium,* November, 1988, p. 349.
35. I. Fukuzawa, "Moisture Resistance Degradation of Plastic LSIs by Reflow Soldering," *Proceedings of the International Reliability Physics Symposium,* 1985, p. 192.
36. O. Nakagawa, "High Density Packaging Technology," *J. Electron. Matls.,* **18(5),** 633 (1989).
37. T. Nishioka, "Special Properties of Molding Compound for Surface Mounting Devices," *Proceedings of the 40th Elect. Components and Technol. Conference,* May, 1990, p. 625.
38. S. Ito, "Molding Compounds for Thin Surface Mount Packages and Large Chip Semiconductor Devices," *Proceedings of the 41st Elect. Components and Technol. Conference,* May, 1991, p. 190.
39. M. Kitano, "Analysis of Package Cracking during Reflow Soldering Process," *Proceedings of the International Reliability Physics Symposium,* 1988, p. 90.
40. J. H. Lau, *Solder Joint Reliability: Theory and Applications,* Van Nostrand Reinhold, New York, 1991.
41. J. H. Lau, and G. H. Barrett, "Stress and Deflection Analysis of Partially Routed Panels for Depanelization," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **10(3),** 411, (Sept., 1987).
42. H. D. Solomon, "Solder Joint Acceleration Factor," *J. Electronic Packaging, Trans. of ASME,* **113,** 186 (Jun., 1991).
43. H. D. Solomon, "Fatigue of 60/40 Solder," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **9,** 423 (Dec., 1986).
44. J. H. Lau, "Experimental and Analytical Studies of 208 Pin Fine Pitch Quad Flat Pack Solderjoint Reliability," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **9,** 122 (Jun., 1991).
45. J. H. Lau, "Solder Joint Reliability of Fine Pitch Surface Mount Technology Assemblies," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **13(3),** 534 (Sept., 1990).
46. J. H. Lau, "Thermal Stress Analysis of TAB Packages and Interconnections," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **13(1),** 182 (March, 1990).

Micron Ex. 1029, p. 613
Micron v. YMTC
IPR2024-00911

47. J. H. Lau, "Thermal Stress Analysis of SMT PQFP Packages and Interconnections," *J. Electronic Packaging, Trans. of ASME,* **111,** 2 (Mar., 1989).
48. M. L. White, "Attaining Low Moisture Levels in Hermetic Packages," *Proceedings of the 20th International Reliability Physics Symposium,* 1982, p. 253.
49. M. L. White, "The Removal of Die Bond Epoxy Bleed Material by Oxygen Plasma," *Proceedings of the 32nd Elect. Components and Technol. Conf.,* 1982, p. 262.
50. W. Paulson and R. Kirk, "The Effect of Phosphorous Doped Passivation Glass on the Corrosion of Aluminum," *Proceedings of the International Reliability Physics Symposium,* 1974, p. 172.
51. T. C. May and M. H. Woods, "Alpha-Particle-Induced Soft Errors in Dynamic Memories," *IEEE Trans. Electron Dev.,* **26,** 2 (1979).
52. K. Yamaguchi and M. Igaroshi, "Screen Printing Grade Polyimide Paste for Alpha-Particle Protection," *Proceedings of the 36th Elect. Components and Technol. Conf.,* May, 1986, p. 340.
53. M. M. White, "The Use of Silicone RTV Rubber for Alpha Particle Protection on Silicon Integrated Circuits," *Proceedings of the International Reliability Physics Symposium,* 1981, p. 43.
54. R. L. Clain, "Beam Tape Carriers—A Design Guide," *Sol. State Technol.,* March, 1978, p. 53.
55. K. Hatada, "New Film Carrier Assembly Technology: Transferred Bump TAB," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **10(3),** 335 (1987).
56. M. Kohara, "New Development of Thin Plastic Package with High Terminal Counts," *IEEE Trans. Components, Hybrids, Manuf. Technol.,* **13(2),** 401 (1990).
57. T. Ueda, "High Pin Count and High Power Package," *IEEE Proceedings of the VLSI & GaAs Chip Packaging Workshop,* 1991, p. 13.
58. T. Inoue, "Micro Carrier for LSI Chip Used in the HITAC M-880 Processor Group," *Proceedings of the 41st Elect. Components and Technol. Conf.,* 1991, p. 704.
59. P. A. Totta and R. P. Sopher, "SLT Device Metallurgy and its Monolithic Extension," *IBM J. Res. Devel.,* **13,** 226 (1969).
60. J. Sloan, "Over Molded Pad Array Carrier (OMPAC): A New Kid on the Block," *Proceedings of the 1st VLSI Packaging Workshop,* Nov., 1992, p. 17.
61. K. Hatada, "Applications of New Assembly Method—Micro Bump Bonding Method," *Proceedings of the 1989 Japan Intl. Elect. Manuf. Technol. Symposium,* 1989, p. 47.
62. H. Fujimoto, "Bonding of Ultrafine Terminal-Pitch LSI by Micron Bump Bonding Method," *Proc. IMC1992,* 1992, p. 115.
63. K. Nakamura and Y. Nagahiro, "Emerging MCMs Behave Like System on Silicon," *Nikkei Microdevices,* 28 (April, 1993).

PROBLEMS

1. Estimate the necessary package pin count for a 40,000-gate, logic-gate-array chip using Rent's rule (Eq. 10.2). Assume that $\alpha = 1.5$, $\beta = 0.5$, and that the chip needs one pair of power/ground package pins per 10 signal pins (or signal terminals).

2. Suppose you have an Au-Al bond, and the metallurgical diffusion thickness (with Al as a solvent metal and Au as a diffusion metal) increased by 1 μm during a 150°C time-temperature storage operation. Calculate the period of time using Eqs. 10.3 and 10.4. Assume $D_o = 2.2 \times 10^{-4}$ (m^2/sec), $Q = 134$ (kJ/mol), and $R = 8.31$ (J/mol-K).

Micron Ex. 1029, p. 614
Micron v. YMTC
IPR2024-00911

3. A high-temperature storage considerably accelerates Au-Al diffusion and affects the bond quality. Calculate the diffusion thickness increase caused by a 200°C, 60-day storage operation. Use the same assumptions as in Problem 2.

4. What chip size is required to place 352 pads (I/Os) on a pad-limited chip? Assume that the chip is square, the bonding-pad pitch is 120 μm, and the chip size is defined by the loci of bonding-pad centerlines. Use Fig. 16 to find the answer.

5. Derive Eq. 10.13. (Note that $l \gg a$).

6. Suppose you have a 5-volt, 1-nsec output-rise-time silicon chip with 68 signal leads and you want to package it so that inductive noise in the ground line is kept to 0.2 volt. You want to switch 18 lines simultaneously, and you want each package lead to have approximately 7 nH inductance. The output buffers draw 25 mA. How many package leads do you need for this device, assuming that there are eight power leads?

7. A high-power plastic packaged device that allows a maximum junction temperature of 85°C dissipates 10 watts in an ambient of 40°C. Calculate the required thermal resistance (θ_{ja}) of the package.

8. Discuss how the package described in Problem 7 affects the next-level (PWB) packaging and the final system construction.

Micron Ex. 1029, p. 615
Micron v. YMTC
IPR2024-00911

CHAPTER 11

Wafer Fab Manufacturing Technology

T. F. Shao and F. C. Wang

11.1 WHAT IS MANUFACTURING?

Semiconductor manufacturing is often referred to as a *high-tech industry.* High-tech industries are typically characterized by a multidisciplinary approach and the quick adoption of advances in science and technology to commercial production. This chapter describes how technologies are implemented for manufacturing.

Manufacturing is one link in the iterative economic chain of providing goods for society. The following is a typical flow of goods:

Customer's need → Conception of products → Product design →
Product prototyping → Test marketing → Sales → Manufacturing →
Customers → Payment → Customer satisfaction → More customer's needs →

In this economic chain, the key role of manufacturing is to produce products the customer wants. This role must be supported by the timely infusion of the necessary technologies to permit a manufacturer to continue serving its customers and make a profit.

11.1.1 Critical Success Factors

To serve the customers, the critical success factors considered important are these:

- *Quality:* Quality is only the first step in satisfying customers' demand.[1–6] Products must meet the customer's application requirements. These requirements typically are presented in the form of a product specification. To produce an acceptable product, the manufacturing process must have built-in quality control to ensure that the output meets the customer's requirements.

Micron Ex. 1029, p. 616
Micron v. YMTC
IPR2024-00911

A modern integrated circuit is produced through hundreds of steps that involve sophisticated equipment and precise control either by an automatic or a manual discipline. Quality can be achieved, not by inspection and screening, but only through well-proven and controlled processes. The typical methodology for process control will be described later in this chapter.

- *Cost:* Given the same quality of product, customers invariably will compare the price of the product. Generally, the lower-priced, but equal-quality, product is bought. Thus, a manufacturer must constantly make sure that cost is low enough that the market price results in a profit.
- *Delivery:* Customers expect goods with the promised quality within the promised delivery time.[2] In many instances, quick delivery can result in premium prices or higher market share.

  Thus, the manufacturer must put in place the necessary infrastructures and training to produce integrated circuits within short cycle times and in the quantities that customers order. For commodity products such as DRAMs, the quantity can be in the millions; for prototype circuits, the quantity can be under 10.
- *Service:* Service is the "soft" side of hardware manufacturing in providing goods for the customers. The manufacturer must be able to provide services,[6] such as, how to select the right product for the customer's application; how to apply the product in the customer's application; how to troubleshoot; how to adjust delivery when customer's need for the product changes; etc. The customer service related to the product can become more and more important when the quality/cost/delivery characteristics of all manufacturers are similar.

11.1.2 Manufacturing Flow

An integrated circuit provides integration of active and passive components on a common substrate to perform desired functions. Typically, the structure is manufactured by a process involving photopatterning, film growth, implantation, diffusion and film etching, etc. A typical process flow is given in Table 1 to illustrate the process classifications. A measure of the complexity of the manufacturing flow is the number of photopatterning levels. It is not uncommon to see ICs made with more than 20 levels, and the total number of process steps can easily exceed 200.

11.1.3 Manufacturing Organization

The organization for manufacturing ICs is determined by the relation of manufacturing flow and its functional blocks as shown in Fig. 1. Based on the functional relationships, the goals of the total organization, and other factors such as the experience level of manpower, local perceptions, etc., many forms of organization can be structured. Two of these forms are known as classical organization, Fig. 2, and modular organization, Fig. 3. Each organizational structure has its pros and cons. The classical organization enhances professional development along its special func-

Micron Ex. 1029, p. 617
Micron v. YMTC
IPR2024-00911

TABLE 1
Schematic illustration of a typical wafer fab manufacturing process flow

Process steps	Equipment used	Photo	Thin film	Ion implant	Diffusion	Etch
Wafer ID marking	Laser marker					
First oxidation	Furnace				X	
First nitride deposition	CVD furnace		X			
n-tank pattern	Coater/stepper/developer	X				
First nitride etch	Plasma etcher					X
n-tank implant	Ion implanter			X		
Stripe nitride	Hood				X	
p-tank pattern	Stepper	X				
Tank drive	Furnace					X
Second oxidation	Furnace					X
Second nitride deposition	CVD furnace		X			
Moat pattern	Coater/stepper/developer	X				
Second nitride etch	Plasma etcher					X
Field oxidation	Furnace				X	
Stripe nitride	Hood				X	
Gate oxidation	Furnace				X	
Poly 1 Deposition	CVD furnace		X			
⋮						
BPSG deposition	CVD reactor		X			
Reflow	Furnace				X	
Contact pattern	Coater/stepper/developer	X				
Contact etch	Plasma etcher					X
First metal deposition	Sputtering machine		X			
⋮						
Protective overcoat	CVD reactor		X			
Pad pattern	Coater/printer/developer	X				
Wafer test	Tester				X	

tional line. Experts can be developed in depth. The modular organization enhances the team effort by the process, equipment, and manufacturing in a given machine group such as photo, etch, diffusion, etc. Generally speaking, for any organizational structure, an interorganizational team is necessary to minimize the barriers that develop along the organizational lines and to enhance the cross-organizational teamwork.[2,3] The interplay and effectiveness of the organization cannot be rigidly prescribed. Management should comprehend the objectives and other background factors, then structure the organization appropriately.

11.1.4 The Cost Structure of Manufacturing

Semiconductor manufacturing is capital-intensive. Each wafer fab typically requires investment of several hundred million dollars. The physical building and facilities

Micron Ex. 1029, p. 618
Micron v. YMTC
IPR2024-00911

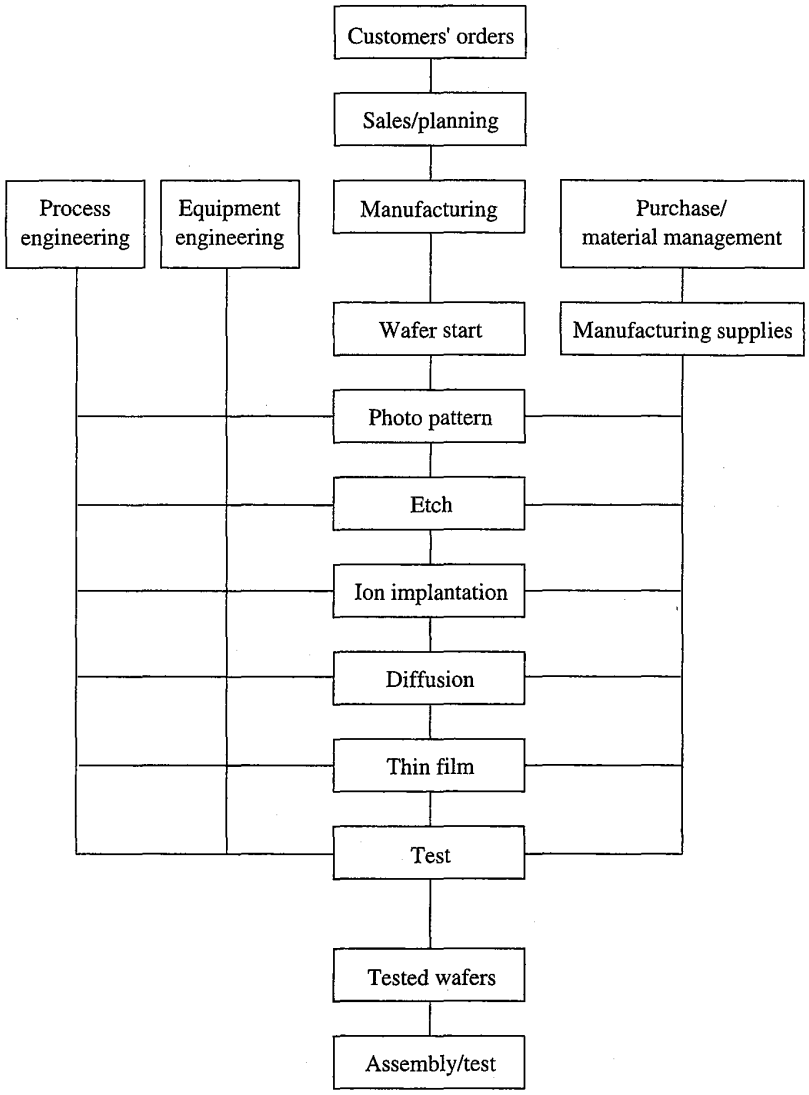


FIGURE 1
Manufacturing flow and functional block diagram.

can cost $100 million or more. Major process equipment may be priced on the order of $1 million. Beyond the capital cost, a representative monthly operating cost of a factory is several hundred thousand to more than $10 million consisting of material, labor, and overhead.

The function of manufacturing is to utilize the capital-intensive factory to produce the integrated circuits in a cost-effective manner. The speed with which a factory can move from start-up to full capacity is becoming increasingly important to meeting financial expectations and reducing risk of technology turnover.

How to manage a company for profit and growth is itself a subject requiring serious study by the management people.[1–3,5,6] For students starting on a career in

Micron Ex. 1029, p. 619
Micron v. YMTC
IPR2024-00911

FIGURE 2
Structure of a classical organization.

Micron Ex. 1029, p. 620
Micron v. YMTC
IPR2024-00911

FIGURE 3
Structure of a modular organization.

Micron Ex. 1029, p. 621
Micron v. YMTC
IPR2024-00911

engineering and manufacturing, it is important to know that each innovation or change must be reviewed with economic judgment before implementation in manufacturing.

Wafer fab manufacturing is very complicated with many dependent variables. A wafer fab commonly has several hundred pieces of equipment consisting of more than 30 different machine types. This equipment is operated continuously by several hundred operators in shifts around the clock. Under these conditions it takes months to complete one product cycle after hundreds of processing steps. Yet, the finished products must consistently meet customers' standards of quality and reliability. For commodity volume products, such as DRAMs in particular, several million units must be fabricated a month. To deal with such complexity, built-in manufacturability is most important, with sufficient margin at the start-up of a wafer fab and good process control to sustain volume production. The control and reduction of all variations[4,6] are emphasized for wafer fab manufacturing, which may not be the case for other chapters. To ensure that any subsequent changes enhance manufacturability but do not have a detrimental impact requires a control system for the improvement process. Those are the subjects to be considered in subsequent sections.

11.2 WAFER FAB MANUFACTURING CONSIDERATIONS

In this section, several common manufacturing considerations are introduced. These considerations are aimed at establishing the supporting foundation for designed-in manufacturability.

11.2.1 Typical Process Layout Considerations

It is difficult to discuss process layout without using an actual wafer fab design. However, several manufacturing concerns for layout follow:[2]

- *Relationship to utility connection:* Since the nature of each process group (diffusion, photo, etching, thin film) is different, the utility requirements are also different. For example, deionized water (DIW) and chemicals are used primarily by the wet-hood operation. Grouping similar processes in the process layout provides advantages by simplifying cost savings in acquiring supplies and by reducing maintenance efforts.
- *Wafer moving path:* Wafers move constantly during the entire manufacturing cycle from one equipment unit to another, from one process (area) to another process (area), and from a process to measurement/inspection. The moving paths based on the process flow (Fig. 1) should be planned carefully to minimize the movement of the wafers so as to reduce cycle time and defects.
- *Routing for manufacturing supplies:* In addition to the wafers that are being produced, other manufacturing supplies such as pilot/dummy wafers, quartzware, etc. also move constantly inside the wafer fab to support process needs. Routing considerations are similar to the foregoing.

Micron Ex. 1029, p. 622
Micron v. YMTC
IPR2024-00911

- *Inventory storage:* Wafers waiting to be processed are stored somewhere inside the wafer fab. Sufficient storage space at appropriate locations must be provided in the layout to furnish easy access and clear segregation to avoid improper operation. Inventory storage of manufacturing supplies requires the same layout considerations.
- *Cross-contamination:* Processes that use little or no phosphorus and those that use much phosphorus should be segregated in the layout of equipment, process tools (wafer cassette, wafer transfer, etc.), and measurement instruments so as to avoid cross-contamination. Nonmetal and metal processes should be segregated as well.
- *Other considerations:* The layout must account for other manufacturing needs such as storage space for instructions (specifications, procedures, notes), housekeeping materials, maintenance parts/tools, etc. to ensure that the working area is clean and orderly.

Readers are encouraged to gain personal experience by studying actual wafer movement inside a wafer fab to appreciate the problems that can result from improper layout.

11.2.2 Automation

With the increased complexity of ICs and the sophistication of process equipment, automation is almost a necessity. The degree of automation is a trade-off of cost and benefit and is limited by the maturity of the automation hardware and software. Automation is also an important planning and control tool for wafer fab manufacturing. Automation can be classified into two categories:

- *Physical automation*

  Factory utility control: Temperature and humidity are usually controlled automatically (see Chapter 1 on cleanroom technology). The level of airborne particulates and the purity of chemicals are monitored, and equipment can be set to issue warning and shutdown signals.

  Wafer transfer: Wafers can be automatically transferred from one process equipment to another (e.g., cluster tools).

  Wafer processing: Wafers arriving at a given process equipment can be automatically loaded then unloaded after completion of processing. The process recipe for each process can be automatically downloaded from a central computer, which sets up the equipment for processing. Cross-checking by the computer can practically eliminate all errors that otherwise may be caused by human operator mistakes.

- *Information automation*

  Inventory control: All wafers or lots are given an identification and can be tracked throughout the processes. The movement of the inventory can be controlled by computer based on customer's need, equipment conditions, special instructions from engineering, etc.

Micron Ex. 1029, p. 623
Micron v. YMTC
IPR2024-00911

Equipment control: The performance and maintenance status of the production equipment can be tracked by computer. This control will facilitate both production scheduling and equipment monitoring for capacity planning.

Cost collection: For purposes of financial decisions the cost of wafer processing can also be determined as each wafer passes through the process steps.

Technical data: At each process point, the input and output data, as well as real-time process conditions, can be collected and compared against the requirements (specifications). In case of a deviation, either automatic adjustments can be made or a signal sent as a warning or for shutdown. Technical data can be used not only for engineering purposes but also for process control.

11.2.3 Material Requirement

The quality and consistency of the materials greatly affect the yield and performance of the devices. Typically, materials can be separated according to whether they do or do not have direct contact with wafers during processing. Materials in contact include the wafers themselves, chemicals, gases, metal targets, and photo masks; these have more stringent requirements. Some materials do not have direct contact with wafers but still impact the process results, for example, quartz, wafer cassettes, cleanroom clothing, and cleanroom cleaning cloth for equipment. These critical materials should undergo incoming quality checks and specification controls to ensure that requirements for quality and consistency of supply are met. Other materials used inside the wafer fab, but not specifically mentioned here, are less critical except for considerations of minimal particle and contamination generation, cost, and supply convenience.

The material requirements for modern wafer fab manufacturing are extremely difficult to discuss because of the enormous variety of types of materials and broad ranges of specifications used for the IC industry. Instituting a requirement without understanding the exact process details and device parametric design is, in fact, dangerous. We will provide some basic material considerations in light of the importance of the subject to manufacturing success. Those generic requirements will attempt to balance both today's and near future advanced wafer fab needs over a broad scale of product varieties. The suppliers' capabilities are an integral part of material considerations. We do not want to make impractical demands and bear unnecessary high costs, but on the other hand, we do not want to compromise on actual process needs.

Direct material considerations

This section addresses wafers, chemicals, gases, metal targets, and photomasks. In Table 2, several common requirements[7–11] for 150- and 200-mm wafer processes are listed. ASTM[7] and SEMI[8] have on-going activities to discuss, upgrade, and standardize wafer requirements. For more details, please refer to ASTM/SEMI documents or attend meetings of the respective committees.

Unlike the wafer, there have been fewer attempts to standardize the chemical property requirements.[12] This is understandable since the effects of chemicals depend more on process sequence and the criticality of the process to device

Micron Ex. 1029, p. 624
Micron v. YMTC
IPR2024-00911

TABLE 2
Typical wafer requirements for manufacturing

Item	Bare or EPI wafer	Purpose	Typical specification
General			
Diameter	B and E	Wafer size	150 or 200 ±0.5 mm
Orientation	B and E	Film growth and implantation	(100) ±1
Conductivity type	B and E	Type of doping for conductivity	P(boron), N(phosphorus)
Electrical			
Resistivity	B	Process and control device parameter	1–20 ohm-cm
	E		0.001–0.01 ohm-cm
Radial variation	B and E	Control of variation across wafer	< 15%
Minority carrier lifetime	B	Control of impurity concentrations	> 200 μsec
Mechanical			
Thickness	B and E	Mechanical strength and process equipment features	600–750 ±40 μm
Bow and wrap	B and E	Mechanical stress and photoprocess	> 60 μm
Global flatness	B and E	Control of thickness variation across water	< 20 μm
Site variation	B and E	Control of local thickness uniformity for photo exposure field	> 1 μm focal plane deviation Min. size 20 mm× 20 mm
Surface finish	B	Specify surface preparation technique for mechanical specifications and particle/contamination control	Chemical/mechanical polish
Backside finish	B and E	Particle and gettering control	Soft blasted/etched, oxide or poly-Si coated
Edge profile or chem/mech	B and E	Particle control and microcrack elimination	Edge beveled/etched polished
Structural			
Oxygen-induced stacking fault (OISF)	B	Minimize additional process-included defect generation	< 100 /cm^2
Dislocation	B	Same as above	< 50 etch pits/cm^2
Slip line	B	Same as above	None

Micron Ex. 1029, p. 625
Micron v. YMTC
IPR2024-00911

TABLE 2 *(continued)*
Typical wafer requirements for manufacturing

Item	Bare or EPI wafer	Purpose	Typical specification
Chemical			
Oxygen	B	Control of internal gettering effectiveness	20–32 ppma
Carbon concentration	B	Control of carbon-related nucleation for gettering	< 1 ppma
Surface metal concentration	B	Minimize variation of device electrical parameters related to surface contamination	$< 5 \times 10^{10}$ atoms/cm^2 for Fe, Ca, Na, Ni, Cr, Zn, Cu, Mg, etc.
Surface features			
Particle	B and E	Starting material defect control for yield consideration	< 0.16 /cm^2 @ 0.2 μm
Haze	B	Same as above	None detectable
Other surface defects (pits, dimples, cracks, chips, scratches, organic peels, contaminations, etc.)	B and E	Same as above	None visible

performance. For example, the importance of the impurity levels in chemicals varies according to whether the NH_4OH–H_2O_2–H_2O/HF-based or HCl–H_2O_2–H_2O-based wet cleanup is used as the last cleanup sequence for critical processes such as pregate oxidation. Yet, there is more demand for lower impurity levels for NH_4OH, H_2O_2, HF, or HCl than for H_2SO_4 or H_3PO_4 as a result of the process criticality itself (H_2SO_4-based photoresist or HNO_3-based quartz cleanups naturally are dirtier). In addition, not all impurities in the same chemical are important. Concerns about typical inorganic impurities (such as Al, Ca, Cu, Cr, Fe, Na, Ni, Mg, and Zn) are based on the known impact of these elements on device electrical performance [such as threshold voltage V_t shift, weak gate oxide integrity (GOI), and degraded minority carrier lifetime].[13] Because an element is not mentioned above does not imply that it has no effect on device performance or makes no contribution to manufacturing variations. The reader is advised to look into the contribution of specific elements to specific device parameters or under sporadic abnormal situations (e.g., excessive contamination). Refer to Chapter 2 for details on cleanup.

In Table 3, typical specifications for several elements of interest in some common chemicals are provided. The actual level of impurity for those elements, as well as the consistency of contamination from one lot to another, vary by orders of magnitude between different industrial suppliers. Readers must check the actual quality and consistency performance levels and cost with respect to individual process needs before selecting a particular vendor to supply a specific chemical. Impurities

Micron Ex. 1029, p. 626
Micron v. YMTC
IPR2024-00911

TABLE 3
Typical chemical requirements for manufacturing

Impurity	LDL†	IPA†	Acetone	HF	NH_4OH	H_2O_2	HNO_3	HCl	H_2SO_4	H_3PO_4
Particles	> 0.5 μm	50/mL	20/mL	5/mL	30/mL	50/mL	50/mL	50/mL	30/mL	50/mL
					Parts per billion					
Chloride	< 1	50	50	10	200	100	100	—	100	1000
Nitrates	< 100	1000	—	10	—	2000	—	—	200	5000
Phosphates	< 10	50	50	5	100	200	50	50	500	—
Sulfates	< 100	500	—	10	500	200	100	50	—	10000
Aluminum	< 0.1	2	5	1	1	1	1	5	5	300
Calcium	< 0.1	1	0.5	1	3	3	10	5	5	500
Chromium	< 0.1	1	0.5	1	1	1	1	2	2	—
Copper	< 0.1	1	0.5	1	1	1	1	1	2	50
Iron	< 0.1	1	0.5	1	1	1	10	10	5	500
Magnesium	< 0.1	1	0.5	1	1	1	2	2	5	500
Nickel	< 0.5	1	0.5	1	1	1	1	2	2	50
Potassium	< 0.1	1	1	1	1	1	2	5	5	100
Sodium	< 0.1	5	5	1	1	5	5	10	5	100
Zinc	< 0.5	1	0.2	1	1	1	5	2	2	—

† LDL:lower detection limit
IPA: isopropyl alcohol

Micron Ex. 1029, p. 627
Micron v. YMTC
IPR2024-00911

not listed in Table 3 are commonly below detection limits, however, detection limits may vary based on the vendor's analytic capability. Similarly, the particle content in chemicals varies greatly. This is related not only to the manufacturer's process capability and control of particles in chemicals but also to additional particle generation arising from packaging materials and transportation and to the sampling and measuring techniques used. The particle quality is typically guaranteed through extensive filtration in the distribution system rather than from direct control on incoming chemicals.

Gas requirements are identical to those for chemicals.[14] In Table 4 some typical purity specifications for common wafer fab gas supplies are listed. Specifications in Table 4 refer to values at the outlet of the wafer fab distribution systems after filtration. Once again, actual performance may be well below the specification value. Metallic impurities in gases are typically not specified and rely on the vendor's guarantee.

Specialty chemicals are primarily those used for photo and metal cleanup-related processes. A common requirement for photoresists, developers, SOG, and polymide is the transportation and storage temperature. The photosensitivity and lifetime of a photoresist varies with temperature, hence the need to maintain temperature of materials within specifications of 5 to 20°C for photoresist/developer and −20 to 10°C for SOG/polyimide. The specialty gases are primarily those used for deposition (SiH_4, DCS, PH_3, NH_3, HCl, etc.) and plasma etching (NF_3, CF_4, C_2F_6, BCl_3, HBr, etc.). Typically, common impurities such as N_2, O_2, H_2O, and THC (total hydrocarbon) are in the sub- to several-ppm range, as are metallic elements of concern. The required limits of the manufacturing by-products in a specific specialty gas—for example, $SiCl_4$, $SiHCl_3$, and SiH_3Cl in SiH_2Cl_2 (DCS)—can be handled only on a case-by-case basis. The interior coating materials for the cylinder, the connection valve assembly, and the vendor's handling and control procedures for cylinder preparation and gas filling are also areas of concerns in providing quality specialty gases. In most cases one would rely solely on the manufacturer's guarantee to meet the material requirements for specialty chemicals and gases routinely.

In Table 5 typical metal-target material requirements are listed. These are primarily guaranteed by the manufacturers at the time of delivery. As the materials are used, continuous cooperation with the manufacturer is necessary to understand the target degradation characteristics and target-to-target consistency.

Manufacturing considerations for photomasks emphasize defects and consistency. Defects and critical dimensions (CD) control within a mask and from photomask to photomask are important parameters, since they transfer to the patterned wafer every time it prints. Table 6 lists several material requirements for photomasks.

Deionized water (DIW) of high purity is actually an important material supply[15] but is often forgotten as a direct material, since it is generated at the wafer fab rather than being delivered by a vendor. DIW has been used extensively for all the wafer fab manufacturing process cleanup activities and for supply material preparation. In Table 7, typical DIW requirements are listed. The parameters in Table 7 should be monitored routinely at the DIW supply system and at the point of use.

All the requirements discussed in this section are for reference only. One should not apply them blindly.

Micron Ex. 1029, p. 628
Micron v. YMTC
IPR2024-00911

TABLE 4
Typical gas requirements for manufacturing†

Gas	Purity (ppb) Ar	N_2	O_2	H_2	CO	CO_2	THC‡	H_2O	Particles #/FT³ at ≥ 0.1μm
Nitrogen	—	—	50	50	50	50	10	50	5
Oxygen	500	100	—	50	50	50	10	50	5
Hydrogen	10	20	10	—	50	50	10	50	5
Argon	—	50	50	50	50	10	50	50	5
Helium	—	50	50	50	50	50	10	50	5

† Specifications refer to values at the outlet of the distribution system after filtration
‡ THC: Total hydrocarbon

TABLE 5
Typical metal target requirements for manufacturing

Item	Purpose	Typical specification
General		
Dimension and shape	Match sputter machine configuration	Per sputter machine drawing for bonded target thickness and diameter
Bonding	Specify bonding requirements between target itself and backing plate for bonding area, backing plate materials, and minimum withstanding power	Per sputter machine specifications
Chemical		
Composition	Specify major constituents of target	Per process specification
Purity	Total purity level	> 99.99%
Total metallics	Control of metal impurity	< 100 ppm
Metallic impurity	Same as above	< Few ppm per each element
Other impurity	Control of nonmetallic impurity	< 1000 ppm per each element (H, N, O, C, etc.)
Visual defects		
Appearance	Quick check on surface feature nonuniformity	Uniform appearance
Stain, surface contamination	Quick check on surface quality	None visible
Chip, pit, crack, scratch, etc.	Same as above	None visible

Micron Ex. 1029, p. 629
Micron v. YMTC
IPR2024-00911

TABLE 6
Typical photomask requirements for manufacturing

Item	Purpose	Typical specification
Reticle		
Reticle ID	Ensure proper device levels and test patterns are fabricated	Vendor code, device name, level code, part no., reticle lot no., serial no., alignment mark no.
Reticle material	Material specification	Cr/Cr_2O_3 or $Cr_2O_3/Cr/Cr_2O_3$ on quartz
Reduction ratio	Specify reduction ratio	Per stepper specification
Pattern defect	Define defect types (contaminant, opaque spot, hole, excess Cr, Cr line intrusion and scratch) and sizes that affect printing results	None $\geq$ 2 μm
Registration type	Specify needs for position accuracy	Position accuracy $<$ 0.5 μm
CD criteria	Control reticle CD variation within photomask and mask to mask	Per design CD tolerances
Tone	Specify tone convention	Normal: black Inverse: white
Alignment mark and chip coordinate	Align photomask to stepper	Per design and stepper specifications
Border line width	Define actual printing area on wafer	6 mm from edge
Pellicle		
Pellicle material	Specify pellicle source material	Per vendor specification
Printable particle	Define minimum defect size that affects printing	$\geq$ 40–60 μm

TABLE 7
Typical DIW requirements for manufacturing

Item	Typical specification	Item	Typical range
Resistivity	> 18 Megohms	Inorganics	
Particles	< 0.1/mL at $\geq$ 0.1 μm	Ions	< 0.1 ppb
Total organic carbon	< 10 ppb	(Cl, F, Br, NH_4, NO_2,	
Dissolved oxygen	< 20 ppb	NO_3, PO_4, SO_4, etc.)	
Silica	< 1 ppb	Metallics	
Bacteria	< 1 colony/mL	(Al, Ca, Cr, Cu, Fe, K, Li, Na, Ni, Pb, Zn, etc.)	< 0.05 ppb per element

Micron Ex. 1029, p. 630
Micron v. YMTC
IPR2024-00911

Indirect material considerations

There are literally hundreds of materials used indirectly in a wafer fab. The criticality of those materials is reduced compared to that for direct materials. It is impossible to discuss them all in this chapter owing to their large quantity and wide range of variation. Readers should make selections based on their own needs and common cleanroom practices.

11.2.4 Specification Systems and Operation Procedures

To handle a complicated process and to control wafer fab manufacturing, maintenance of a stable operation is very important. Stability can be viewed from both the hardware (equipment/process) and the software (system/procedure/people) points of view. This section addresses the issue of systems and procedures.

System vs. people

What are the systems and procedures? Systems and procedures are ways to describe the sequences, requirements, and standards of how to do things.[16] For wafer fab operation with a high volume of output, everyone in the organization must know the detailed standards and what to do at all times. In this way mistakes are kept to an absolute minimum and deviations, if they occur, are realized early, preventing large losses.

There are two opposing schools of thought regarding systems and procedures. One believes that people's stability is not important so long as the systems and procedures are well-defined and implemented. This approach is possible for an operation with a limited number of products and rigorous management. The other school believes that the operation's stability comes from people's stability, because experience and judgment are required routinely and standards are sometimes difficult to define. Wafer fabs employing people with extensive experience and long tenure (10 years plus) find less need for detailed systems and procedures. Typically, wafer fab manufacturing requires both the well-defined system/procedure and experienced people to produce consistent, quality products; the system and people complement each other.

Specification system structure

Figure 4 shows the structure of a typical hierarchical specification system with examples. It starts with company policy and guidelines as the highest level and ends with standard operating procedures (SOPs) as the lowest level. The lower-level specifications must be in accord with higher-level specifications. For example, all equipment must have a calibration procedure in the manufacturing specification (lower-level specification) that is consonant with the calibration requirement in the quality specification (higher-level specification). Since the higher-level specifications provide the guiding principles for all operations, they can lead to an excellent company culture if they are properly established or be detrimental to the company's long-term existence if improperly set. Management must exercise extreme caution in defining such high-level specifications.

Micron Ex. 1029, p. 631
Micron v. YMTC
IPR2024-00911

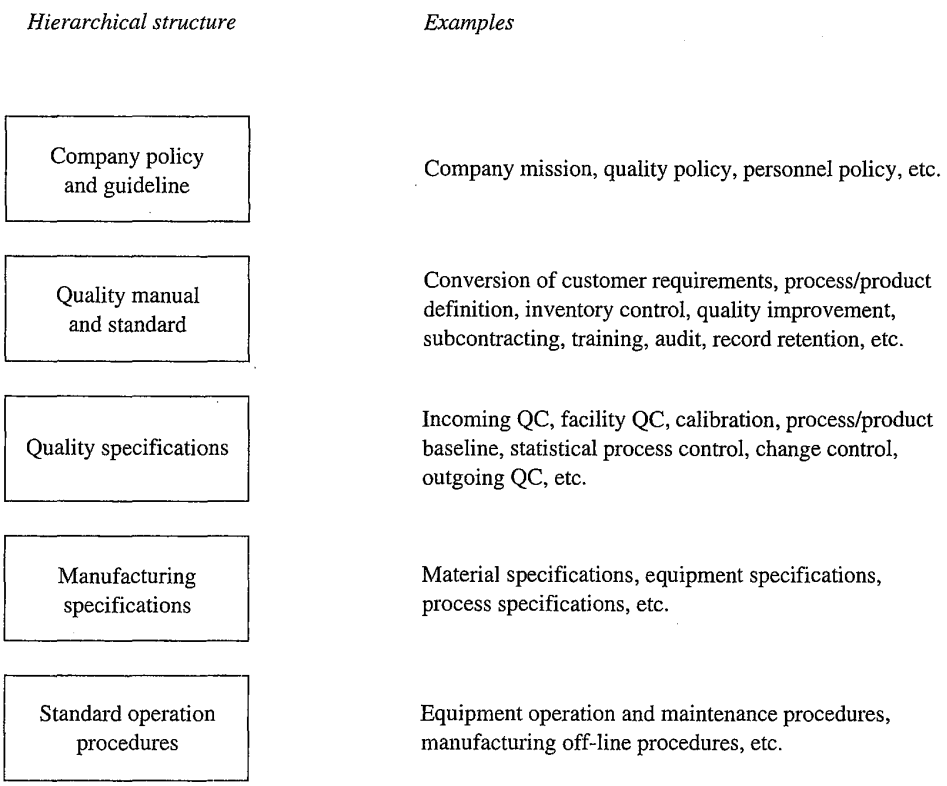


FIGURE 4
Illustration of a typical hierarchical specification system.

Along with a good system structure, proper system implementation requires a hierarchical sign-off control procedure. The sign-off not only acknowledges the specification but also provides a commitment from everyone to perform in agreed ways to carry out the requirements. For any high-level policy/guideline, top management is involved in the sign-off prior to the announcement to company employees. For other specifications, the sign-off follows the sequence: initiator, initiator's functional managers, other related functional managers, and the quality control (also quality and reliability assurance) manager. The QA manager is responsible for the specification activation. The fully signed-off and activated specification is then distributed to all involved personnel for immediate implementation.

Specification format

Since the specification/procedure defines the operational framework, its content needs to be detailed and clear so that it can be easily understood and followed.

Table 8 shows a typical formal content of a specification/procedure. Basically, it should describe in detail why (the purpose) the specification is needed and what will be done, how, by whom, and when. The scope of the specification and references should be given. Common words should be used for easy understanding and clarity because the specification is intended for all levels of the organization. An alternative format is to include a flow diagram for the cover page indicating the sequence of the

Micron Ex. 1029, p. 632
Micron v. YMTC
IPR2024-00911

TABLE 8
Specification contents

	Format	Designating
Purpose	To state the purpose of the spec/procedure	Why
Scope	To define the applicability of the spec/procedure	—
References	To list specs/procedures related to this spec/procedure	—
Definition	To define terminologies to be used in the spec/procedure	—
Requirements	To present a detailed description of what, how, who, where, and when to do for the spec/procedure	What, how, who, where, when
Others	To indicate other requirements not included above, e.g., supporting documents/forms, record retention, etc.	What

procedures. Developing a flow diagram before writing the actual specification can be a good practice because the flow diagram clarifies the procedure.

Important systems

In Fig. 5 several important systems needs for wafer fab manufacturing are illustrated for the input, process, and output stages. For the input systems, the incoming and facility quality control ensure the proper incoming quality of all the materials and infrastructure environments (Sections 11.2.3 and 11.3.2). For the process-related systems, specifications for sustaining operations and process control are important. To sustain operations, one must identify and control the process/product baseline (Section 11.3.6), changes from the baseline (Section 11.4.3), and baseline abnor-

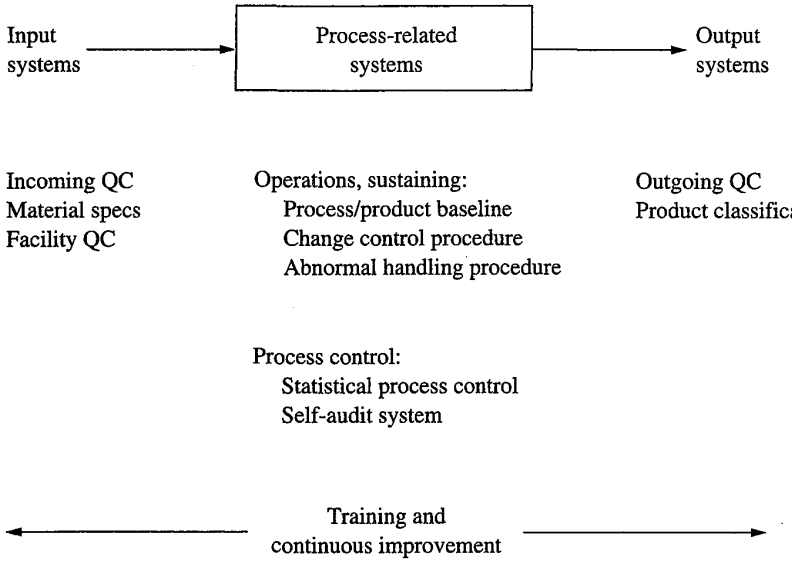


FIGURE 5
Schematic diagram for important operation specifications.

Micron Ex. 1029, p. 633
Micron v. YMTC
IPR2024-00911

malities. For process control, a statistical process control (SPC)[17] self-audit program is commonly used to control drift in equipment and process parameters, as well as workers' adherence to procedure, in real time (Section 11.4.3). For output systems, the product classification and outgoing quality control are categorized for various customer needs and guarantee quality products for continuous customer satisfaction (Sections 11.3.5 and 11.4.3).

11.2.5 Metrology and Analytic Capability

Both measurement and analytic capability are required for wafer fab manufacturing. Inaccurate measurements, particularly of critical parameters, are detrimental, especially in high-volume production, and can have a large financial impact. Analytic capability, on the other hand, is like the eyes and ears of the wafer fab. It can serve either as a routine monitor to prevent problems or as a problem-solving tool. However, the establishment of those capabilities is a balance between capability and cost. In this section, several metrology and analytic capabilities for wafer fab manufacturing applications are recommended. If some of the suggested capabilities or, even better, more advanced analytic capabilities are available in the city where the wafer fab is located, the wafer fab can minimize its own cost yet have access to the capability. Sometimes, an independent source of analysis to verify internal analytical results is also locally available. This is the subject of infrastructure (Section 11.2.6).

Wafer fab metrology considerations

Accuracy and consistency are two key factors for manufacturing metrology considerations. Process control is not possible without accurate measurement. To achieve accuracy, absolute standards certified by a national body on calibration are required.[18] In Table 9, metrology needs along with the available absolute standards are provided for typical wafer fab manufacturing metrology applications.[18–21] For advanced processes, absolute standards are not available in several cases as shown in Table 9. If we use thickness measurement as an example, a gate oxide thickness of less than 100 Å with variations less than ± 10 Å are commonly required. Absolute standards in this region (both thickness and variations) are not yet available. In this situation, a very thorough characterization is necessary using reference standards and other characterization means. For the gate oxide, the thickness is characterized, using gate oxide produced at the pilot line, with ellipsometer, transmission electron microscopy (TEM), and electrical measurements. Wafers with known thickness are then transferred to the wafer fab and used as the reference standards for calibration (or correlation) using similar characterization means. Once all the results are confirmed, the new ellipsometer, which will be used as a manufacturing metrology tool, is considered calibrated. It is, of course, preferable that this ellipsometer has already been calibrated against the available thicker absolute standards and the extrapolated results match the actual data using the thinner reference standards. This kind of approach, especially for metrology without an absolute standard, is essential for wafer fab start-up (Section 11.3). Where absolute standards are available, the metrology instruments must be calibrated against them directly.

Micron Ex. 1029, p. 634
Micron v. YMTC
IPR2024-00911

TABLE 9
Typical wafer fab metrology instruments

Equipment	Purpose	Standard	Capability
Thickness			
Ellipsometer	To measure thickness and refractive index of oxide and nitride films	500 Å–1 μm	SiO_2: 10 Å–6 μm Si_3N_4: 10 Å–30 μm
Nanospec®	To measure thickness of poly Si, oxide, nitride films	500 Å–1 μm	25 Å–20 μm
Alpha step®	To measure film thickness	400 Å–25 μm	Low range < 13 μm High range > 13 μm–< 300 μm
Dimension			
SEM	To measure CD	0.5–5 μm at 1–20 K magnification	0.1–100 μm
Electrical			
Resistivity meter	To measure the film sheet resistance	0.1–150 ohm/sq cm	5 milli–5 meg ohm-cm
Carrier lifetime tester	To measure minority carrier lifetime in Si	Reference standard	1–5000 μsec
C-V (capacitance-voltage) tester	To measure device *C-V* characteristics such as gate-oxide V_T (threshold voltage)	1–500 pF	0–2000 pF
Particle/Defect			
Particle Counter	To measure particles on bare Si wafers	Reference standard	0.07–900 μm
	To inspect defects on patterned and nonpatterned wafers	Reference standard	0.2–10 μm
Defect detection and classification tool	To inspect defects on patterned wafer	Reference standard	0.15–10 μm Ability to classify defects
Inspection microscope	To inspect surface defects on grown/deposited films	Reference standard	0.15–100 μm
Others			
Stress measurement	To measure stress of films and finished wafers	Reference standard	Thickness: 400–1000 μm Bow/warp: ±350 μm
X-ray fluorescence	To measure boron and phosphorus content in PSG and BPSG film	Reference standard	B: 0–1%/1000 Å P: 0–1.5%/1000 Å
Fourier transform infrared spectrometer	To measure boron and phosphorus content in PSG and BPSG film	Reference standard	1–10 wt %
Stepper alignment inspection tool	To measure stepper alignment accuracy	Reference standard	0.5–20 μm

Micron Ex. 1029, p. 635
Micron v. YMTC
IPR2024-00911

The consistency of the metrology instruments is the next key manufacturing consideration. To maintain consistency, periodic monitoring using the absolute standards (or reference standards) is necessary. Once the procedure and sampling frequencies are determined, one can control the consistency through statistical process control (SPC).[17] In Fig. 6 two charts are shown to demonstrate metrology accuracy using SPC. The example in Fig. 6*a* illustrates ellipsometer measurements with a thickness of 504 Å and variations of ±4 Å against an absolute standard of 506 Å. In Fig. 6*b* are measurements using the same ellipsometer with a 56-Å thickness and variations of ±4 Å using a reference (not absolute) standard of 56 Å. The variations shown in Fig. 6 arise primarily from the standard itself and from handling, i.e., placing the wafer on the stage, moving the stage to five preset locations, etc. The consistency of the metrology, however, can be maintained through SPC to keep manufacturing processes under control.

Other commonly needed analytic capabilities for wafer fab

Owing to process complexity (individual processes are dependent on each other) and long cycle time, various process monitors, as well as quick problem-solving capabilities, are essential when problems occur. The best approach is to equip the wafer fab with the analytic capabilities required most often.[11] Tables 10 and 11 list the commonly used analytic capabilities. Table 10 details the inspection needs for the wafer and reticle. The instruments listed in Table 10 are located inside the cleanroom so that the inspected wafer/reticle can return to production immediately without contamination. Of course, any inspection step can potentially introduce contamination, such as particles due to handling. Instruments listed in Table 11 are typically located outside the cleanroom even though samples may be taken from equipment inside the cleanroom. Sampling procedures and preparation are very important in advanced sub-ppb level analysis since any small handling error can introduce contamination and result in a large analytic error.

For problem solving and yield improvement,[9,11] the analytic capabilities in Tables 11 and 12 are needed. Instruments in Table 11 are used to identify potential contamination from the facility or wafer process. Instruments in Table 12 are tools for failure analysis of the electrical and physical properties of the finished or partially finished chips or wafers. Some of the most common applications using the analytic capabilities in Tables 11 and 12 are these:

1. Process control and improvement
   - Equipment/process release
   - DIW/chemical point-of-use monitor
   - B and P composition control in BPSG films
   - Cross-sectional process structural characterization
2. Problem solving
   - Facility quality problems
   - Abnormal process or low yield analysis
3. Incoming quality control (IQC)
   - IQC for direct materials (Section 11.2.3)
   - Evaluation and qualification of indirect materials (Section 11.2.3)

Micron Ex. 1029, p. 636
Micron v. YMTC
IPR2024-00911

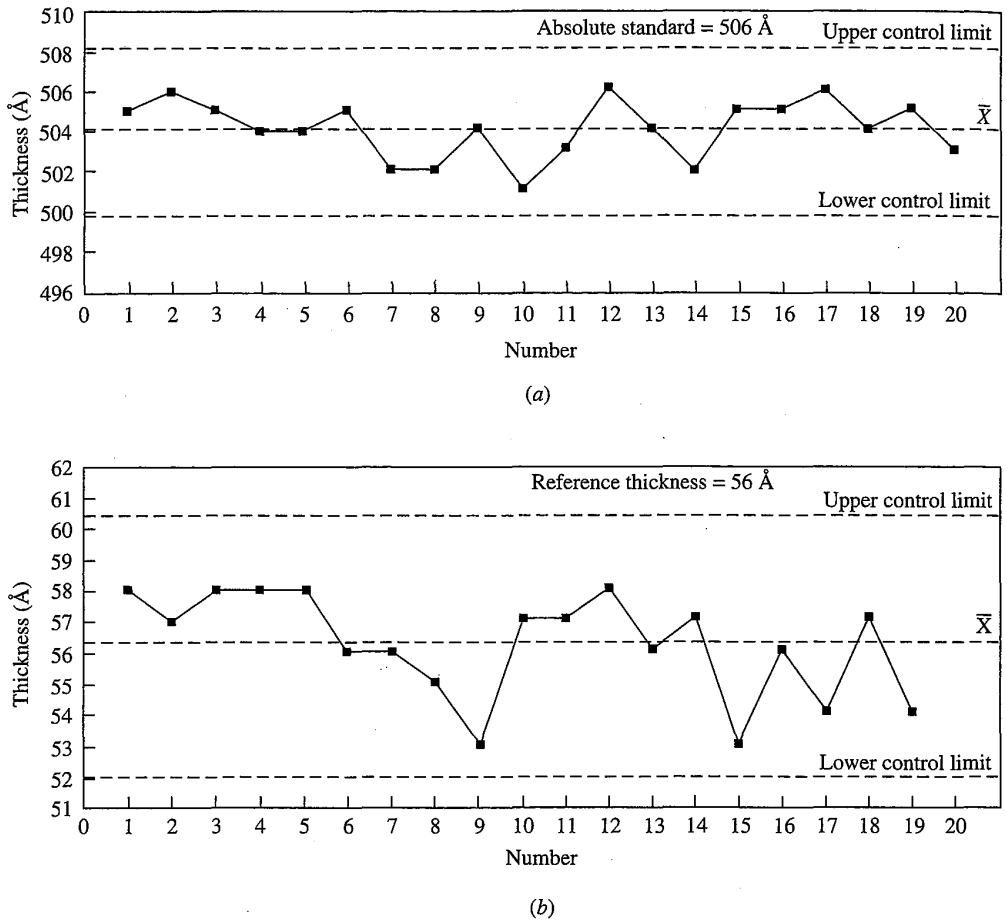


FIGURE 6
Ellipsometer metrology consistency indicated by SPC charts for (*a*) absolute standard with 506 Å and (*b*) a reference thickness of 56 Å.

Volume testing is a unique characteristic of these applications. Typically, hundreds or thousands of analyses are performed on a monthly basis; these are intended for everyday support of continuous manufacturing with quick turnaround time, sometimes in a matter of hours. Advance characterization techniques[9,11] in addition to those listed in Tables 11 and 12 are definitely helpful when the need occurs. Once again, it is preferable to have these available locally from outside services (Section 11.2.6).

11.2.6 Other Infrastructure Considerations

Even though a wafer fab may be equipped with many supporting facilities and analytic capabilities, it needs still other services because of the complexity and the nature of technological advancement as well as for daily manufacturing logistics. Many of these services will not be cost-effective if they are owned by one wafer

Micron Ex. 1029, p. 637
Micron v. YMTC
IPR2024-00911

TABLE 10
Analytic capability for incoming wafer and reticle

Equipment	Purpose	Capability
Wafer		
Flatness tester and resistivity meter	To measure characteristics of the incoming wafer such as bow, warp, thickness, and resistivity	THK: 400–1000 μm Bow/warp: ±350 μm Resist.: 0.2–200 ohm-cm
Fourier transform infrared spectrometer	To measure the content of oxygen and carbon in silicon	O: 2–50 ppma C: 0.5–15 ppma
Particle counter	To measure particles on bare Si wafers	0.07–900 μm
Reticle		
Photomask defect inspection tool	To inspect defects on reticles	Defect size > 0.25 μm
Photomask particle and defect counter	To inspect particles on reticles	Particle size 0.5–60 μm

fab. It makes more sense to have those requirements satisfied by available local services, that is, the infrastructure. On the other hand, one wafer fab may not have enough work to support the service profitably. Wafer fabs and local infrastructure hence rely on each other. In any case, since wafer fabs, such as those producing advanced DRAM, are the driver, they play a leading role in local infrastructure development. Table 13 lists some of the basic infrastructure considerations required to support manufacturing. Of course, infrastructure availability is not simple because it is related to individual business needs and often is subject to local government policy and regulation. However, locations with wafer fabs should aim to provide the supporting infrastructure as suggested in Table 13.

11.3 MANUFACTURING START-UP TECHNOLOGY

The start-up of a new wafer fab is the most important, exciting, and work-intensive stage of a wafer fab history. It sets the foundation of the manufacturing facility. It also establishes the tone of future operations. A good start-up establishes positive reinforcement between manufacturability (process capability and people's practice) and continuous improvement. In this case, all resources contribute to capability improvement instead of troubleshooting of problems. Conversely, a poor start-up pushes the wafer fab into a negative cycle where one is constantly fighting to solve the remaining problems, which keep on manifesting themselves because of incomplete start-up checks. The start-up determines the fate of the wafer fab: either *designed-in manufacturability* or *designed-in problems.*

In this section, we introduce a start-up approach that has successfully brought in more than one advanced 0.5- and 0.8-μm wafer fab. Figure 7 is a schematic diagram

Micron Ex. 1029, p. 638
Micron v. YMTC
IPR2024-00911

TABLE 11
Suggested wafer fab chemical analysis capability

Equipment	Purpose	Capability
Atomic Absorption spectrometer (AAS)	Detect cationic elements in liquid samples Analyze 17 common elements in samples (Al, As, B, Ca, Co, Cu, Cr, Fe, K, Mn, Na, Ni, P, Se, Si, V, and Zn)	Sub ppb to ppm
Inductively coupled plasma mass spectrometer (ICP/MS)	Determine 70% of the elements in periodic table in liquid samples Analyze more than 30 elements simultaneously	Sub ppb to ppm
Ion chromatograph (IC)	Detect trace anionic impurities in DIW or process water Analyze Cl, Br, NO_3, SO_4, PO_4, etc.	Sub ppb to ppm
Gas chromatograph		
TCD (Thermal conductivity detector)	Determine impurities in gases	ppm
DID (Discharge ionization detector)	Determine impurities in gases including H_2, O_2, N_2, CO, and CO_2, hydrocarbons, noble gases, and inorganic acid, ammonia vapors, etc.	ppb
Gas chromatograph mass spectrometer (GC/MS)	Identify all components in a gas system	0 to 650 amu
MSD (Mass detector)	Analyze gaseous and volatile liquid materials Provide qualitative information for unknown compounds	
Trace O_2 analyzer	Monitor trace oxygen content in gases	2 to 100 ppb
Trace moisture analyzer	Measure the trace moisture content in gases	10 ppb to 20 ppm
TOC analyzer	Detect total organic carbon Measure both the temperature and resistivity of DIW Monitor the organic contamination source in DIW system	Sub ppb to ppm
Titrators	Measure the components in liquid solution by titration method	ppm to at %
	Measure trace moisture content in organic solvents	ppm
	Check consistency of liquid acid mixtures	ppm to at %
Liquid particle counter	Measure particle counts both in DIW and chemicals (including corrosive solvents)	0.1 to 0.5 μm
Bacteria counter	Check total bacteria and particulate counts in DIW or liquids	Per DIW system specification

Micron Ex. 1029, p. 639
Micron v. YMTC
IPR2024-00911

TABLE 12
Suggested wafer fab failure analysis (FA) capability

Equipment	Purpose	Capability
Physical FA		
Wet deprocessing (solvent/acid hoods and clean bench)	Strip device layer by layer Decorate defects for observation	Protective overcoat to substrate
Dry deprocessing (plasma etcher-RIE)	Remove certain device layers	Protective overcoat, metals, layers
Microsection station	Prepare device cross section of failed locations Device structural analysis	Identify failed locations
SEM and sputter coater	Detect microdefects, morphology, and dimensions	20 to 300 K magnification 150 Å resolution
Energy dispersive X-ray (EDX)	Determine material constituents qualitatively or quantitatively	Atomic %
Microscope	Observe gross defects, morphology, or structures	5–3500 magnification
Electrical FA		
Device tester	Test device AC/DC parameters, determine electrical failure phenomena	Test pattern and algorithm generation
	Test the sensitivity of failure under various electrical operation conditions	Test full device operations and stress conditions
Bias and temperature test station	Test sensitivity of failure under bias and temperature stresses	0–150°C 0–50 V
Liquid crystal analysis	Identify failed location through hot spot Detect smaller DC, AC, and ICC device failures	Leakage > 100 μA
Emission microscope	Identify dielectric breakdown and transistor failure through detection of photoemission	Leakage > 100 μA
Laser spot size	Isolate metal and poly-Si lines for DC/AC testing	1 μm minimum laser spot size
Internal probe	Analysis of failure unit internal signal and waveform	

illustrating the major activities of wafer fab start-up. Some details of these activities are discussed in this section.

Because of the complexity of start-up, a very comprehensive project schedule and tracking plan must be developed first. Milestones and subtasks for each of the activities described in Fig. 7 should be identified. The interrelationships among activities also need to be comprehended to identify potential bottlenecks. In our experience, target time for starting wafer production is the major milestone. Activities

Micron Ex. 1029, p. 640
Micron v. YMTC
IPR2024-00911

TABLE 13
Infrastructure considerations for manufacturing

Infrastructure service	Considerations
Laundry	Facility with DIW and cleanroom Low-Na and I-Cl raw materials (detergent) Thoroughly rinsed to remove all impurities Handling procedures for drying and packaging for particle control Anti-static bag
Quartz work	Facility with DIW and controlled environment Quartz cleanup capability Precision laser work, control, and measurement capabilities Stress annealing and inspection capabilities Established handling and control procedures, e.g., IQC, flame welding, outgoing inspection, and packing Experienced workers for precision high-temperature works
Calibration	Certification by national body Wide range of calibration capabilities Established procedures to ensure calibration accuracy and repeatability Issuance of proper certificate of calibration On time and fast turnaround Low cost Capability to develop and set up new calibration needs
Analytic service	Wide-range analytical availability to analyze physical, chemical, electrical, structural, and surface properties of materials Sufficient instrumentation with proper detection sensitivity Established procedures for calibration, sample preparation, and measurement accuracy and repeatability assurance Experienced analysts in different fields Reports issued with data analysis and conclusions On-time and fast turnaround Low cost Ability to develop and set up new analytical requirements
Chemical and gas supply (minimum requirements)	Facility with DIW, warehouse, and analytical lab Capability for purification, drum/cylinder cleaning, and drum/cylinder filling Capability for sampling and analysis by experienced analysts Established handling procedure, tracking, analysis, and outgoing control procedure Waste and used drum/cylinder treatment Ability for service, fast response, and problem solving
Cleanroon furniture	Supplier capability to clean up oil, grease, and fingerprints on finished furniture Good metal workmanship Supplier ability to pack cleaned furniture Electro-polished surface finishing preferred

Micron Ex. 1029, p. 641
Micron v. YMTC
IPR2024-00911

FIGURE 7
Wafer fab start-up activities.

Micron Ex. 1029, p. 642
Micron v. YMTC
IPR2024-00911

required before the start, primarily facility construction and the acquisition of utilities and equipment, can be derived. Activities after the start of wafer production, such as the process release, product qualification, and volume ramp-up plan can also be determined.

11.3.1 Specification Generation

The first start-up activity, as shown in Fig. 7, is to generate various specifications from product to process to production, to meet the final manufacturing goal. This task is very important because all future activities are bound by these specifications. Our intention here is not to address the details of how the specifications should be derived. However, a generic methodology,[22] shown in Fig. 8, illustrates the process. For deployment, the characteristics of each item related to each block in Fig. 8 are decided first. Target values can then be determined based on their relationships. This process starts with understanding customer needs and continues until production requirements are defined. Details in other chapters of this book can be used for product, process, equipment, material, and facility considerations in the deployment process.

For a manufacturing wafer fab start-up, the more desirable situation is that only production planning (Fig. 8) is needed. However, owing to the short semiconductor product cycle time, sometimes it is inevitable that both process planning and production planning (Fig. 8) are required for start-up. In both cases, the following items should be given consideration in generating specifications:

- Process specification must accommodate design tolerances, demonstrated equipment performances, and manufacturing variations for good process margins.
- Equipment, material, and facility specifications should consider proven technology, industrial availability, and suppliers' continuous support and problem-solving capability to facilitate fast start-up and ramp-up with minimal resources.
- Connection and material specifications, and their quality standards with respect to equipment and facility, should be consistent so that they are not overspecified yet cause no loss of quality at the transition between the two.
- In all cases, the lessons learned from previous process and production experience should be incorporated as much as possible to eliminate costly known mistakes.

11.3.2 Utility Requirements and Control

For a start-up wafer fab, facility construction of the building and the utility systems (cleanroom, chemical, gases, and DIW distribution systems) starts first. A modern advanced wafer fab contains miles of pipes, thousands of valves, thousands of filters, and thousands of connections, weldings, and installations. All have high purity and cleanliness requirements.[23–25] The construction of high-purity utility systems can take over half a year. If work is not done carefully in terms of the system certification and cleanliness requirements, the results are either delayed completion or sacrificed system quality and reliability. In Chapter 1, details of the utility systems' design and specifications are discussed. In this section, we discuss the certification[26] and

Micron Ex. 1029, p. 643
Micron v. YMTC
IPR2024-00911

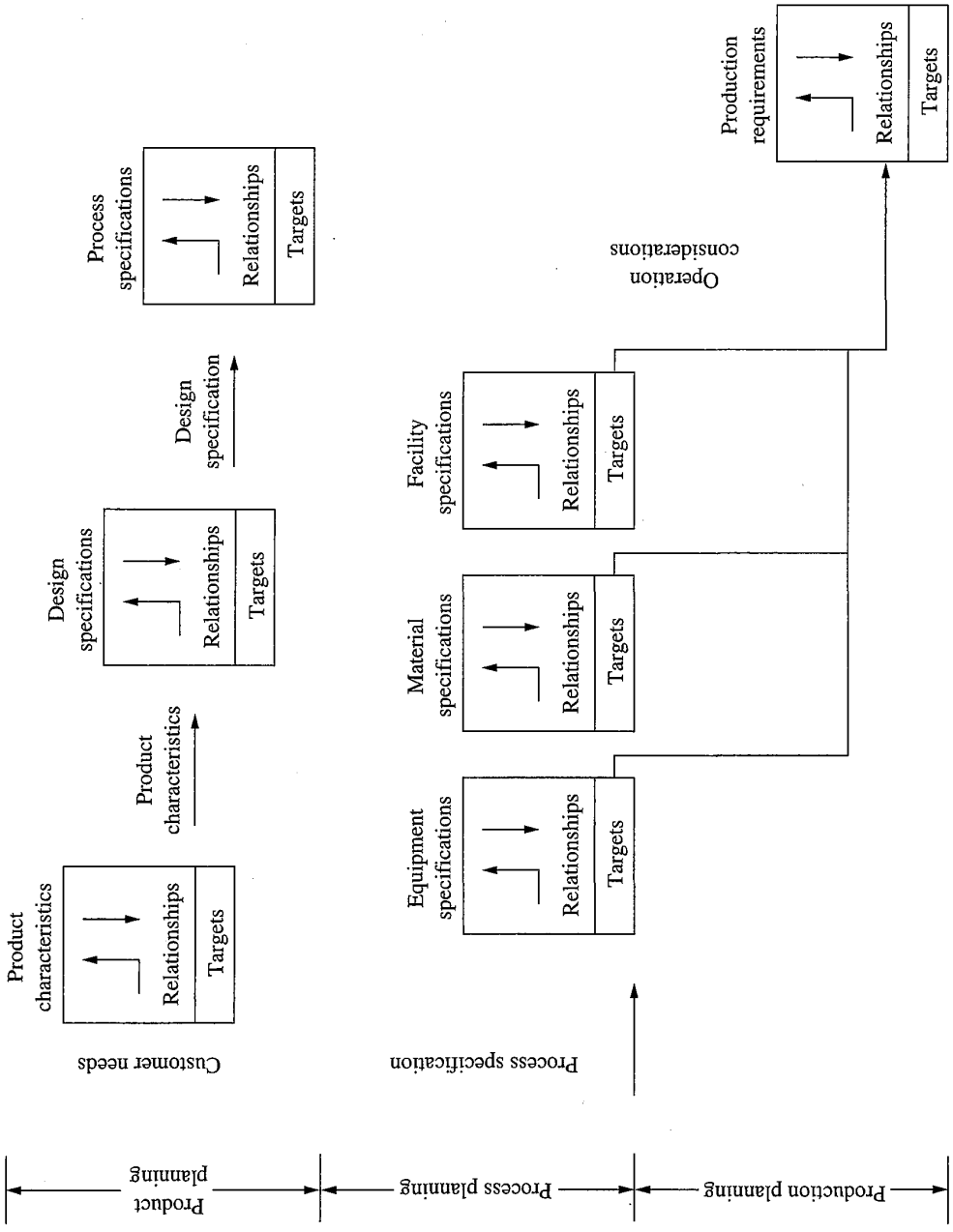


FIGURE 8
Deployment for specification generation—planning for quality.

Micron Ex. 1029, p. 644
Micron v. YMTC
IPR2024-00911

control aspects of the utility systems. The suggestions in Section 11.3.1 are important to ensure fast system start-up and long-term stable system performance. Additional considerations are as follows:

- Detailed checklists for parts (pipe, valve, filter) and works (welding, pipe cleaning, filter installation) are prepared prior to construction based on experience and system requirements.
- Selection of parts and their material requirements is based on proven quality and reliability performance and is consistent with equipment specification.
- There is continuous worker training on procedures and cleanliness requirements.
- Routine and thorough checks by the construction teams using the prepared checklists and corrective actions are implemented on a real-time basis.

Utility systems certification

The certification of the utility systems ensures that the minimum process requirements of the utility supplies are met at time zero to eliminate supply variations at start-up. Certification will not be successful if the facility construction itself is poor. It is most important that certification be viewed only as the last stage of the construction quality control, whereas the main body of the quality control is actually done during the construction. If this is not the case, costly reworks, extended systems flushing (DIW and chemicals), and a delay of the supply operation are typical consequences. Another aspect is to certify each subunit of the system on-site on a real-time basis so that utility systems–related causes can be readily eliminated, which will avoid subsequent troubleshooting.

Figure 9 is a flow diagram illustrating the sequence for utility system certification. Two stages are involved: integrity tests and functional tests. The integrity tests are primarily to check that the system's mechanical strength under pressure, vacuum, and fluid conditions meets the system's mechanical design criteria. The functional tests are aimed at checking the system's quality performance in meeting system quality standards. In Table 14 are listed several parameters for each utility system. The actual acceptance value for each parameter is not listed. However, specifications discussed in Section 11.2.3 and Chapter 1 can be used for a particular application.

Maintaining facility systems stability

In addition to the mechanical and functional (quality) systems certification, the reliability of the systems' performances is also important. An unreliable system causes a great deal of supply fluctuation and a system supply shutdown in the worst case. Continuous monitoring of all utility systems ensures consistent system supplies for manufacturing. In today's wafer fab, the majority of the monitoring efforts are made automatic by using the continuous quality control (CQC) approach (Section 11.2.2). CQC consists of hardware (sensors and instruments) and software (data collection/analysis and control actions) allowing real-time and continuous monitoring, therefore minimizing human effort. The parameters not yet covered under CQC are marked with a "†" in Table 14. Therefore, they require periodic sampling and analysis by designated and trained personnel. In Fig. 10 an example of Fe impurity

Micron Ex. 1029, p. 645
Micron v. YMTC
IPR2024-00911

FIGURE 9
Schematic diagram of facility certification procedure.

Micron Ex. 1029, p. 646
Micron v. YMTC
IPR2024-00911

TABLE 14
Parameters for facility system certification

		Gas			Chemical	
	Cleanroom	**Plant**	**Specialty**	**DIW**	**Acid**	**Solvent**
Integrity check	Filter grids and filter gross leak	Visual inspection	Visual inspection	Visual inspection	Visual inspection	Visual inspection
		Leak integrity, vacuum and pressure tests	Leak integrity, vacuum and pressure tests	Leak integrity, pressure decay with N_2	Leak integrity, pressure decay with N_2, DIW, and actual chemicals	Leak integrity, pressure decay with N_2
Functional certification	Clean classification Temperature Humidity Air flow velocity Air flow laminarity Vibration† Room pressure† Lighting level† Sound level† Air quality† Make-up air flow and recovery	Moisture Trace O_2 Particle† Impurities (Table 4)	Moisture‡ Trace O_2‡ Particle†	Temperature Pressure Resistivity D.O.§ TOC¶ Particle† Bacteria† Silica Impurities (Table 7)†	Particle using DIW Particle using actual chemical† Impurities (Table 3)†	Particle using N_2 Particle using actual chemical† Impurities (Table 3)†

† Parameters not monitored by CQC.
‡ Using plant gas, e.g., N_2.
§ D.O.: Dissolved oxygen
¶ TOC: Total organic content

Micron Ex. 1029, p. 647
Micron v. YMTC
IPR2024-00911

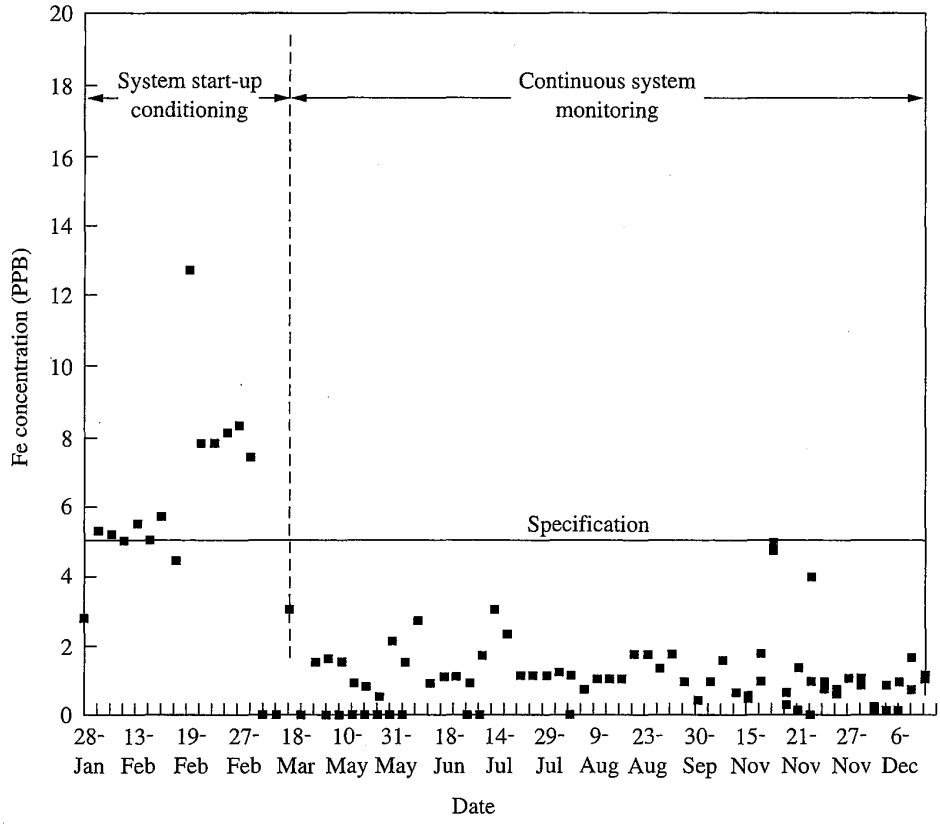


FIGURE 10
Measured level of Fe in HCl at the output filters of the chemical-dispensing system.

content in HCl at the point of delivery from the HCl distribution system is shown on a time scale. The system was under start-up conditioning before March. Leaching of Fe from the distribution system and contamination from HCl drums were observed. These problems were corrected and the system was certified on March 18 to meet the Fe specification of 5 ppb and maintained below 3 ppb. On November 17, a high Fe level of 5 ppb was again detected and traced back to contamination of the high-purity tubing connecting the chemical drum to the distribution system. The contamination was cleaned out, and system performance was restored. In other cases, one may not be able to resolve problems in facility systems with this efficiency and effectiveness. The effort, however, must be made to ensure that systems are under control for stable and consistent supplies.

11.3.3 Equipment Checks

The performance of all equipment is extremely important to a successful start-up. Equipment functional performances must first meet the targeted process requirements to demonstrate the capability. Subsequently, equipment stability and reliability are key to fast volume ramp-up and in-control manufacturing. To ensure good equipment performance, the following checks can be instituted.

Micron Ex. 1029, p. 648
Micron v. YMTC
IPR2024-00911

Equipment source inspection

The concept of source inspection (inspection conducted at the vendor's site before shipment) is to eliminate problems prior to the equipment shipment. Table 15 suggests a checklist for source inspection. The purchasing specifications are the basis for the acceptance criteria for items 1–5 in Table 15. Items not meeting the acceptance criteria must be identified during source inspection and corrected prior to shipment.

Equipment cleanliness

Despite the high level of cleanliness (better than class 10 at 0.12 μm; see Chapter 1) required for today's advanced wafer fabs, equipment may carry gross contamination (dirt, grease, and fingerprints) into the cleanroom if it is not cleaned properly. Cleaning the equipment before moving it into the cleanroom is one way of eliminating contamination sources.

There are two main causes for equipment contamination: the equipment fabrication process and packaging. During the former, components and subassemblies can be contaminated by constant handling with bare hands, cutting/drilling/welding, or having dirty parts initially. To eliminate these contamination sources, a vendor must establish a good equipment-cleaning procedure, covering component and subassembly cleaning, handling parts with gloves, removing contaminants (dirt, drill/cutting debris) constantly, and continuously wiping down grease and oil. Subsequently, the exterior of the fabricated equipment should be wiped clean and checked for defects in workmanship, such as painting scratches, mechanical misalignment, and corrosion. A 5% solution of IPA (isopropyl alcohol) in DIW is commonly used for cleanup. For packaging, double or triple wrapping with precleaned plastic wrapping materials is used. Double (or triple) wrapping isolates outside contaminants (crate materials, dirt, dust) from the cleaned equipment. Equipment cleanliness is merely a good common sense practice rather than advanced technology. Most major vendors are equipped with Class 10,000 cleanrooms for assembly and have defined cleanliness and packaging procedures.

Equipment integrity

Once the equipment has been moved into its proper location inside the cleanroom, the next step is to assemble the equipment and make the utility connections. The equipment assembly includes mechanical assembly of the main body, leveling, connection to control units and vacuum pumps, and connection to stand-alone supply units (rf generator, ozone generator, cooling units, etc.) specific to that equipment.

Utility systems to be connected include electricity, chemicals, gases, DIW, process cooling water, plant vacuum, exhausts, and drain/waste-handling systems for processing and removal of the used/residue materials and by-products. The utility hookup should be consistent in materials and connector type between utility system and equipment, as shown in Table 16, to maintain high-purity supplies of chemicals, gases, and DIW.

The purpose of the equipment integrity check is to ensure the mechanical and point of use (POU) quality. The mechanical integrity check emphasizes visual

Micron Ex. 1029, p. 649
Micron v. YMTC
IPR2024-00911

TABLE 15
Equipment source inspection checklist

Items	Description
1. Equipment specifications	Reconfirm specifications
2. Utility connection	Utility connection types and dimensions Piping and connector materials, e.g., PVDF for DIW piping Welding requirements Exact locations of the connection
3. Cleanliness requirements	Cleanliness procedure and setup Cleanliness inspection before packing Packing requirements
4. Safety issue	Common safety features, e.g., emergency shutoff device Specific safety features, e.g., laser safety interlock, specialty gas (SiH_4, etc.) safety interlock
5. Basic equipment function checks	Mechanical functions, e.g., stage accuracy, wafer transfer mechanisms, safety alarms, etc. Process functions, e.g., stepper lens performances, deposited film uniformity on pilot wafers, etc.
6. Calibration methodology	Vendor's calibration standards and procedures Vendor-provided certificates, as for mass flow controllers and thermocouples Understanding of total equipment on-site calibration requirements
7. Work out detailed equipment on-site checklist and schedule	List detailed equipment check items and estimate time required for scheduling
8. Confirm equipment shipping/arrival date and vendor on-site support resources	Confirm equipment actual shipping date to targeted shipping date Arrange for vendor on-site support items

inspection and pressure/vacuum tests that use pressurized inert gas and also tests lines and chambers for proper connection and sealing. The POU integrity check is to ensure that the supply quality is maintained from supply source to point of use and also for process quality assurance. Table 17 provides some POU quality requirements for gases, DIW, and chemicals.

Equipment functional check

The purposes of the equipment functional checks are to verify the equipment's ability to meet process margin requirements and to ensure the equipment's reliability and capability to sustain operation with the required consistency and capacity.

Before running the tests, the equipment needs to be calibrated (Section 11.2.5) so that the equipment parameters (temperature, gas flow, pressure, power, etc.) can be

Micron Ex. 1029, p. 650
Micron v. YMTC
IPR2024-00911

TABLE 16
Typical materials and connectors used for equipment hookup

Utility system	Supply or drain	Materials	Connector type
Plant gas	S and D	316EL	VCR
(N_2, O_2, H_2, Ar, He)		316SS (nonprocess gas)	Swagelock (nonprocess gas)
Specialty gas	S and D	316EL	VCR
Hot/cold DIW	S	PVDF, PFA	Buttwelded union
	D	PP (cold), PVDF (hot)	Union
Chemicals			
Acid	S	PFA, PTFE	Teflon™ fittings
Solvent	S	316EL	VCR
Photoresist	S	304SS	—
Process cooling water	S and D	PVC	Swagelock/PVC
Plant vacuum	S	SS/PVC	Swagelock
Exhaust			
General and caustic	—	—	—
Acid	—	PP	Flange
Solvent	—	SS	Flange
Industrial waste	D	HDPE	Flange

controlled accurately. For new equipment, components such as thermocouples, mass flow controllers (MFC), pressure and vacuum gauges, etc. should be precalibrated at delivery and a certificate of calibration provided by the vendor. The whole equipment calibration is then performed on-site (using prearranged calibration standards) by the equipment vendor's field engineers. An example of equipment calibration is the stepper calibration for mechanical stages and optical/lens assembly.

Subsequent to calibration, the equipment is subject to verification for quality and reliability performances. Quality performance refers to equipment operation results associated with the process parameters at one instant of time. Reliability performance refers to failure characteristics of the equipment with time, which impacts the capacity and stability. With a deposition furnace as an example, illustrated in Table 18, film thickness and uniformity within a wafer, wafer to wafer, and run to run are parameters for quality performances. On the other hand, mean time between failures (MTBF), mean time to repair (MTTR), etc. are parameters for reliability performance. It is most important that equipment performance in both categories (quality and reliability) be verified to establish the foundation for good process margins and long-term manufacturability. In Table 19 several typical parameters are shown for functional verification of various types of equipment. Most of the reliability parameters for different types of equipment are similar to those listed in Table 18, and therefore are not repeated in Table 19.

Equipment operating procedure

Equipment operating procedure is not part of the equipment check but is a good practice to establish during the start-up stage. Clear operating procedures on equipment start-up and shutdown and loading and unloading procedures are minimal needs. Operating procedures for short-term (daily or weekly) and long-term (monthly or yearly) preventive maintenance[5,27] should be established after start-up

Micron Ex. 1029, p. 651
Micron v. YMTC
IPR2024-00911

TABLE 17
Typical equipment point of use requirements

Utility system	Check item	Requirement	Check Location†
Plant gases	Leak check		
	Pressure decay	As required	POU
	Helium leak check	As required	POU
	Particle count	< 1 particle/ft^3; particle ≥ 0.1 μm	POD
	Moisture	< 50 ppb	POD
Specialty gases	Leak check		
	Pressure decay	As required	POU
	Helium leak check	As required	POU
	Particle count	< 1 particle/ft^3; particle ≥ 0.1 μm using N_2 or Ar	POD
	Moisture	< 50 ppb	POD
	Oxygen	< 50 ppb	POD
	THC	< 50 ppb	POD
Deionized water	Leak check	Visually inspect for leaks when equipment is turned on	POD
	Particle count	< 2 per mL; particles ≥ 0.085 μm	POD
	Bacteria	< 1 colony per liter	POU
	Impurities	< 0.05 ppb per metallic element	POU
Solvents	Leak check	As required	POD
	N_2 purge	As required	POD
	Particle count	< 20 mL > 0.2 μm	POD
	Impurities	< 1 ppb for Al, Ca, Cu, Fe, K, Ni, Cr, Zn	POU
Acids	Leak check	As required	POD
	DIW Flush	As required	POD
	Particle count	< 20 particles/mL > 0.2 μm	POD
	Impurities	< 1 ppb for Al, Ca, Cu, Fe, Na, K, Ni, Cr, Zn	POU

†POU: Point of use at the equipment
POD: Point of delivery to the equipment

but no later than the completion of the capacity ramp-up. These operating procedures are not only for the manufacturing operation but also for verification and improvement of the equipment availability for capacity planning and monitoring. An example is provided in Table 20.

11.3.4 Process Characterization

Process characterization at start-up is aimed at establishing actual process performances[10,11,28] against the design criteria and pilot production references. If the process step has never been studied owing to process changes or the use of new equipment, the process should be *characterized* to understand the key control parameters of the process requirements. Characterization also ensures that the individual process

Micron Ex. 1029, p. 652
Micron v. YMTC
IPR2024-00911

TABLE 18
Typical equipment check parameters for a deposition furnace

Quality category	Reliability category
Temperature (set and ramp rate)	Equipment-dependent uptime (%)
Gas flow rate	MTBF (hrs) or MWBF (pieces)
Furnace pressure	MTTR (hrs)
Defect density (particle)	Component or subassembly
Film properties	Reliability, e.g., wafer transport mechanism
Deposition rate (Å/min)	
Thickness target	
Thickness and composition	Preventive maintenance (%)
Uniformity	MWBC (pieces)
Within wafers	Wafer boat
Wafer to wafer	Furnace tube
Run to run	
Resistivity	
Index of refraction	
Process yield loss (%)	
Max water scrap due to equipment (%)	
Furnace loading capacity & wafer size	
Maximum contamination level	

performances[10,11] meet design criteria for that process step and can be integrated[10,28] with other individual processes to meet overall device parameter requirements.[13] In this case, it is important that the process and equipment be characterized, as much as possible, at the vendor's location to minimize the unknowns during start-up. Most preferably, start-up process characterization should not be a *characterization* for new process/equipment but rather a *verification* of the process performances with respect to the process performances already achieved during pilot production. This is one approach for establishing a new wafer fab process baseline and fast start-up. Sometimes, the use of new equipment is inevitable so that running recipes for equipment parameters may not be transferable from the pilot line to production line. In this case, the performances of process output but not process recipe should still be maintained strictly compared to that at pilot running by fine-tuning equipment performances. Once again, this approach is to ensure no loss of process margin so that the equivalent pilot production device and yield can be achieved readily at the completion of start-up in the shortest time.

With this understanding of the characterization/verification of the start-up process, the release of a specific process, that is, the completion of its checks so that it can be moved to the production phase, becomes a matter of preparation and execution. In Fig. 11, a flow diagram illustrates the process release procedure. In it, the process release can be separated into two release steps: capability verification and stability check. The capability check is to verify that the process can perform the pilot production references (actual margin but not the spec) or the design criteria (process spec) if no references are available. The stability check, on the other hand, is to ensure that the process can be maintained without degradation over time after capability has been demonstrated.

Micron Ex. 1029, p. 653
Micron v. YMTC
IPR2024-00911

TABLE 19

Typical equipment functional check parameters

Photo

Stepper	Coater/Developer
1 Facility/safety check	1 Safety/facility check
2 Alignment (reticle/wafer)	2 Machine performance (uniformity, speed)
3 Illumination	3 EMO/interlock check
4 Stage/lens performance	4 Dispense/rinse cycle
5 Focus/leveling	5 Wafer handing
6 Wafer handling	6 Particle check
7 Particle check	7 Temperature/humidity
8 EMO/interlock check	
9 Process yield loss	

Plasma

Etchers	Ashers
1 Facility/safety check	1 Facility/safety check
2 Gas check (MFC)	2 RF/temp control test
3 MFC check	3 Leak check
4 Leveling	4 Strike plasma test
5 Purge/temperature/ pressure check	5 Wafer handling
6 Calibration forward power	6 Particle check
7 EMO/interlock check	7 Magnet check
8 Particle check	8 MFC check
9 Wafer handling	

Metrology

1 Facility/safety check
2 Repeatability and reproducibility
3 Accuracy
4 Correlation to pilot production
5 Calibration certificate
6 EMO/interlock check reference
7 Particle check
8 Wafer handling

Diffusion

Furnace	Wet
1 Facility/safety check	1 Facility/safety check
2 Gas check (MFC)	2 Host computer communications check
3 External torch check	3 Leak check
4 Leveling	4 Compressed air check
5 Heater baking	5 Wafer handling
6 Temperature check	6 Particle check
7 EMO/interlock check	7 Filtration panel check
8 Particle check	8 Running test
9 Thermocouple calibration/check	9 POU certification
10 Automechanism	

Thin-film

Sputter	CVD	Back grinder	Implant
1 Facility/safety check	1 Facility/safety check	1 Facility/safety check	1 Facility/safety check
2 Robot indexer check	2 MFC/temp/rf calibration	2 Spindle angle	2 Beam energy capability /stability
3 Exhaust flow	3 Leak test	3 Chuck uniformity	3 Mass resolution
4 Target burn-in	4 Robot settings	4 Robot settings	4 High-voltage calibration
5 Vacuum performance	5 Wafer handling	5 Wafer handling	5 Wafer handling
6 EMO/interlock check	6 Particle check	6 Particle check	6 Particle check
7 Particle check	7 EMO/interlock check	7 EMO/interlock check	7 Wafer cooling
8 Wafer handling	8 Baratron gain/linearity		8 Beam Parallelism
	9 Load lock		9 Dose uniformity
			10 EMO/interlock check

Product

Tester	Prober	Laser repair	Software Defect Analysis
1 Facility/safety check	1 Facility/safety check	1 Facility/safety check	1 MTBF
2 Calibration	2 Alignment	2 Configuration check	2 Hardware linking
3 Host controller communications check	3 Chuck X-Y accuracy	3 Temperature check	
4 System burn-in	4 EMO/interlock check	4 EMO/interlock check	
5 EMO/interlock check	5 Particle check	5 Particle check	
6 Particle check	6 Wafer handling	6 Wafer handling	
7 Wafer handling			

Micron Ex. 1029, p. 654
Micron v. YMTC
IPR2024-00911

TABLE 20
An equipment operating procedure (furnace tube change)

Item	Time (hr) Estimated	Actual
1. Process stop and heater off	6.0	6.0
Heater breaker off		
Cleaning of main valve	1	
Checking and cleaning of pumping parts for by-products	1	
Furnace cool down	6	
2. Tube/vacuum exhaust pull	4.0	2.0
Removal of shutter, nozzle, inner/outer tubes, thermocouple manifold	2	
Removal of vacuum piping	2	
3. Furnace/vacuum exhaust clean	3.0	2.5
4. Tube clean	0.0	0.0
5. Tube installation	4.0	2.7
6. Room temperature leak check	1.0	1.5
Performance of leak check for all seals	1	
Pumping and purging to degas		
7. Boat setup/registration	1.0	2.8
8. Heater on and heating up	2.0	1.0
9. High-temperature leak check	3.0	5.5
Process temperature leak check	0.5	
Pumping and purging to degas	2.5	
10. Temperature profile	8.0	7.0
11. Tube coating (process recipe with extended time)	3.0	3.0
12. Manufacturing test runs	7.0	7.0
Total time	42.0	41.0

In Table 21, a list of parameters for process checking is presented. The user decides which part of the list or whether additional specific parameters should be engaged in the user's own application. In Table 22, the process release steps just discussed are tabulated. A minimum of 25 tests reflects the statistical needs for meaningful control limit computation.[17] A solid process foundation can be established by combining the needs in Tables 21 and 22 and by thorough execution of the process release procedure shown in Fig. 11. As an example, process release data for the 4-Mbit DRAM trench capacitor critical dimension after etching are shown in Fig. 12 to illustrate the verification of the process capability and stability between transferred and received wafer fabs.

As a reminder, for manufacturing consideration, the process-related specifications and operating procedures (discussed in Section 11.2.4) also should be prepared in this stage to permit unambiguous execution by the manufacturing personnel.

11.3.5 Product Characterization

In the previous two paragraphs, we have discussed how to establish running and processing baselines for wafer fab equipment through characterization and verification

Micron Ex. 1029, p. 655
Micron v. YMTC
IPR2024-00911

TABLE 21
Control parameters of process release

Photo			Plasma		Diffusion	
Stepper	**Resist**	**Coater/developer**	**Etcher**	**Asher**	**AP furnace**	**Wet**
1 Lens performance Resolution Mix and match Depth of focus 2 Process CD Average Uniformity (within wafer to wafer)	1 Characterization	1 Particle test 2 Thickness check 3 Uniformity check Within wafer Wafer to wafer 4 Resist type	1 Particle 2 Etch uniformity 3 Etch selectivity 4 Etch profile 5 Etch rate 6 Resist burning 7 CD etch bias 8 Corrosion	1 Particle 2 Ash rate 3 Ash uniformity 4 Resist removal capability	1 Particle test 2 Oxide growth Thickness Uniformity 3 MCLT† 4 GOI‡ 5 Slipline 6 Bias temperature stress 7 Oxide growth during push 8 Doping uniformity for in-situ doped 9 Cross-sectional profile 10 Refractive index 11 Pinhole	1 Chemical assay 2 Particle test 3 Etch rate 4 MCLT† 5 Water mark 6 Impurities

Thin-film					Product	Metrology
CVD-Oxide	**Scrubber**	**Implanter**	**CVD-W**	**Sputter**	**Tester**	
1 Particle test 2 Thickness nonuniformity 3 Stress 4 Refractive index 5 Conformity 6 Pinhole 7 Peeling characteristic 8 Corrosion	1 Particle removing efficiency	1 Particle test 2 Sheet resistance uniformity 3 Metal contamination 4 Implant angle Check 5 GOI‡ 6 Implant species check 7 Dose and energy	1 Particle test 2 Resistivity uniformity 3 Reflectivity Center to edge 4 Step coverage 5 Mass gain 6 Grain size	1 Particle test 2 Reflectivity Center to edge 3 Stress 4 Step coverage 5 Resistivity Uniformity 6 Grain size	1 Calibration 2 Correlation to pilot reference wafer Parametric Parameter Laser Software defect analysis Repair success rate Multiprobe Fail label	1 Calibration 2 GRR§ study

† MCLT: Minority carrier lifetime
‡ GOI: Gate oxide integrity
§ GRR: Gauge repeatability and reproducibility

Micron Ex. 1029, p. 656
Micron v. YMTC
IPR2024-00911

TABLE 22
Process release steps

Purpose	Type of parameters	Material	Minimum # of tests
Capability	Parameters related to this process only	Bare pilot wafer	25
	Process integration parameters related to other processes	Preprocessed pilot wafers up to this process or	25
		Actual production lots	3–5
	One-time check parameters	Pilot or production wafers	3–5
Stability	Production control or monitoring parameters	Actual production wafers or production monitor wafers	25

of each process. With good preparation and execution, the approach has proven to be effective in our new advanced wafer fabs in fabricating good production units from the very beginning of the wafer fab start-up. The dependence of each process on the other (process integration) has been considered and checked through identified integration process parameters during the individual process releases. This approach minimizes the need to perform lengthy process integration checks using short-loop tests (combination of a few process steps but not the entire process flow) to verify the effectiveness of the process combination. It does not, however, eliminate the need for device/product characterization to ensure that not only process but also device margins are achieved. In addition, short-loop tests do serve other purposes—as a process monitor (e.g., gate oxide integrity) and as a device/product monitor (e.g., wafer-level reliability control)— and are commonly implemented on the manufacturing line for control. This paragraph will discuss several start-up device/product characterization considerations.

Parametric parameters

The parameteric parameters are those that are important to device performances.[13] With DRAM as an example, several parameters of interest to DRAM are the transistor threshold voltage (V_t), transistor GOI, cell dielectric integrity, cell leakage, and cell-to-cell isolation. In addition to these fundamental device parameters, there are also parametric parameters for process monitoring. Examples of parametric parameters are the electrical thickness of various dielectric layers (gate oxide and interleave isolation dielectrics) and the deviation of critical dimensions from design targets such as delta widths for gate and metal levels.

The approach/methodology for parametric parameter characterization is similar to that for the process release discussed in paragraph 11.3.4. It is equally important that the actual data from the pilot line or design criteria be clearly specified and collected. These data are compared with the data obtained from the start-up wafer fab. Often, fine-tuning of a few process parameters may be required to bring the parameter of concern back to the reference value if an actual shift is observed. The fine-tuning is aimed at adjusting device performance back to the designed margin. Even though poor performance margins may not impact the product salability,

Micron Ex. 1029, p. 657
Micron v. YMTC
IPR2024-00911

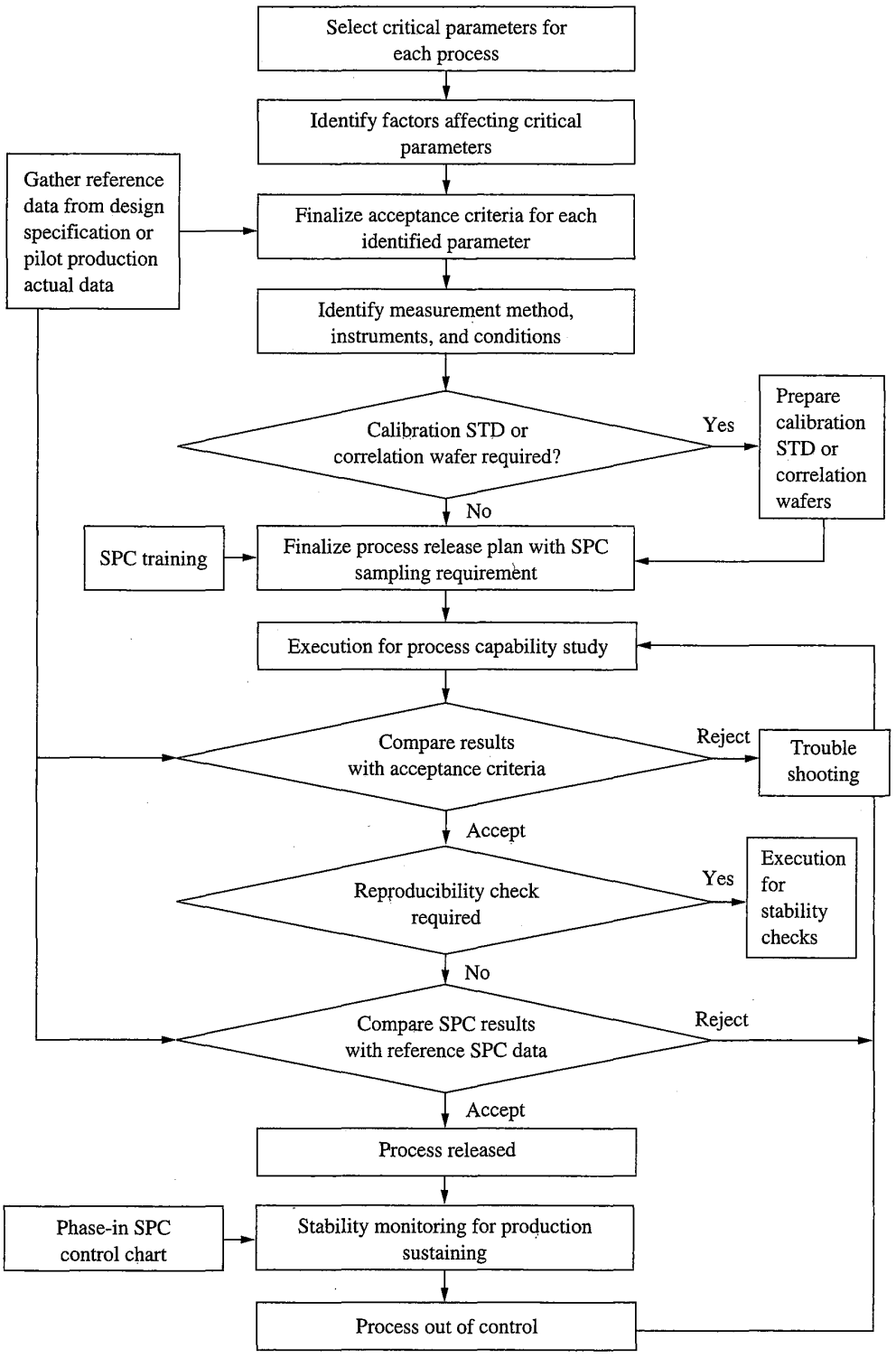


FIGURE 11
Schematic flow diagram for process release procedure.

Micron Ex. 1029, p. 658
Micron v. YMTC
IPR2024-00911

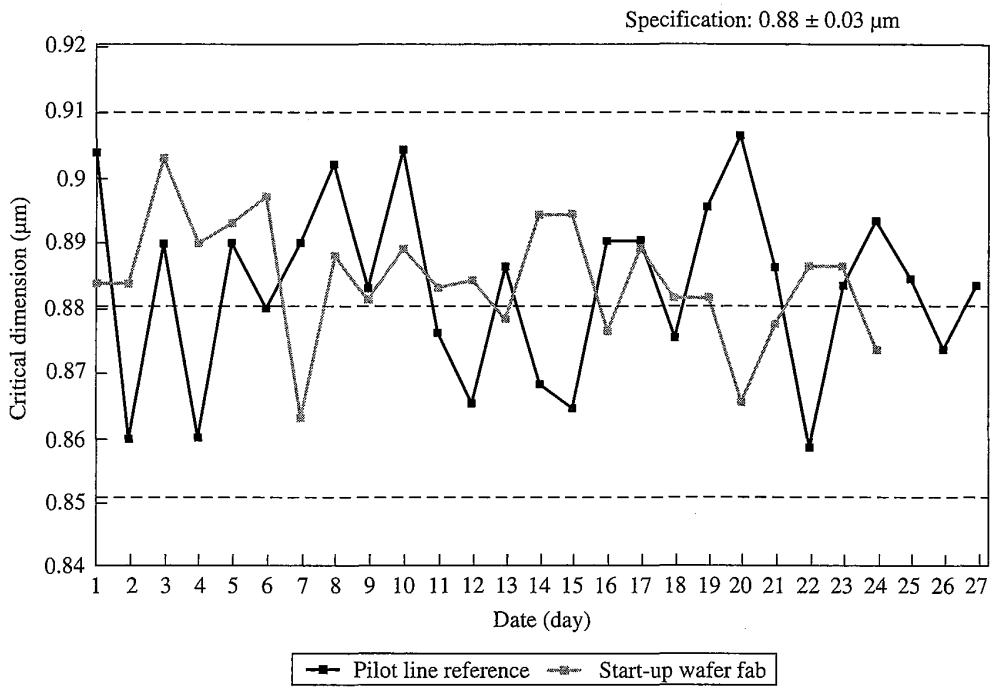


FIGURE 12
Verification of start-up 4-Mbit DRAM trench capacitor CD to reference performance.

maintaining a good margin is critical to guard against all variations in long-term volume manufacturability.[6] With DRAM as an example again, Fig. 13 illustrates a slightly worse device margin on 4-Mbit DRAM trench capacitance for the start-up wafer fab compared to that for the pilot reference. This variation indicates the need to fine-tune the capacitor dielectric thickness and critical dimensions.

Wafer-level reliability

In the preceding paragraph, we discussed the characterization of the functional parameters for the purpose of checking on the quality of the device at time zero. In this paragraph, we will consider aspects of the device reliability from the wafer-level point of view[29–31] for early detection of potential reliability problems. Wafer-level reliability control (WLRC) is implemented simply by physical or electrical means during the wafer fab processing. Some of the key factors for WLRC are these:

- Maximize in-situ reliability controls to replace product reliability control.
- Perform wafer-level reliability tests for fast process feedback, abnormality identification, and product classification.
- Develop device structure modeling to determine product life; investigate the feasibility of real-time test structures and wafer-level burn-in.

Typical WLRC tests and parameters[29–39] are summarized in Table 23. The selection of the WLRC parameters should be based on the historical problems, potential reliability concerns about the device of interest, and specific wafer fab process problems that affect product reliability.

Micron Ex. 1029, p. 659
Micron v. YMTC
IPR2024-00911

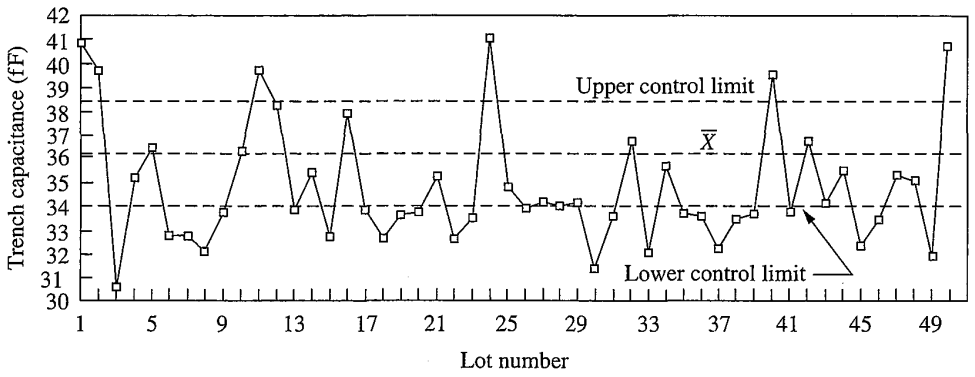


FIGURE 13
The slightly worse margin of the 4-Mbit DRAM trench capacitance for the start-up wafer fab requires process fine-tuning.

Probe testing

Wafer fab testing is not intended to replace final product testing after assembly but instead is aimed at studying the chip characteristics for process feedback and at classifying chips for various degrees of quality and reliability so as to simplify assembly/test operations for product sorting.

Typical wafer fab test programs include the following major categories:

- Parametric for testing the device parameters.
- Test of gross problems such as electrical continuity (open and short circuits).
- Test of various leakage mechanisms for device components (diode, transistor, capacitor, etc.), device isolation, and chip-level leakage in operating and standby modes.
- Test of chip functionality. The functionality test items, of course, are different for different devices. With DRAM as an example, functionality tests may include pause distribution, device timings (various circuit access times), test patterns (disturb, page, etc.), reliability in the form of voltage stress, and items specific to customer requirements.

Determining the test conditions for each of the test items is very important. The test must distinguish device quality levels but not eliminate good chips. Typically, the test conditions are decided on the basis of design, modeling, past experience, detailed correlation with packaged-product quality and reliability performances, and customer requirements. In many cases, a wafer fab test is conducted at higher than room temperatures (70 to 100°C) or at cold temperatures, because the typical guaranteed packaged-product temperature performance is in the range of −65 to 150°C or, a less stringent condition, −10 to 125°C. Unless there is a known low-temperature problem, wafer fab testing at high temperature is sufficient, because device performance is generally worse at high temperature.

One of the most important test parameters is, of course, the yield. Wafer fab test yield is not absolute but varies with different test items and conditions. For the same set of chips, a tighter test program may give a lower wafer fab yield but may provide

Micron Ex. 1029, p. 660
Micron v. YMTC
IPR2024-00911

TABLE 23
Typical wafer level reliability control tests

Wafer level reliability test	Physical or electrical	Application (Refs. 30–39)	Judgment criteria
Metal			
Notch/void	P	EM induced by PO mechanical stress	% to minimum metal line width
Grain size	P	Sputter machine leak Al metal migration	Range of grain size
Si nodule	P	Metal (AlSi) sintering temperature control	Range of nodule size
Side hillock bridging	P	Sputter machine control	% to minimum metal line spacing No bridging
Step coverage	P	Control of via and contact metal deposition	Minimum of step coverage
Adhesion	P	Surface condition and contamination Film stress control	No metal lifting Range of stress
Corrosion	P	Post-metal-etch cleanup	No visual metal corrosion
Electromigration (EM)	E	Metal line EM	% Metal resistance change after electrical/temperature stress
Metal via and contact resistance	E	Via and contact EM Si Buildup at via and contact	% Via and contact resistance change after electrical/temperature stress
Protective overcoat (PO)			
Stress	P	EM induced by PO stress Metal peeling	Range of stress
Thickness	P	PO process control	Range of thickness
PO integrity	P	Corrosion due to moisture penetration	Minimum defect density (pinholes, etc.)
Wafer PCT (temperature and humidity)	P	Corrosion resistance	No corrosion

Micron Ex. 1029, p. 661
Micron v. YMTC
IPR2024-00911

TABLE 23 *(continued)*

Typical wafer level reliability control tests

Wafer level reliability test	Physical or electrical	Application (Refs. 30–39)	Judgment criteria
Interlevel dielectrics			
Stress	P	Film cracking and delamination	Range of stresses
Thickness	P	Deposition process control Metal cracking relating to poor planarization	Range of thickness Reflow control
% phosphorus in PSG and BPSG	P	Phosphorus-induced corrosion	Maximum % of phosphorus
Wafer temperature cycle (TC) test	P	Interlevel strength	No peeling/delamination
Breakdown voltage cycle (VC) test	E	High breakdown strength	Minimum electrical field strength
Gate oxide integrity (GOI)			
Breakdown voltage	E	Gate process contamination, surface roughness, and dielectric quality control	Minimum electrical field strength
Device component			
Transistor channel hot carriers (CHC)	E	Short channel degradation	% Leakage current/transconductance increase after electrical/temperature stress
Diode leakage	E	Diode leakage kinetics	% Leakage current increase after electric/temperature stress

Micron Ex. 1029, p. 662
Micron v. YMTC
IPR2024-00911

a higher final test yield after assembly, burn-in, and reliability checks. There are no set rules on whether a tighter or looser test program should be used except for the fine balance between the test and correlation with product performance, as well as a thorough understanding of the customer requirements.

In Fig. 14, four histograms show the distributions of test items in each category using DRAM devices. These items and a few others are used as the wafer fab outgoing quality control items for a commercially available product. With this kind of distribution, product classification becomes possible based on statistical variations. The quality and reliability performances of the sorted chips, using the distributions shown in Fig. 14 can then be correlated with those of the final packaged products. As a final-stage check for start-up at a wafer fab, the sorting criteria for quality and reliability classifications of a specific test item and distribution should be referenced to design criteria and/or pilot production data. If all checks are carried out thoroughly, the performance of the fabricated wafers at the start-up wafer fab are already within design tolerances or in line with the actual performance of the pilot production. Hence, the outgoing QC criteria from the reference pilot line wafer will be accurate (80 to 90%) indicators of the actual final quality and reliability performance of the new start-up. This is important for the fast qualification of the new wafer fab facility to produce and sell products.

11.3.6 Product Qualification

Finished wafers are subjected to next-stage processes, namely, assembly, burn-in, testing, and reliability checks. Some details of the assembly and the reliability are discussed in other chapters of this book and therefore are not covered here. However, we will briefly introduce the considerations for packaged product qualification with an emphasis on the wafer fab tests. The subject deserves some attention, because the successful completion of product qualification signals an important start-up milestone. Qualification is not an easy task and depends a great deal on product design, process margin, and the lessons learned (particularly from reliability-related problems) from past product generations and processes. The minimum qualification cycle time itself can take up to 2 months since several reliability tests require 1000 hours or more of test time. If the product fails, another attempt at qualification would easily delay the start-up milestone by 3 to 6 months, counting the time from wafer start to the completion of qualification tests. This time does not include the problem-solving efforts. In the late 1970s and early 1980s, product qualification took up to two years (as with 16K or 64K DRAMs), which is very expensive because of the large number of units needed, the operation cost, and loss of revenue. Products with designed-in reliability based on experience gained from past lessons learned and manufacturing start-up technology with thorough process/product checks (Section 11.3) are key to a robust qualification process. As a result, the qualification time has been reduced to less than 6 months (e.g., 4-Mbit DRAM) in the 1990s.

Internal product qualification

There are two test categories for qualification: a test for quality (the device is functional at time zero) and a test for reliability (the device is functional at a

Micron Ex. 1029, p. 663
Micron v. YMTC
IPR2024-00911

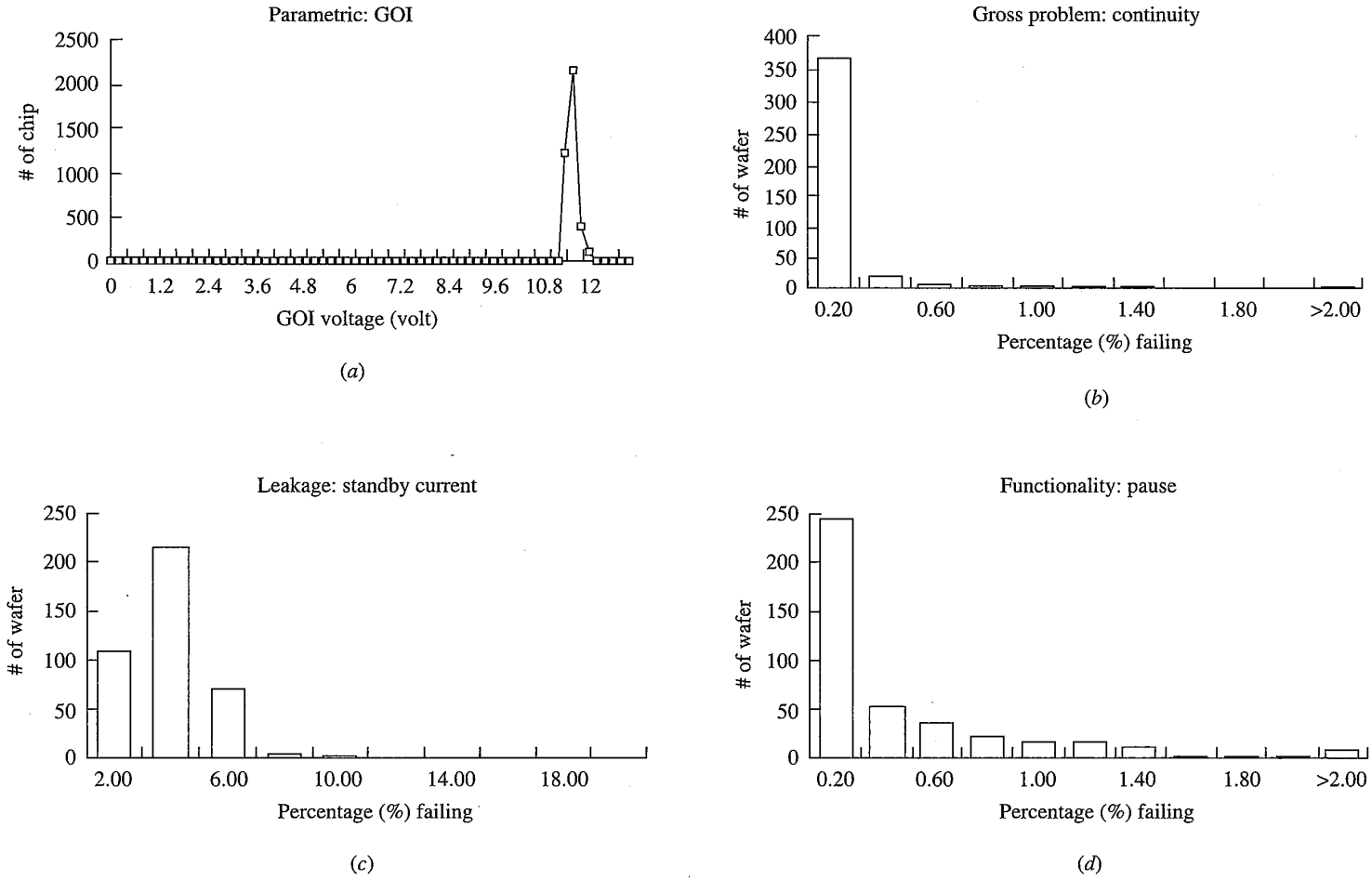


FIGURE 14
Examples for probe testing for (*a*) parametric, (*b*) gross problem, (*c*) leakage, and (*d*) functionality using DRAM devices and a sample size of 400 wafers from 20 lots.

Micron Ex. 1029, p. 664
Micron v. YMTC
IPR2024-00911

future guaranteed product life). The tests of quality are tests against the product data sheet specifications. The tests themselves are not different from those discussed in Section 11.3.5, hence we shall not consider them further. The tests of reliability are summarized in Table 24. The exact stress conditions for each test may vary based on customer requirements and guaranteed product life. By knowing the customer requirements and the guaranteed product life, the acceptance criteria for each test can be generated based on a typical reliability failure rate (e.g., failure rate at a 60% confidence level, or FR60) calculation[40] with the stress conditions and a large enough sample size to judge qualification success. Last, in Table 25, a product qualification approach is provided. As one can see, good practice involves not only a few lots of snapshot testing (one can only do limited reliability tests because of time constraints) but also continuous monitoring of more lots over a period of time to ensure volume manufacturability.

Customer qualification

The ultimate goal of manufacturing, of course, is to sell the products that it produces. To do that, products need be qualified by customers first. There are two

TABLE 24
Typical tests of product reliability

Test	Purpose	Typical conditions
Dynamic operation life (op Life)	Test of product operation functionality under stress	−10°C, 125 to 150°C 5 to 8 V 500 to 2000 hours
Early failure rate (EFR)	Check product "infant mortality"	125°C, 7 to 8 V 20 to 200 hours
Humidity	Moisture test for mechanical strength and corrosion	65°C, 85°C 65%, 85% relative humidity for 168 to 1000 hours
Autoclave or pressure cooker test (PCT)	Test of mechanical strength and corrosion under severe conditions of total saturation and temperature stress	120 to 125°C 2 atm pressure with absolute 100% relative humidity for 96 to 240 hours
Temperature cycle (TC)	Test of mechanical strength	−65 to 0°/125 to 150°C 500 to 2000 cycles
Electrostatic damage (ESD) sensitivity	Test of ESD damage to products	MIL STD 883 D Method 3015.7
Soft-error rate (SER)	Test of radiation-induced soft failures	Test parameters per product data sheet, 1 μs/access timing, external checkerboard and inverse checkerboard data pattern
Package integrity (not discussed in this chapter)	Test of package-related reliability concerns	
Others		

Micron Ex. 1029, p. 665
Micron v. YMTC
IPR2024-00911

TABLE 25
Product qualification approach

	Initial qualification		Volume ramp-up control	
Purpose	Capability demonstration		Manufacturability confirmation	
Tactics	Customer driven Process baseline establishment Assess intrinsic and early failure rate statistically Initial production start		Internally driven Process baseline sustaining Define product control range Final production release	
Source	Few lots with statistical validity		Every lot for a hundred lots or more	
Test	Highly accelerated stress margin test		Minimum lot coverage	
	Test	**Sample size**	**Test**	**Sample size**
	Op life	300–800	Op life	50–200
	EFR	500–2000	EFR	50–100
	Moisture	300–800	Moisture	50–100
	Thermal stress	300–800	Temp. cycle	50–100
	System soft error	200		
	ESD	20–50		
	Package integrity	200–300		
	Others	—		
Acceptance criteria	Customer requirements		Guaranteed product life	

possible levels of customer qualification: component and system. The exact customer qualification requirements are different for each customer. At the component level, however, qualification requirements are similar to those discussed in Section 11.3.5.6. Functional tests are based on the manufacturer's product data sheets, and reliability tests may include part or all of the tests listed in Table 24 as well as additional tests. The conditions for each test, as specified by each customer, may be different. However, they typically lie within those in Table 24. Hence, the component-level customer qualification is very similar to internal qualification from the test point of view. On the other hand, the customer system-level qualification is less known and predictable to manufacturers. For example, the same DRAM unit may produce quite different results in two different PC boards of the same type from two different system manufacturers. This outcome is possible because of the differences in the board design, materials used, processes, and system qualification test conditions. There are no general rules except to work with the customer and understand each specific customer requirement.

What would be an effective approach to deal with the whole range of test variations from many customers? One suggestion is simply to adopt the most stringent test conditions among all requirements for each test at the internal qualification stage. One can then work with the customers to have them accept the internal qualification results instead of the customer qualification at the component level. This is a win-win solution, because it reduces the qualification effort, cycle time, and cost for both the customer and manufacturer.

Micron Ex. 1029, p. 666
Micron v. YMTC
IPR2024-00911

11.3.7 Process and Product Baseline

With all these efforts, we are finally in a position to establish the process and product baseline for the start-up wafer fab. Baseline is a document that describes all important factors of the process required to produce the product. This document is aimed at maintaining the process procedures and parameters and at controlling of future process changes of the product. Thus, the process and product baseline is extremely important and should be tightly controlled to ensure high-quality, high-reliability manufacturing. Table 26 is a summary of the typical contents of the baseline.

The bottom line in establishing the baseline is protecting our customers who will buy and use the products. Therefore, the baseline specification is customer auditable,

TABLE 26
Baseline specification description

Subject	Description
Header	Product name, device name, design revision, process flow, ID/revision, sign-off information
Purpose	Purpose of the baseline specification
Scope	Definition of applicability of the baseline
References	List of references to baseline specification
Change definition	Define changes that require notification and qualification approval, such as these: Major process steps Chip active circuit design change (photomask) Process flow changes Major process-technology-related equipment changes Change of direct materials used for processing (materials in contact with wafer during processing) Unspecified/unqualified reworks Changes of outgoing QC specifications
Notification and qualification approval	Specify in detail the requirements and procedures of notification and qualification approval for baseline changes defined in the previous entry
Baseline specification	Details of the baseline specifications Definition of major process steps Listing of process steps based on process flow with reference to process specification IDs and key parameter control ranges Photomask IDs and specifications Critical equipment related to key process technologies, with equipment vendor and model specified Critical materials such as wafer, chemicals, and gases with reference to each material specification and a brief description Allowed rework guidelines Test and qualification requirements Specification revision record
Others	If any

Micron Ex. 1029, p. 667
Micron v. YMTC
IPR2024-00911

meaning that the customer can audit the wafer fab compliance based on this specification. Failures in baseline compliance found by a customer audit may result in the termination of the customer's order to purchase this product. Approval of the baseline document signals the commitment of the wafer fab to supply reliable, quality products in adequate quantities.

11.4 VOLUME RAMP-UP CONSIDERATIONS

Long-term success depends not only on a solid start-up foundation but also on the ability to produce products in volume. In Section 11.3, we discussed the foundation-building processes. In this section, we will address issues on volume production, which include equipment utilization, cycle time and inventory control, and sustaining production (baseline and change control).

The principal objective of a production wafer fab is to produce chips at a profit. For a new fab, the volume ramp-up is the key to profitability. Consider the following simple relationships. Conceptually, the profit can be expressed as

$$\begin{aligned} \text{Profit} &= \text{revenue} - \text{cost} \\ &= \sum(\text{output} \times \text{price} - \text{cost}) \\ &= \sum \text{output} \times \text{price} - \sum \text{cost} \end{aligned} \tag{11.1}$$

where Output = (wafer input) × (process yield) × (test yield)

The wafer input rate is limited by the bottleneck equipment, that is,

$$\text{Wafer input} = (\text{bottleneck equipment throughput}) \times (\text{equipment utilization}) \tag{11.2}$$

Thus, the output or profit can be raised if the wafer input rate, process yield, and test yield can be increased. Of course, at the same time, the cost must be controlled and reduced without sacrificing quality/reliability.

11.4.1 Equipment Availability and Utilization

The wafer input rate can be increased for a given set of process equipment if more wafers can be processed through the bottleneck equipment. Typically, the utilization of the equipment must be increased. In this section, we examine the limitations of equipment in terms of production usage.

Process equipment, like any equipment, will experience trouble and require repair, resulting in the loss of production time. Process equipment also contains components that wear out, and equipment must be shut down periodically for preventive maintenance,[5,27] which also results in lost time.

Micron Ex. 1029, p. 668
Micron v. YMTC
IPR2024-00911

A typical breakdown of the process equipment may be as follows:

Production	150 hours
Setup	20
Repair	5
Preventive maintenance (PM)	10
Idling	10
Total	195 hours, or 100%

Typical terminology for equipment used in wafer manufacturing can be expressed as follows:

$$\text{Availability} = \frac{\text{Time available for production}}{\text{Total time}} \tag{11.3}$$

$$= \frac{195 - (\text{repair} + \text{PM})}{\text{Total time}}$$

$$= \frac{180}{195} = 92.3\%$$

$$\text{Utilization} = \frac{\text{Time utilization for production}}{\text{Total time}} \tag{11.4}$$

$$= \frac{\text{Total time} - (\text{idling time available for production})}{\text{Total time}}$$

$$= \frac{150}{195} = 76.9\%$$

In this example, we see that even though availability for production is 92%, only 77% is actually utilized for production.

To improve utilization, one can break down the time loss further and then establish corresponding improvement actions (Section 11.5.1).

Time loss	**Improvement actions**
Repair	
Inexperienced equipment engineering staff	Increase training
Waiting for parts	Improve spare parts management
PM	
Inexperienced staff for maintenance	Increase training
	Reduce setup time
Idling	
Waiting for product	Increase inventory level as a buffer
No operator	Increase automation effort
	Cross-training of operators

Micron Ex. 1029, p. 669
Micron v. YMTC
IPR2024-00911

11.4.2 Cycle Time and Inventory Control

In Section 11.4.1, we pointed out that profitability can be achieved by volume (output) ramp-up. The key factors listed were wafer input rate, process yield, and test yield. All of these are basically technical or engineering factors.

We will introduce in this section other nontechnical factors that are also key to successful manufacturing.[41]

- *Cycle time:* For a given process step, the cycle time is defined as the elapsed time from input to output. For a typical IC process flow, there may be tens to hundreds of process steps. The total cycle time is the sum of the cycle time of each step.
- *Work in process* (WIP) *inventory:* The quantity of wafers that reside in a given process step, whether actually in process or waiting to be processed, is called WIP.
- *Process speed:* Process speed is the rate of input being processed to the output (good units produced) plus the scrap due to defects.

These three parameters are related by the following equation:

$$\text{Processing speed} = \text{WIP/cycle time} \tag{11.5}$$

Ideally, the WIP can be reduced to the point that wafers are waiting to be processed, provided the process equipment never breaks down; the input rate of the wafer exactly equals output; supplies to the process step never run out; etc.

In real life the equipment does break down, the input rate of the wafer fluctuates, and some supplies can be interrupted. The result is that the WIP level will fluctuate above the ideal. In fact, one typical action to counter input rate fluctuation is to increase inventory level deliberately to ensure that the bottleneck equipment will never run out of production material.[41] The general relationship between inventory level and equipment utilization is shown in Fig. 15. Note that high utilization requires high inventory. However, if the fluctuations or variability (V) of operations (process, equipment, supplies, etc.) is reduced ($V_2 < V_1$), the curve of cycle time vs. utilization shifts to the right as shown in Fig. 15. By reducing variability (better process/equipment control, tight inventory movement control, etc.) the output can be increased and cycle time can be reduced. The reduction of manufacturing variability is thus of fundamental importance.[4,6,17,41]

High inventory, hence high cycle time [Eq. (11.5)], can enhance equipment utilization and increase factory output to generate more revenue. However, high inventory and cycle time result in some problems.[2,41]

- *Cost:* All inventory requires money and increases the cost of manufacturing.
- *Risk of defective inventory:* Figure 16 illustrates the inventory distribution of two cases, one with cycle time CT_1 and the other with cycle time CT_2. If a defect generated at point X_g is not detected until point X_d, the inventory at risk is the inventory distribution curve from X_g to X_d. Thus, high inventory or long cycle time results in an increased risk of accumulating defective inventory.
- *Slow response to customer demand:* For a commodity IC such as a standard memory device, the ICs can be made and sold without much consideration to cycle time, except for inventory cost. For ICs that are made to order in response to the

Micron Ex. 1029, p. 670
Micron v. YMTC
IPR2024-00911

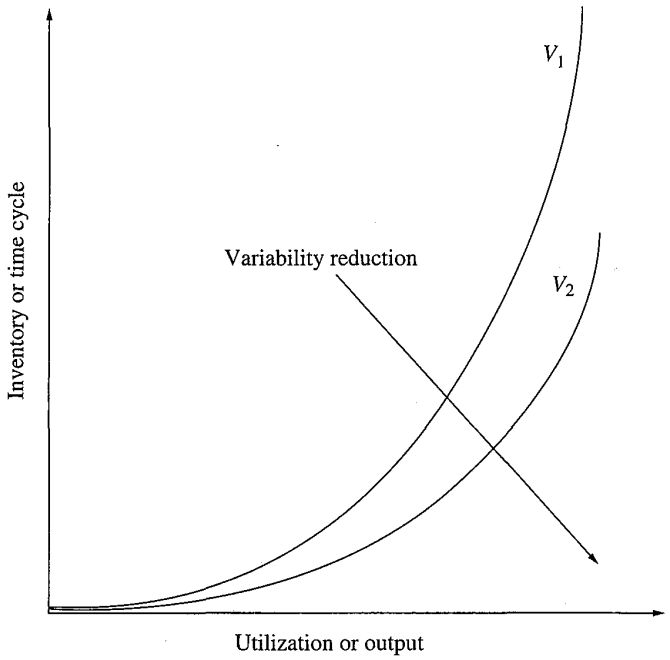


FIGURE 15
Relationships between cycle time and output.

quick pace of the market demand, which may be up or down, long cycle time results in long delivery time for the customer and potential excess inventory for the manufacturer.

Typical cycle time varies from 10 to 90 days. By going to all single-wafer processing, it has been reported that the cycle time of a complex CMOS IC can be less than 3 days![42]

11.4.3 Sustaining Production

We have repeatedly emphasized that minimizing variations is key to stable manufacturing.[4,6,17,41] Many of the wafer processes are interdependent and each may take a long time to complete, meaning that one may not know the exact impact of one change on others for a certain period of time. Typically, a large loss is incurred if one change adversely affects other processes. There are many painful real-life lessons where a large yield reduction (in excess of 10%) and loss of profit (in excess of $10M) were encountered when several changes or unknown variations were added, resulting in the loss of process margins. Perhaps the only way to deal with this situation is to control the variations proactively so that processes are kept stable to sustain continuous volume production. This situation is illustrated in Fig. 17 in terms of the cycle time needed for feedback (Section 11.4.1). It is obvious that only real-time control within an individual process provides the fast feedback needed for quick problem solving.

Micron Ex. 1029, p. 671
Micron v. YMTC
IPR2024-00911

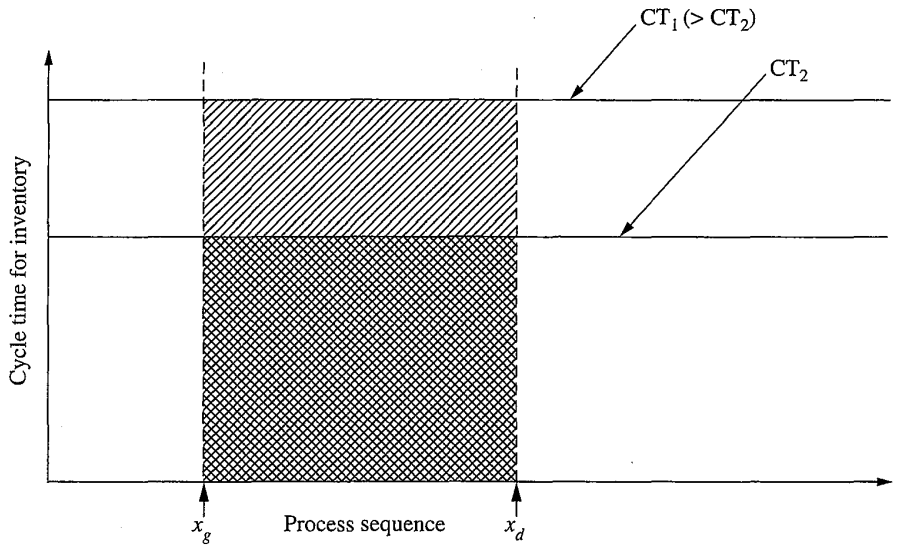


FIGURE 16
Risk for defective inventory (marked area) as affected by cycle time.

A scheme to achieve in-control manufacturing has been developed, as shown in Fig. 18. The concept is to implement designed-in capability and leading Q&R control parameters rather than to reject product based on lagging Q&R parameters. In addition to the use of the control strategy, as shown in Fig. 18, in-control manufacturing also depends on full compliance to procedures at all levels, which can only be ensured through repeated self-checkings.

Lagging parameters

Lagging parameters are those parameters that are the result of previous actions. They include assembly/test/reliability parameters (Table 24), wafer fab test probe parameters, and WLRC or short-loop process parameters (Table 23).

In all cases, the use of lagging parameters is less desirable for control because of the long feedback cycle time, as shown in Fig. 17, during which status information needed for diagnosis is lost. The lagging parameters are used to control the final net results, but they have more severe consequences (often, line shutdown) when an out-of-control situation occurs.

Concurrent quality (Q) and reliability (R) parameters

The concurrent parameters have real-time characteristics. They refer to property parameters such as film thickness (Table 21), process parameters themselves (Table 21), and equipment parameters (Table 19).

Concurrent Q&R parameters should be the main body of the process control and should be used heavily. One should lean toward using as many of the equipment parameters as possible for earliest problem detection. SPC is used to determine the out-of-control situation for each individual parameter.

Micron Ex. 1029, p. 672
Micron v. YMTC
IPR2024-00911

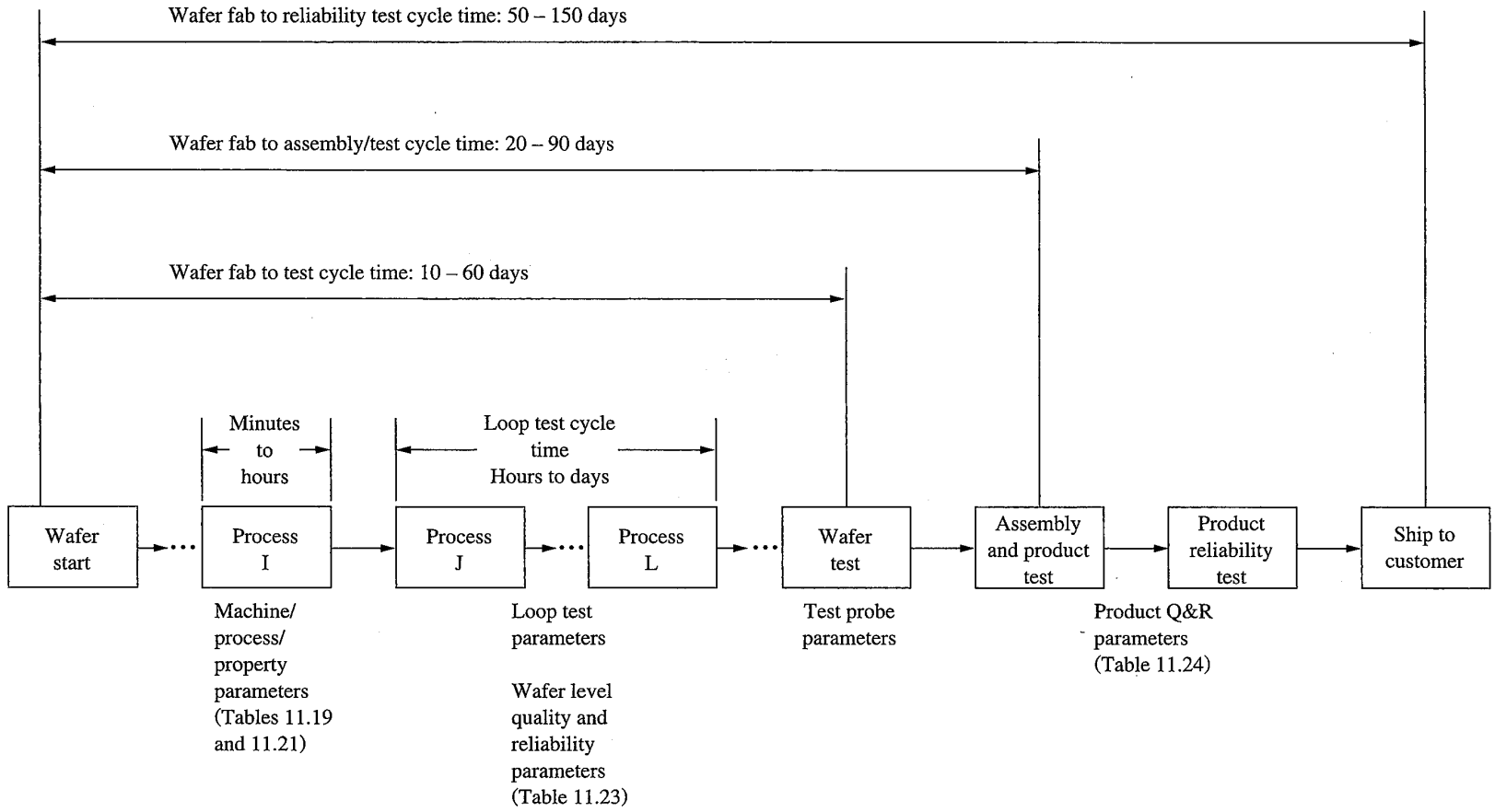


FIGURE 17
The long processing time indicates the importance of real-time control.

Micron Ex. 1029, p. 673
Micron v. YMTC
IPR2024-00911

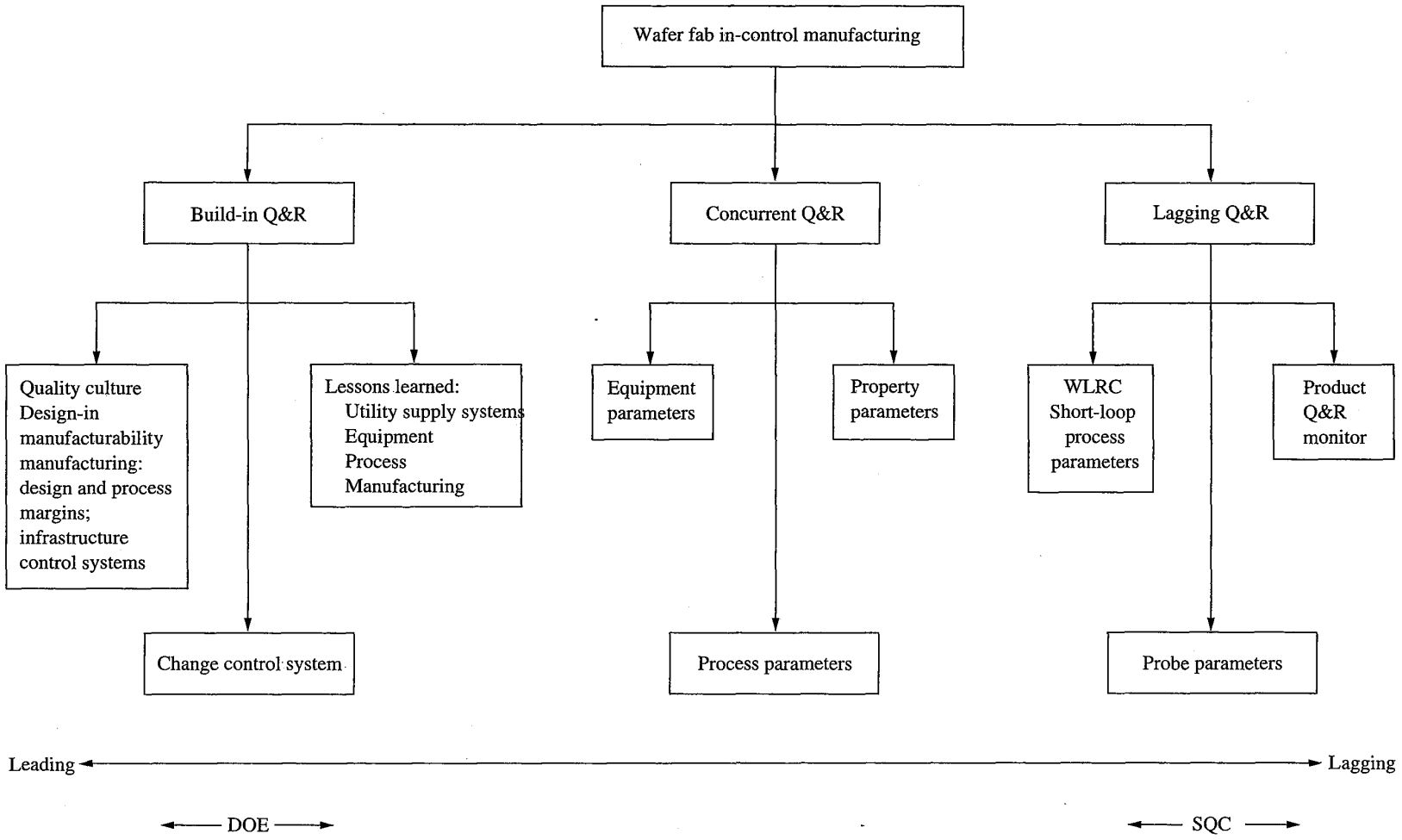


FIGURE 18
Scheme for an in-control wafer fab manufacturing strategy.

Micron Ex. 1029, p. 674
Micron v. YMTC
IPR2024-00911

Build-in quality and reliability

The most powerful strategy is to *build* quality and reliability *in.* Control activities can be significantly relaxed if the process/product margins are so tolerant that variations cause little or no impact on the product quality and reliability. The build-in strategy can be further classified based on its impact. The first build-in category refers to newly established technology, infrastructure, practice, or methodology based on historical problem-solution lessons. The material requirements, facility systems start-up, and equipment cleanliness control are examples of build-in Q&R based on past lessons.

The second build-in category refers to control of change. We have repeatedly warned about the potentially adverse impact of changes on manufacturability. However, since changes are inevitable, we must learn how to control changes or make the best out of them. Indeed, a controlled change, when executed properly, improves process capability. The change control system of a wafer fab therefore has critical importance for continuous improvements in manufacturability. The system should utilize statistical methods on design of experiments (DOE)[4] and estimate the impact of the change on reliability. In Fig. 19 a flow diagram illustrates the approach for change control that builds upon the start-up process release, using the process release plan and actual data base as its foundation.

Quality culture and designed-in manufacturability are the most important build-in strategies. Regardless of the hardware setup, wafer fab will not be successful without a total quality culture[1,2,5] encompassing people who want to do the right things. This desire is something that no control systems can replace. The establishment of the quality culture reflects management's commitment to excellence and must be carefully thought out and implemented. On the other hand, the purpose of designed-in manufacturability is to develop proactive design-in activities for functionality, reliability, and infrastructure so as to establish the margins for simple and easy manufacturing prior to the wafer fab start-up. Designed-in manufacturability is desirable for optimum yield and quick volume ramp-up to achieve maximum financial results.

In summary, start-up emphasizes using build-in Q&R activities to establish the baseline foundation. Subsequent production emphasizes using concurrent Q&R parameters for sustaining the baseline and using the change control system for further continuous improvements. The strategy has produced profound results,[43] which have led to short- and long-term success in our advanced wafer fabs.

11.5 CONTINUOUS IMPROVEMENT

In previous sections, we focused on how to start up and establish capacity control to achieve volume wafer manufacturing. The emphasis has been to reduce variations of all kinds to sustain the baseline process for stability. Only in Section 11.4.3 was the change control system introduced (Fig. 19).

In this section, we introduce the concept of continuous improvement[1–3] through the constant use of problem-solving methodology.[4] The result is improvement in

Micron Ex. 1029, p. 675
Micron v. YMTC
IPR2024-00911

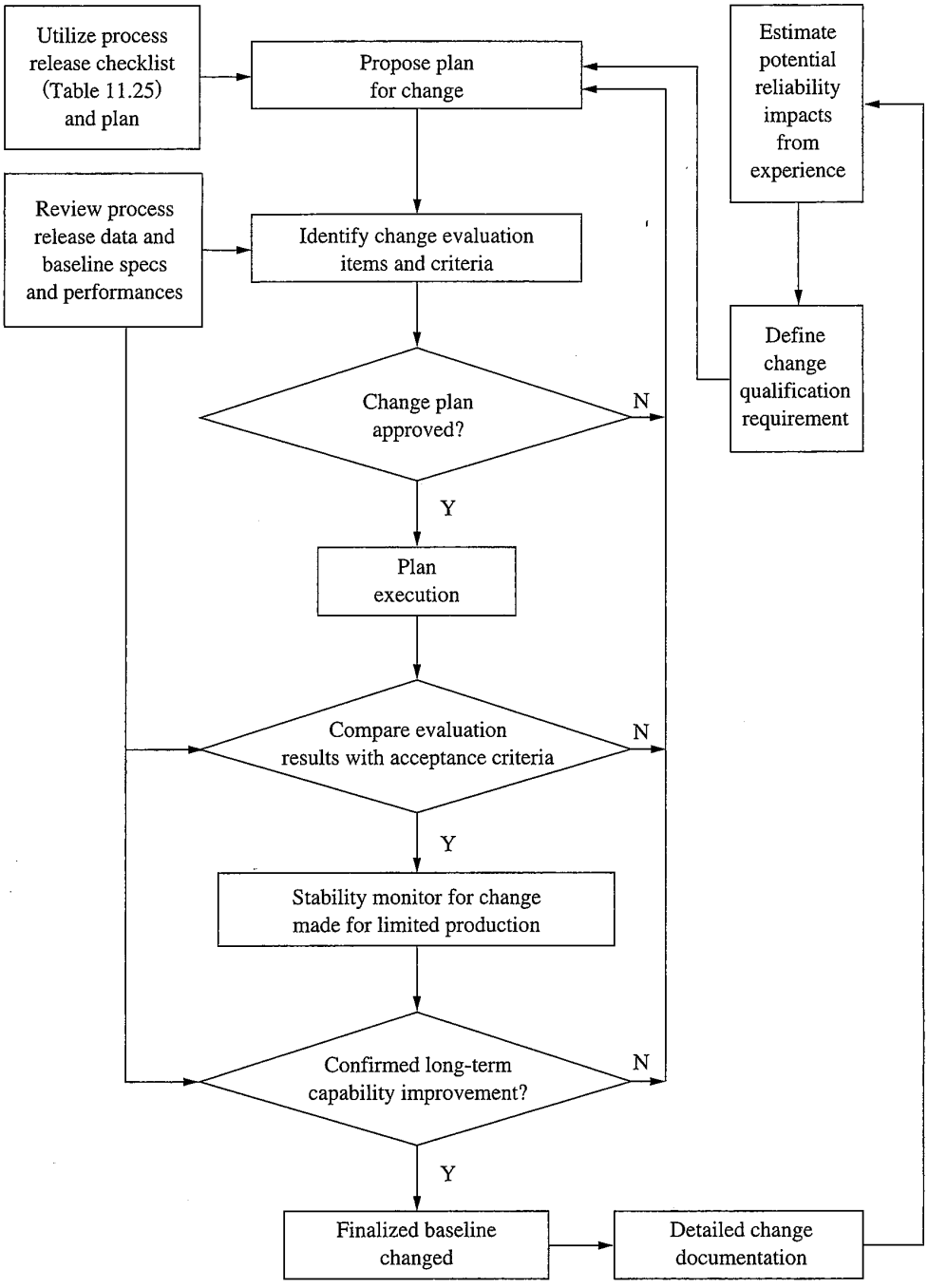


FIGURE 19
Flow diagram for change control.

Micron Ex. 1029, p. 676
Micron v. YMTC
IPR2024-00911

yield, capacity, Q&R, and the cost/profit margin. Methodology for continuous improvement is different from the yield models[11,44] that are discussed in the technology books. Readers should realize that improvement opportunities are everywhere. Many small improvements can accumulate to a big gain.[3] The nature of stable manufacturing is not luck but hard work and good control. There are many potential >10% yield drops but very little opportunity for one 10% yield gain. This kind of improvement is often overlooked by most technologists yet has great impact on real-life manufacturing considerations. Continuous improvement is the key to operational excellence.

11.5.1 Improvement Process

Figure 20 shows a continuous improvement process using problem-solving methodology.[3,4] The methodology uses commonly accepted QC tools. Repeated, honest, and thorough use of the methodology is the secret of success. To illustrate the improvement process, we will discuss a case study using the problem-solving tool to improve the LPCVD nitride furnace utilization and thus to improve capacity. The case is chosen to be consistent with the manufacturing theme of the chapter.

CASE STUDY

Step 1: Select problem.

- Poor nitride furnace utilization

Step 2: Understand present situation.

- Target vs. actual utilization on monthly basis

	Target	**Actual**
MTBF (hr)	>300	360
Equipment utilization (%)	> 80	60

- Pareto analysis for downtime breakdown (Total: 288 hours): Fig. 21
- Major causes of downtime: 1. 144 hr, or 50%, due to maintenance
  2. 94 hr, or 34%, waiting for parts
- Breakdown of maintenance activities

Activity	**# of times performed**	**Total downtime (hr)**
Tube change	2	82
Boat change	3	33
Waiting for dummy preparation	3	29

- Details of tube change procedure: Table 20
- Details of boat change requirements:

Tube clean cycle	5 μm of nitride deposited
Boat change cycle	2 μm of nitride deposited
Dummy wafer reclean	1 μm of nitride deposited
Boat scrap frequency	After 5 chemical etches

Micron Ex. 1029, p. 677
Micron v. YMTC
IPR2024-00911

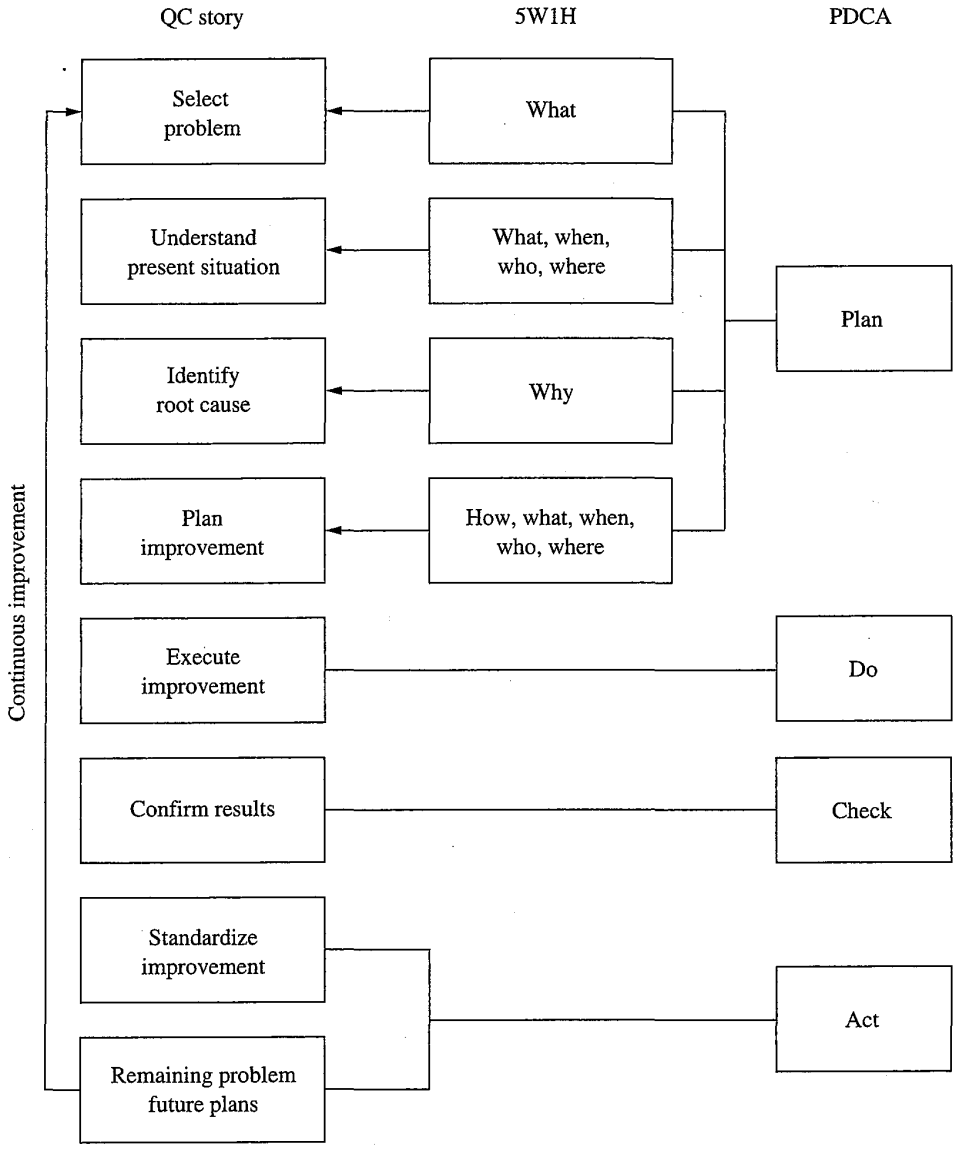


FIGURE 20
Continuous improvement through problem-solving methodologies.

Step 3: Identify root cause.

- Maintenance time loss

1. Poor PM procedure
2. Poor people skills in performing maintenance work such as proper sealing for leak-tight installation
3. Long boat setup/registration time due to alignment shift caused by change of quartz dimension with uncontrolled boat etching
4. Short tube/boat clean cycle

Micron Ex. 1029, p. 678
Micron v. YMTC
IPR2024-00911

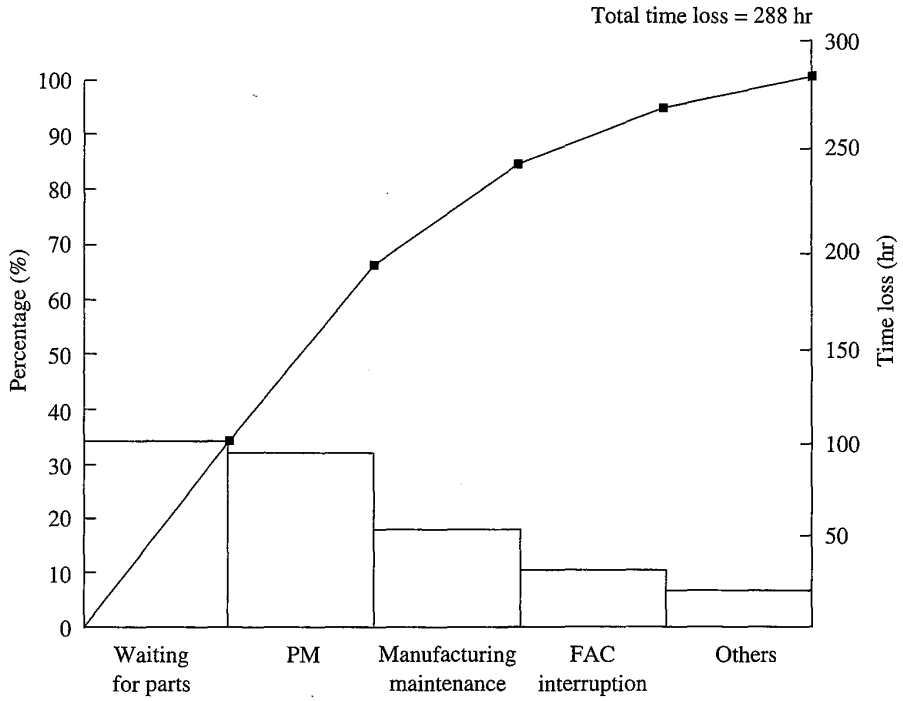


FIGURE 21
Pareto analysis on low furnace utilization.

5. Long waiting time for oxide dummy wafers used for nitride, due to poor peroxide furnace scheduling

- No spare parts, such as manifold, available
- Furnace interruption related to gas shutoff, due to improper alarm setting

Step 4: Plan improvement.

- Maintenance improvement

1. Revise tube change PM procedure based on actual experiences by equipment engineering.
2. Retrain furnace maintenance personnel.
3. Establish priority and communication systems for oxide furnace scheduling by manufacturing and planning.

- Setup of spare parts consignment and inventory management systems by purchasing/receiving
- False alarm elimination by facility engineering

Step 5: Execute improvement.

- Carry out plans outlined in step 4.

Step 6: Confirm results (3 months later).

- Target vs. actual utilization

Micron Ex. 1029, p. 679
Micron v. YMTC
IPR2024-00911

	Target	Past actual	Current actual
MTBF (hr)	>300	360	480
Equipment utilization (%)	> 80	60	84

- Total downtime reduction: 172 hr

Improvement plan	Total downtime (hr) Past	Current
Maintenance	144	99
Waiting for parts	98	0
Facility interruption	29	0
Other	17	17
Total:	288	116

- Maintenance improvement: 45-hr reduction in time lost

1. Revised PM procedure shown in Table 27
2. Confirmed maintenance activities

Activity	# of activity	Total downtime (hr) Past	Current
Tube change	2	82	60
Boat change	3	33	33
Waiting for dummy preparation	3	29	6
	Total:	144	99

Step 7: Standardize improvement.
The PM procedure shown in Table 27 has been implemented for all nitride furnaces. Similar procedures are also applied to other LPCVD furnaces.

Step 8: Identify remaining problems/future plans.

- Large time loss due to boat change: 4.6% loss of utilization
- Other losses in Fig. 21

There are numerous wafer fab operations in which the improvement methodology as shown in the case study can be applied for areas such as process capability improvement; defect reduction; failure analysis; test yield improvement; and manufacturing improvements for wafer scrap, misoperation, and equipment capacity.

11.5.2 Supplier Improvement

A supplier improvement program[1–3] is also important to sustain and improve wafer fab manufacturing. Daily activities depend so much on suppliers for materials and services that they contribute significantly to improving or degrading wafer fab performance. Because a good supplier is hard to come by, owing to the high standards, particularly for the advanced wafer fabs, it is only through long-term cooperation with well-managed, capable vendors that wafer fab supplies can be guaranteed.

Micron Ex. 1029, p. 680
Micron v. YMTC
IPR2024-00911

TABLE 27
Revised furnace tube change operating procedure

Item	Time (hr) Estimated	Actual
1. Pre-PM preparation Prepare spare quartz, manifold, O-ring, Dummy wafers, etc.	0	0
2. Process stop and heater off Heater breaker off Clean main valve Check and clean pump parts for by-products Remove and clean vacuum exhausts 3 hrs after heater off Furnace cool down	5.0 —	5.0
3. Tube pull Remove shutter, nozzle, inner/outer tubes, thermocouples, and manifold, and clean up	2.0 —	2.0
4. Furnace clean	1.0	1.5
5. Tube/boat clean	0.0	0.0
6. Tube installation Check tube for less than 5 times etching use previously prepared spare parts	2.5	2.5
7. Room temperature leak check Perform leak checks for all seals Pump and purge to degas	1.5	1.5
8. Boat setup/registration Check boat for less than 5 times etching Update records on part ID and etching time	1.0	1.0
9. Heater on and heating up	1.0	1.0
10. High-temperature leak check Process temperature leak check Pump and purge to degas	2.0	2.5
11. Temperature profile and tube coating	7.0	7.0
12. Manufacturing test runs	6.0	6.0
Total time:	29.0	30.0

We shall not discuss details of the supplier improvement program. However, all of the principles discussed in this chapter are applicable to suppliers, and standards are just as stringent.

11.6 SUMMARY AND FUTURE TRENDS

In this chapter, we have discussed various aspects of wafer fab manufacturing. We have introduced a strategy for in-control manufacturing. We have discussed general manufacturing considerations for organization, layout, materials, systems and

Micron Ex. 1029, p. 681
Micron v. YMTC
IPR2024-00911

procedures, computer-integrated manufacturing, and cost structure. We have also described an approach/methodology for start-up and volume production. Lastly, we have illustrated a continuous improvement process with a case example. All of these are aimed at establishing wafer fab manufacturing to meet customer needs and generate satisfaction.

In practical situations, wafer fab must meet customer requirements and profitability. Automation in hardware to reduce variations in the production process and automation in information to manage the manufacturing dynamics in real time are an important approach to achieve this objective. Design-in manufacturability with sufficient margin so that variations can be easily brought into control is a definite trend. Overall, proactive design and implementation of activities upstream (toward input parameters) are essential to the success of wafer fab manufacturing, in contrast to the current practices of in-line control and screening of noncompliance in finished products downstream.

REFERENCES

1. W. E. Deming, *Out of the Crisis,* MIT-CAES, Cambridge, MA (1982).
2. R. T. Schonberger, *Japanese Manufacturing Techniques,* Collier Macmillan Publishers, London (1982).
3. M. Imai, *Kaizen,* Random House Business Division, New York (1986).
4. G. Taguchi, *Introduction to Quality Engineering,* NIPUB/Quality Resources, New York (1986).
5. H. Naguib, "The Implementation of Total Quality Management (TQM) in a Semiconductor Manufacturing Operation," IEEE Trans. Semic. *Manuf.* **6** (2), 156 (1993).
6. M. J. Harry and J. R. Lawson, *Six Sigma Producibility Analysis and Process Characterization,* Motorola University Press (1992).
7. The American Society for Testing and Materials (ASTM) Publication 25, Section 10: Electrical Insulation and Electronics, Philadelphia, PA (1993).
8. Semiconductor Equipment and Materials Institute (SEMI), Materials Volume, Mountain View, CA, 1993.
9. S. Wolf and R. N. Taubeer, *Silicon Processing,* Vol. 1, Lattice Press, Sunset Beach, CA (1987).
10. S. K. Ghandhi, *VLSI Fabrication Principles,* John Wiley & Sons, New York (1983).
11. S. M. Sze, Ed., *VLSI Technology,* 2nd ed., McGraw-Hill, New York (1988).
12. R. Iscott, "The Search for Ultrapure Chemicals," *Semic. Intl.,* 70 (July 1990).
13. S. M. Sze, *Physics of Semiconductor Devices,* 2nd ed., John Wiley & Sons, New York (1981).
14. D. Toy, "Gas Purity and Wafer Fabrication," *Semic. Intl.,* 114 (Apr. 1988).
15. R. Iscott, "Water Purity for the DRAM Generation," *Semic. Intl.,* 60 (Jan. 1991).
16. J. L. Lamprecht, "ISO9000 Implementation Strategies," *Quality,* 14 (Nov. 1991).
17. D. J. Wheeler and D. S. Chambers, *Understanding Statistical Process Control,* Statistical Process Controls, Inc., Knoxville, TN, 1986.
18. B. C. Belanger, "Traceability: An Evolving Concept," *ASTM Standardization News,* **8** (1), 22 (1980).
19. L. Peters, "Measuring Your Thinnest Films," Semic. Intl., 76 (June 1991).

Micron Ex. 1029, p. 682
Micron v. YMTC
IPR2024-00911

20. M. T. Postek and D. C. Joy, "Submicrometer Microelectronics Dimensional Metrology: SEM," *J. Res. Natl. Bur. Stand.* (US), **92** (3), 205 (1987).
21. S. A. Rievi, "New Frontiers in the Metrology of Registration," *TI Tech. J.,* 76 (Jan.-Feb. 1990).
22. *Quality Function Development, Implementation Manual,* American Supplier Institute, Inc. (1989).
23. K. Kuhara, U. Katoh, and M. Shimbo, "Effects of Ultra-Clean Technology on Devices," *Ultra Clean Technol.,* **1** (2), 61 (1990).
24. Y. Motomura, "Evaluation of Piping Materials for Ultra Pure Water System—PVDF, PEEK," *Ultra Pure Technol.,* **1** (2), 86 (1990).
25. P. Burggraaf, "Valves Are Vital in UPDI Water Systems," *Semic. Intl.,* 112 (Sept. 1989).
26. D. C. Grant, S. VanDyke, and D. Wilkes, "Issues Involved in Qualifying Chemical Delivery Systems for Metallic Extractables," *Proceedings of the 39th Annual Technical Meeting,* Institute of Environmental Sciences, Mt. Prospect, IL, 1993, p. 200.
27. Y. Funaki, "Total Preventive Maintenance," *Semic. Intl.,* 162 (Sept. 1989).
28. S. Wolf, *Silicon Processing,* Vol. 2: Processing Integration, Lattice Press, Sunset Beach, CA (1990).
29. M. Davis and F. Haas, "In-Line Wafer Level Reliability Monitors," *Sol. State Technol.* 107 (May 1989).
30. C. L. Hong and D. L. Crook, "Breakdown Energy of Metal—A New Technique for Monitoring Reliability at the Wafer Level," *Proceedings of the 23rd IEEE International Reliability Physics Symposium,* 1985, p. 108.
31. S. K. Fan and J. W. McPherson, "A Wafer Level Corrosion Susceptibility Test for Multilayered Metallization," *Proceedings of the 26th IEEE International Reliability Physics Symposium,* 1988, p. 50.
32. J. W. McPherson, "Stress Dependent Activation Energy," *Proceedings of the 24th IEEE International Reliability Physics Symposium,* 1986, p. 12.
33. J. R. Black, "Electromigration—A Brief Survey and Some Recent Results," *IEEE Trans. Electron Dev.* **ED-16** (4), 338 (1969).
34. J. C. Ondrusek, C. F. Dunn, and J. W. McPherson, "Kinetics of Contact Wearout for Silicided ($TiSi_2$) and Non-Silicided Contacts," *Proceedings of the 25th IEEE International Reliability Physics Symposium,* 1987, p. 154.
35. S. C. Kolesar, "Principles of Corrosion," *Proceedings of the 12th IEEE International Reliability Physics Symposium,* 1974, p. 155.
36. J. W. McPherson and C. F. Dunn, "A Model for Stress-Induced Metal Notching and Voiding in VLSI Al-Si Metallization," *J. Vac. Sci. Technol. B,* **5,** 1321 (1987).
37. R. Moazzami and C. Hu, "Projecting Gate Oxide Reliability and Optimizing Reliability Screens," *IEEE Trans. Electron Dev.,* **37** (7), 1643 (1990).
38. J. W. McPherson and D. A. Baglee, "Acceleration Factors for Oxide Breakdown," *J. Electrochem. Soc.,* **132,** 1903 (1985).
39. E. H. Snow, A. S. Grove, D. E. Deal, and C. T. Sah, "Ion Transport Phenomena in Insulating Films," *J. Appl. Phys.,* **36,** 1664 (1965).
40. P. A. Tobias and D. Trindade, *Applied Reliability,* Van Nostrand Reinhold, New York, 1986.
41. E. M. Goldratt and J. Cox, *The Goal,* North River Press, Croton-on-Hudson, NY, 1986.
42. *Texas Instruments Corporate News,* June 28, 1993.
43. R. Iscoff, "SI Honors Top Fabs of 1994," *Semic. Intl.,* 76 (Apr. 1994).
44. SEMATECH, "Yield Models," Publication 91060594A-TR, Austin, TX (July 1991).

Micron Ex. 1029, p. 683
Micron v. YMTC
IPR2024-00911

PROBLEMS

1. Describe key factors for customer satisfaction. How would those factors impact the success of wafer fab manufacturing?

2. Construct a complete manufacturing process flow following the example shown in Table 1 using a device of your choice.

3. List five to seven important material characteristics for wafer, chemicals, and gases and describe typical application ranges for each characteristic.

4. Describe parameters of concern for each of the utility systems (cleanroom, chemical, gas, and DIW) supplying the wafer fab. Specify those of particular interest to the point of delivery/use from each of the chemical, gas, and DIW systems.

5. Describe five or more of the process-checking parameters for releasing of the furnace, wet clean, stepper, plasma etching, CVD, ion implantation, and sputtering processes.

6. Describe the four major wafer fab test categories. Give an example of the test parameters for each test category using a device with which you are familiar. Also describe the purpose and test conditions for each of the five (op Life, EFR, humidity, PCT, and TC) product reliability tests.

7. Assume a wafer fab can produce 100 wafers a day. What are the wafer inventory levels if the cycle time to produce products A and B are 75 days and 10 days, respectively? By how much can the wafer inventory be reduced by reducing the product A cycle time to 60 days?

8. Prepare an example of your choice following the improvement process (Fig. 20) and case study (Section 11.5.1) discussed in this chapter.

Micron Ex. 1029, p. 684
Micron v. YMTC
IPR2024-00911

CHAPTER 12

Reliability

J. T. Yue

12.1 INTRODUCTION

The reliability of integrated circuits involves many disciplines such as design, process, assembly, and test. Many delicate processing steps and manufacturing procedures are involved in producing a finished IC component. When the part fails in the field, a myriad of reasons could be the cause. Sophisticated and tedious analysis is required to get to the root cause of failure. From the customer perspective, what is really required of the product is continual functionality of the IC during the specified lifetime. Typically, manufacturers design an IC to last 10 years. Product reliability is ensured if each of the elements of the product is reliable for that length of time. There are three major elements of product reliability: design reliability, process reliability, and assembly reliability. If the reliability of each of these elements meets the required lifetime, then product reliability is ensured.

ULSI technology involves scaling of processes to the deep submicron (below approx. 0.35 μm) dimensions, as well as the additions of new process modules not used previously in the VLSI era. The reliability of each new process module, and how it interacts with other modules, will be critical to the final reliability of the entire process. Often, the scaling of a technology is performed without reducing the power supply voltage. Such an approach presents great challenge to the device engineers, reliability engineers, and process integration engineers. As a result, trade-offs are occasionally required among reliability, design, and process development. In the ULSI era, the concept of *design for reliability* is extremely important. Reliability must be designed in during all stages of IC development, including design, process development, and manufacturing. Therefore, process reliability is an extremely important subject.

First, basic process reliability wearout concepts will be reviewed, such as hot carrier injection, electromigration, stress migration, and oxide reliability. Then the

Micron Ex. 1029, p. 685
Micron v. YMTC
IPR2024-00911

effects of scaling on the reliability of these failure mechanisms will be discussed. Next, the relationship of dc to ac lifetime will be discussed. This topic is important since real-world reliability is based on time-varying, alternating stress conditions, not on constant dc stress conditions. Special reliability concerns arising from some of the recent ULSI processing approaches will be discussed, e.g., tungsten plugs and charge damage during manufacturing. The mathematics of reliability calculations and failure distributions will be included for completeness.

Integrated circuits are made of discrete components such as transistors, resistors, interconnections, dielectric thin films, and capacitors. These components are connected in a predetermined scheme (set by the circuit designers) and put together through a sophisticated sequence of process recipes (set by process integration). One way to measure the reliability of the entire IC process is to measure the reliability of each of the major process modules. Some of the key process modules from the reliability point of view are (1) transistor, (2) interconnection, and (3) gate dielectric.

After a transistor has met all of its designed performance, it has been shown that the transistor may not be able to maintain its performance throughout its intended lifetime. The degradation becomes more severe for shorter-channel devices. This degradation is called the *hot carrier injection* (HCI) *problem* and will be covered in Section 12.2. Similarly, a metal interconnection, either a series of vias or a single metal line, does not keep the same *resistance* value while current is continually passed through it. The resistance of the line will degrade by void formation in Al by a phenomenon called *electromigration* (EM). Even in the absence of current, it has been observed that an Al interconnection can suddenly open up as a result of mechanical stress relaxation, causing open circuits. This phenomenon is called *stress migration* (SM). Both EM and SM will be discussed in Sections 12.4 and 12.5. Gate oxide is an important element of the MOS transistor or of a storage capacitor in a DRAM circuit. This thin dielectric can break down, resulting in gate shorts, during a long application of electric field across it. This phenomenon is called *time-dependent dielectric breakdown* (TDDB) and is increasingly important to consider as gate oxide reaches the 100 Å level and below. The TDDB mechanism will be discussed in Section 12.5. The effect of device scaling on reliability is discussed in Section 12.6. The relationship of lifetimes between dc stress and ac stress is discussed in Section 12.7. Recent ULSI reliability concerns in tungsten plug and plasma damage are discussed in Section 12.8. Finally, the mathematics of failure distribution are briefly discussed in Section 12.9.

12.2 HOT CARRIER INJECTION

Hot carrier injection (HCI) is the phenomenon in which carriers (either electrons or holes), excited by the high electric field in the transistor channel region, are injected into the gate oxide of a MOS transistor, causing damage to the oxide/Si interface and/or trapped charges in the gate oxide. The carriers in the channel become overly excited (hence, *hot*) because they gain enough energy to cause impact ionization before they lose their energy by phonon scattering.[1] The oxide damage and trapped charges will cause electrical parameters of the transistor to drift.

Micron Ex. 1029, p. 686
Micron v. YMTC
IPR2024-00911

Figure 1 illustrates the relationship of the carrier velocity of electrons and holes as a function of the channel longitudinal electric field. When the field exceeds 20 kV/cm, the carriers begin to lose energy by scattering with optical phonons, and their velocity saturates. Above 100 kV/cm, the carriers gain more energy from the field than they can lose by phonon or impurity scattering. Consequently, the energy of these carriers builds up. They are no longer in thermal equilibrium with the lattice and therefore are called *hot carriers.* When the excess energy exceeds a certain threshold, impact ionization is initiated. The electron-hole pairs generated in turn produce more pairs, and so on, resulting in the so-called avalanche multiplication process.

The channel electric field in the velocity-saturation region can be expressed[2] as

$$\mathscr{E}(x) = [V(x) - V_{\text{dsat}}]/l, \tag{12.1}$$

where V_{dsat} is the potential at the pinch-off point in the channel, x is taken to be zero at this point, and l is the effective length of the velocity-saturation region. The maximum electric field, $\mathscr{E}_{\text{max}}$, is found by replacing $V(x)$ by the drain voltage, V_{DS}, in Eq. (12.1). As the transistor channel decreases, V_{dsat} decreases and thus the $\mathscr{E}_m$ increases. HCI degradation is known to become worse in shorter channel devices.[3]

Figure 2 illustrates the impact ionization process in an NMOS transistor. For HCI degradation to occur, three events must happen: (1) there must be impact ionization, (2) the created electron (or hole) must be injected into the gate oxide (near the drain end), and (3) the injected carrier must be trapped near the oxide/Si interface or in the bulk oxide, or else break a chemical bond (e.g., Si-H).[2] Each of these events has a probability of occurrence.

$$\text{Event 1} \rightarrow \sim I_d \exp(-\phi_i/q\lambda\mathscr{E}_m) \tag{12.2}$$

$$\text{Event 2} \rightarrow \sim I_d \exp(-\phi_b/q\lambda\mathscr{E}_m) \tag{12.3}$$

$$\text{Event 3} \rightarrow \sim I_g \exp(-\phi_B/q\lambda\mathscr{E}_m) \tag{12.4}$$

The impact ionization energy in the silicon is ϕ_i, ϕ_b is the Si/SiO_2 barrier energy, and ϕ_B is the chemical-bond-breaking energy in the silicon dioxide (e.g., for Si–H). I_d is the drain current, I_g is the gate current, and $\mathscr{E}_m$ denotes the maximum electric field in Eq. (12.1).

The reliability consequence of these trapped charges (or damaged oxide/Si interface) is a degraded transistor. Parameters such as threshold voltage V_T, transconductance g_m, saturated drain current I_{dsat}, and linear drain current I_{dlin} are common parameters used to quantify the degree of HCI degradation. The HCI process also generates substrate currents I_{sub} as indicated in Fig. 2, which, if severe enough, can induce bipolar latch-up in circuits.[4]

12.2.1 NMOS and PMOS Transistor

Interface trap generation is believed to be the main degradation mechanism for nMOSFETs.[5–8] The mechanism of interface trap generation is not well understood and is still an active subject of research. It may involve hot electrons, hot holes,

Micron Ex. 1029, p. 687
Micron v. YMTC
IPR2024-00911

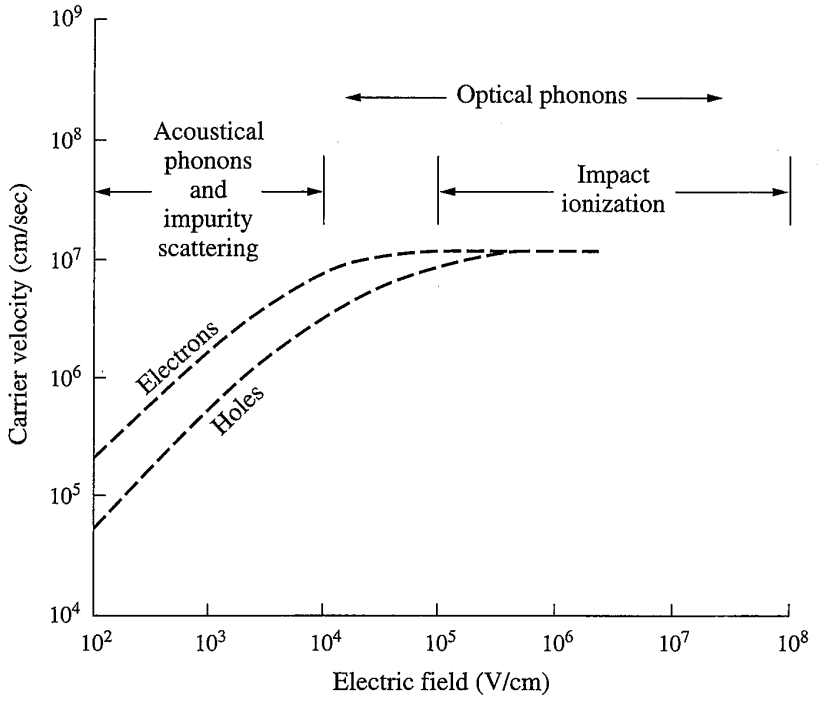


FIGURE 1
Effect of electric field on carrier velocity. (*After Sabnis, Ref. 1.*)

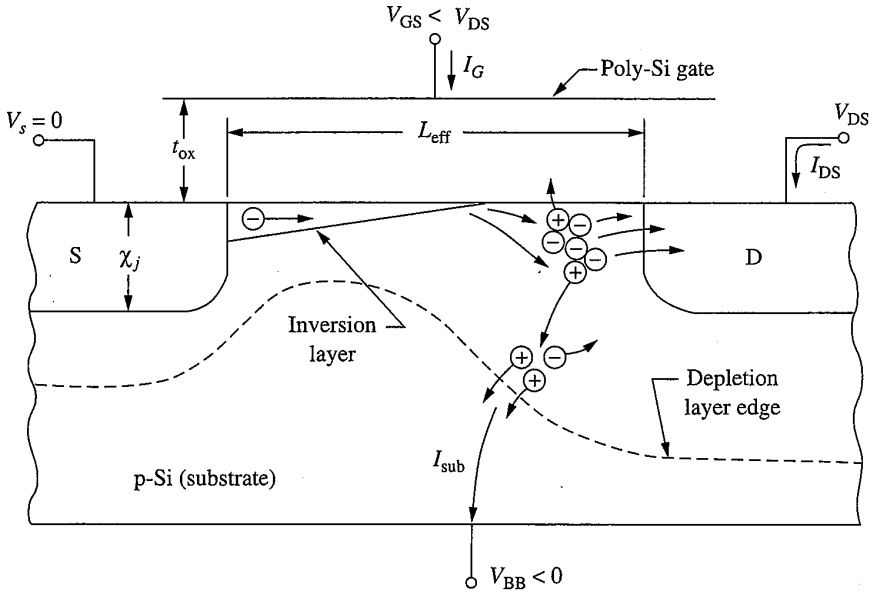


FIGURE 2
Impact ionization in saturation mode for NMOS shown as electron (−) and hole ⊕ generation in the transistor pinch-off region. Holes are swept to a p^- substrate by the electric field as substrate current, I_{sub}. (*After Sabnis, Ref. 1.*)

Micron Ex. 1029, p. 688
Micron v. YMTC
IPR2024-00911

and/or hydrogen.[9] The simplest model is that hot carriers bombard the interface against a retarding electric field. The carriers break the Si–H bonds and cause a trap. This interface damage is shown in Fig. 3. A measure of the interface damage is ΔD_{it}, and it can be measured with the charge-pumping technique.[10]

The maximum interface damage has been shown to occur at the gate/drain bias condition at which the substrate current is a maximum. The gate : drain voltage ratio is typically about 1 : 2. An example of degraded *I-V* characteristics of an n-channel MOSFET is shown in Fig. 4*a*. Notice that the drain current is lower in the degraded curve than the fresh device curve. This decrease in the drain current is due to mobility degradation caused by interface trap generation.

The main degradation mechanism for pMOSFETs is believed to be trapped electrons in the gate oxide near the gate/drain interface.[4,11] In the half-micron or longer channel length transistor, the degradation seems to be greatest at a gate/drain bias condition at which the gate current is a maximum. The gate bias voltage is typically about $\frac{1}{4}$ of the drain bias voltage. However, at shorter channel lengths, degradation under maximum I_{sub} bias conditions could be larger.[12] Degradation parameters measured are similar to those for n-channel MOSFETs.

An example of degraded *I-V* characteristics of a p-channel MOSFET is shown in Fig. 4*b*. Notice the degradation of the drain current is an increase from the fresh device. The increase is due to channel shortening caused by electron trapping near the gate/drain region.[2,11] This trapping causes a decrease in $|V_T|$, an increase in subthreshold leakage current, and a decrease in the punch-through voltage.[3]

12.2.2 Equations of HCI

Degradation of the n-channel MOSFET transistor parameter under HCI stress has been shown[2] to follow the expression

$$\Delta p \sim \{(I_d/W)(I_{sub}/I_d)^{\phi_{it}/\phi_i} \cdot t\}^n \tag{12.5}$$

where Δp = change in a parameter, e.g., V_T, g_m, I_{dsat}, I_{dlin}
I_{sub} = maximum substrate current (A)
I_d = drain current at the maximum substrate current bias condition (A)
W = transistor width (meter)
ϕ_{it} = critical energy for device damage (eV)
ϕ_i = impact ionization energy (eV)
t = time (sec)
n = empirically fitted exponent

The maximum substrate current can be approximated by[2]

$$I_{sub} \approx 2I_d \exp(-\phi_i/q\lambda\mathscr{E}_m) \tag{12.6}$$

where λ = hot electron mean free path
$\mathscr{E}_m$ = maximum electric field

The lifetime τ has been shown[2] to follow the expression

$$\tau \approx (I_{sub}/W)^{-\phi_{it}/\phi_i} \tag{12.7}$$

Micron Ex. 1029, p. 689
Micron v. YMTC
IPR2024-00911

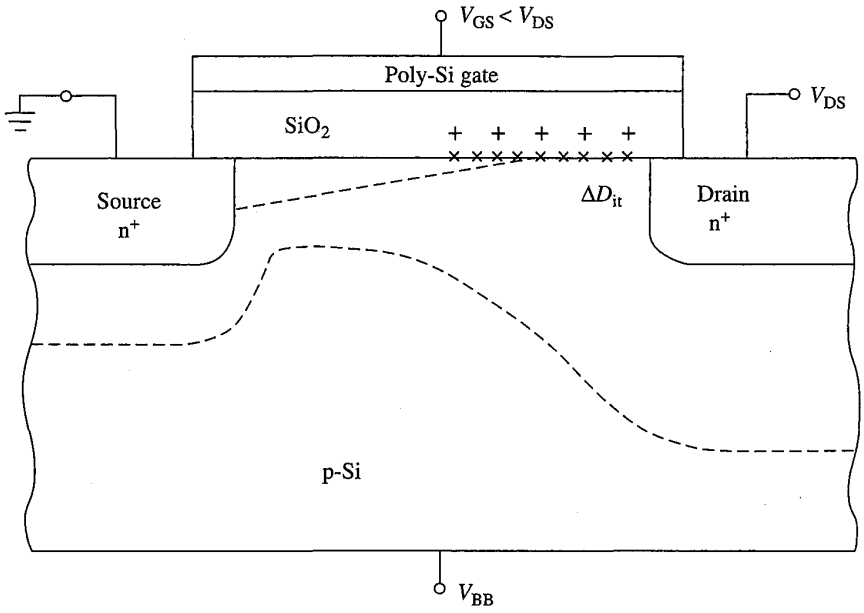


FIGURE 3
Diagram of trapped charges (+) and interface damage ΔD_{it}, or interface traps, during hot carrier injection in an NMOS transistor. Note that the trapped charges and interface damage are toward the drain end. (*After Sabnis, Ref. 1.*)

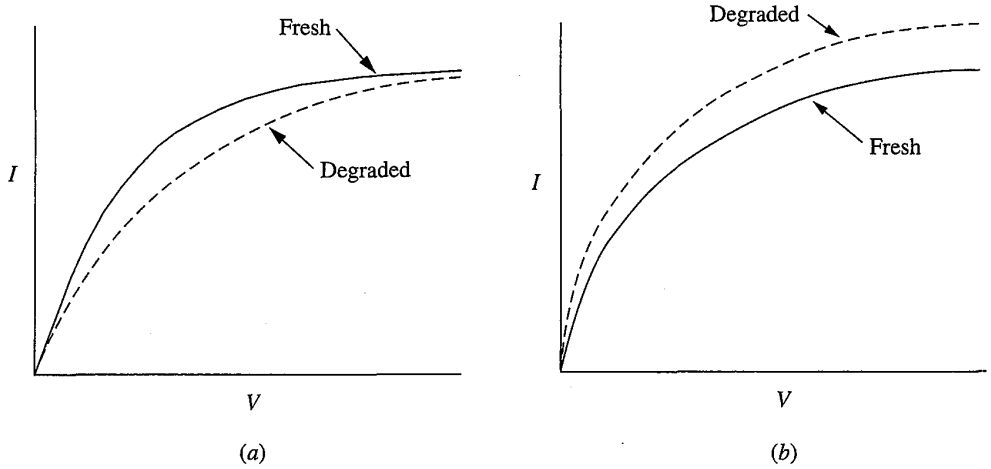


FIGURE 4
(*a*) Degradation of an *I-V* curve of an n-channel transistor after HCI stress.
(*b*) Enhancement of an *I-V* curve of a p-channel transistor after HCI stress.

Micron Ex. 1029, p. 690
Micron v. YMTC
IPR2024-00911

where $\phi_{it}/\phi_i \approx 3$. The parameter τ is the dc lifetime of the device calculated at a given percentage of degradation of the parameter p. Typically a 10% in degradation of p has been used as a measure of τ, but this value has been chosen historically and has no physical meaning. The real value of Δp should be chosen from a circuit degradation viewpoint and could differ for different applications.

From an experimental point of view, Eqs. (12.5) and (12.7) can be expressed as in Eqs. (12.8) and (12.9) for NMOS, that is

$$\text{NMOS}: \quad \Delta p = At^n, \qquad n = 0.3 \text{ to } 0.7 \tag{12.8}$$

$$\tau = B(1/I_{\text{sub}})^m, \qquad m = 2.5 \text{ to } 3.5 \tag{12.9}$$

PMOS transistor HCI degradation is not so widely researched as NMOS, but recent studies have shown that from an empirical point of view, the HCI degradation can be expressed in a format similar to NMOS,[13,14] or

$$\text{PMOS}: \quad \Delta p = A't^{n'}, \qquad n' = 0.3 \text{ to } 0.7 \tag{12.10}$$

$$\tau = B'(1/I_{\text{gate}})^{m'}, \qquad m' = 2 \text{ to } 3. \tag{12.11}$$

Notice that the PMOS lifetime is now a function of the maximum gate current, I_{gate}. [However, for PMOS transistors at the quarter-micron channel length region, the lifetime has been shown to follow the maximum I_{sub} expression, similar to Eq. (12.9).[12]]

In Eqs. (12.8–12.11), the parameters n, m, n', and m' are empirically determined. Examples of Eqs. (12.9) and (12.11) are illustrated in Figs. 5*a* and 5*b* for NMOS and PMOS, respectively.

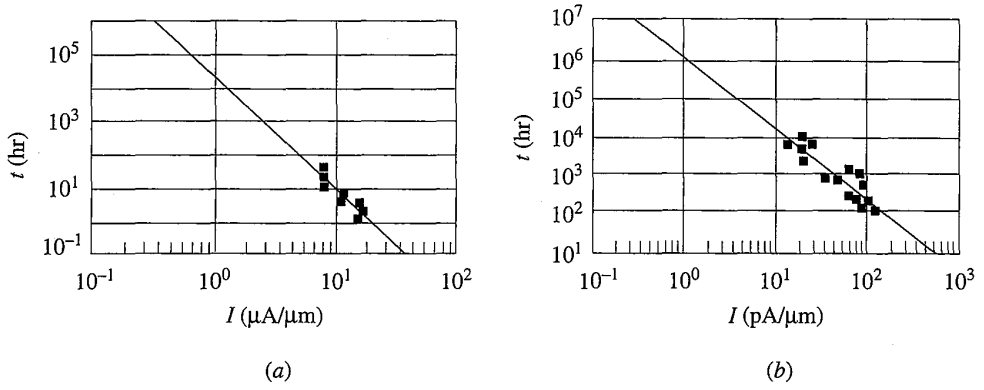


FIGURE 5
(*a*) Example of plotting lifetime data vs. I_{sub} for NMOS under several accelerated HCI stress conditions to project for lifetime to use condition. (*b*) Example of plotting lifetime data vs. I_{gate} for PMOS HCI under several accelerated stress conditions to project for lifetime to use condition. The regression fit of data in (*a*) is $\log t$ (hr) $= -3.34 \log I(\mu\text{A}/\mu\text{m}) + 4.3$. The regression fit of data in (*b*) is $\log t$ (hr) $= -1.81 \log I(\text{pA}/\mu\text{m}) + 5.97$.

Micron Ex. 1029, p. 691
Micron v. YMTC
IPR2024-00911

12.2.3 HCI Summary

In this section, transistor degradation due to HCI effects has been discussed, and basic HCI equations have been given. It will be shown in a later section that HCI degradation is a limiting reliability factor to device scaling. The gate oxide quality has been reported to play a key role in HCI lifetime[15] and is expected to affect the parameters A and B in Eqs. (12.8) and (12.9). A more recently elucidated phenomenon related to HCI is hot electron–induced punch-through degradation.[16] This phenomenon is more pronounced in p-channel MOSFETs and could impose an even greater reliability limitation in the ULSI regime than the traditional HCI degradation. The key to HCI lifetime improvement is to reduce the maximum electric field through *drain engineering* and to produce trap-free, damage-free, and clean gate oxides.

12.3 ELECTROMIGRATION

Electromigration (EM) is the transport of metal ions through a conductor, and it results from the passage of direct electrical current. Electromigration is perhaps the most widely known and extensively researched failure mechanism of integrated circuits. Studies began in the 1950s with the development of the semiconductor.[17,18] With the discovery of crack formation in the aluminum conductor leading to failures in ICs in the late 1960s,[19] electromigration studies in thin Al films began in earnest. Although most metals will electromigrate, Al has by far been the metal of greatest research because of its common usage as the interconnect material of choice.

Electromigration is a combination of thermal and electrical effects on mass motion. The higher the temperature, the easier it is for the metal ions to electromigrate. For bulk metals, electromigration occurs at about three-fourths the melting temperature (T_m in K), whereas for a metallic polycrystalline thin film it occurs at about one-half of T_m. Thus, noticeable electromigration is observed at 423 K and above for Al.

The mass motion of Al ions is visualized in Fig. 6. At temperatures of about $\frac{1}{2}T_m$, the Al ions are sufficiently thermally agitated in the lattice potential well that there is a finite probability that some ions are situated at the top (or saddle point) of the potential well (dictated by Maxwellian statistics). These "activated" ions are essentially free of the lattice and are free to diffuse out of or fall back into the well. In the presence of an electric field, the activated ions experience electromotive forces $F_{\mathscr{E}}$ and F_p, where $F_{\mathscr{E}}$ is the force on the Al^+ ion due to the electric field and F_p is the *electron wind* force exerted on the Al^+ ion due to the momentum exchange of the electrons in the electric current.[20] F_p and $F_{\mathscr{E}}$ are opposite in direction and $F_p \gg F_{\mathscr{E}}$. Thus, a finite portion of the Al^+ ions are "pushed" in the direction of the electron current flow. Al ions move upstream while vacancies move downstream. Agglomerations of vacancies form voids, and large voids can grow completely across the width of the Al lines to cause circuit failures.

Micron Ex. 1029, p. 692
Micron v. YMTC
IPR2024-00911

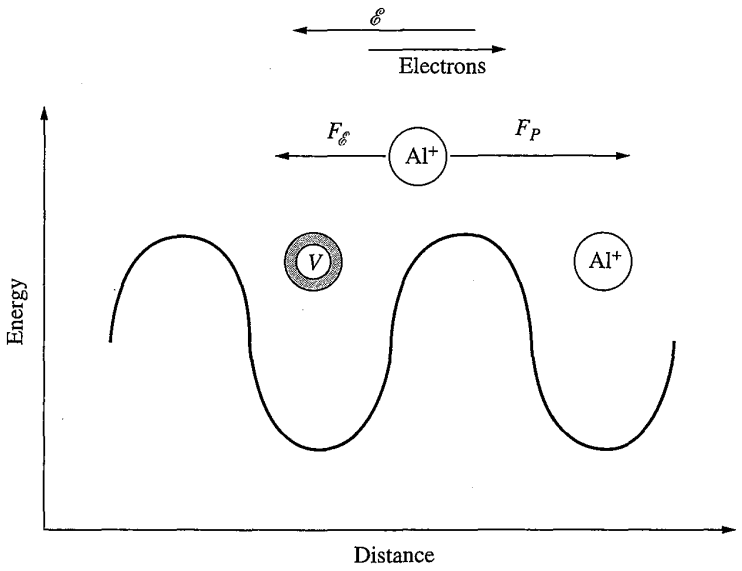


FIGURE 6
Schematic of a positive Al ion (Al^+) in a lattice potential well under opposing electric field ($F_ℰ$) and electron wind (F_p) forces. The letter V represents the vacancy left behind after the Al ion jump.

For Al to move from one well to another, a vacancy must be created, and an ion-vacancy jump process must also occur. The activation energy E_a for this sequence in pure, bulk Al is 1.48 eV.[21] The E_a for Al migration in a grain boundary has been found to be much lower, near 0.5 to 0.6 eV.[22] This difference can be explained by the fact that there is an ample supply of vacancies in grain boundaries and Al/oxide surfaces, so that the ion-vacancy jumping process is made much easier. Hence, electromigration in polycrystalline thin film is predominantly a grain-boundary, diffusion-driven mechanism.

The basic flux equation in electromigration, for an ideal metal in the absence of a temperature gradient, is given by[18,23]

$$F_m = ND_o/kT(Z^*q\mathscr{E})\exp(-E_a/kT) \tag{12.12}$$

where F_m = ion flux
N = density of atoms
D_o = diffusion coefficient
T = temperature in K
k = Boltzmann constant
Z^*q = effective ionic charge
$\mathscr{E}$ = electric field
E_a = activation energy

To have any void growth, there must be vacancy flux divergence, which is related to the ion flux divergence. This is dictated by the continuity equation,

$$dC_v/dt = -\nabla \cdot \boldsymbol{F}_v + (C_v - C_v^o)/\tau \tag{12.13}$$

Micron Ex. 1029, p. 693
Micron v. YMTC
IPR2024-00911

where C_v = vacancy concentration
F_v = vacancy flux
C_v^o = thermal equilibrium vacancy concentration
τ = average lifetime of vacancy

Under a steady-state condition,

$$dC_v/dt = 0$$

so

$$\nabla \cdot \boldsymbol{F}_v = (C_v - C_v^o)/\tau \tag{12.14}$$

If voids are to grow, $C_v \neq C_v^o$, thus $\nabla \cdot \boldsymbol{F}_v \neq 0$.

Figure 7 illustrates how grain boundary triple points and defects in the Al film can cause vacancy flux divergence, which leads to void and hillock formation. Temperature gradients and other structural inhomogeneities (e.g., irregular grain size) can also cause flux divergence.

Examples of electromigration voids in Al/0.5% Cu/1% Si films are shown in Fig. 8. Figure 8*a* shows a large void severing the entire width of the line, causing an electrical open. Figure 8*b*, however, shows that the large void did not cause an electrical open, as the underlying refractory metal barrier, Ti/TiN, is still intact. The temperature of test used for Fig. 8*b* was significantly less than one-half of the melting temperature of the refractory Ti/TiN barrier, so electromigration in the barrier was negligible. Thus, the shunting of the void prevented the line from becoming electrically open. The underlayer barrier of Al is one preventive measure for electromigration in ULSI processing. Other preventive measures will be discussed later.

12.3.1 Effect of Stress on Electromigration

The ion flux in Eq. (12.12) is an oversimplification. In reality, as Al ions move from one end of an interconnect to another under electromigration, a mechanical stress gradient is created in the line. This stress gradient will create an additional ion flux, opposite to the electron wind flux, due to the vacancy concentration difference

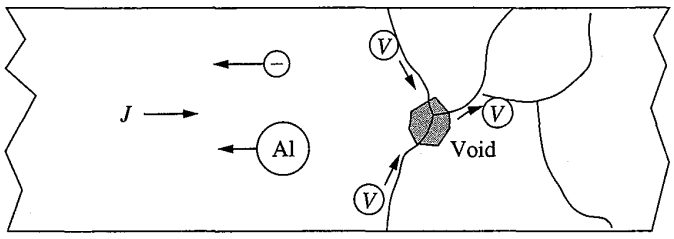
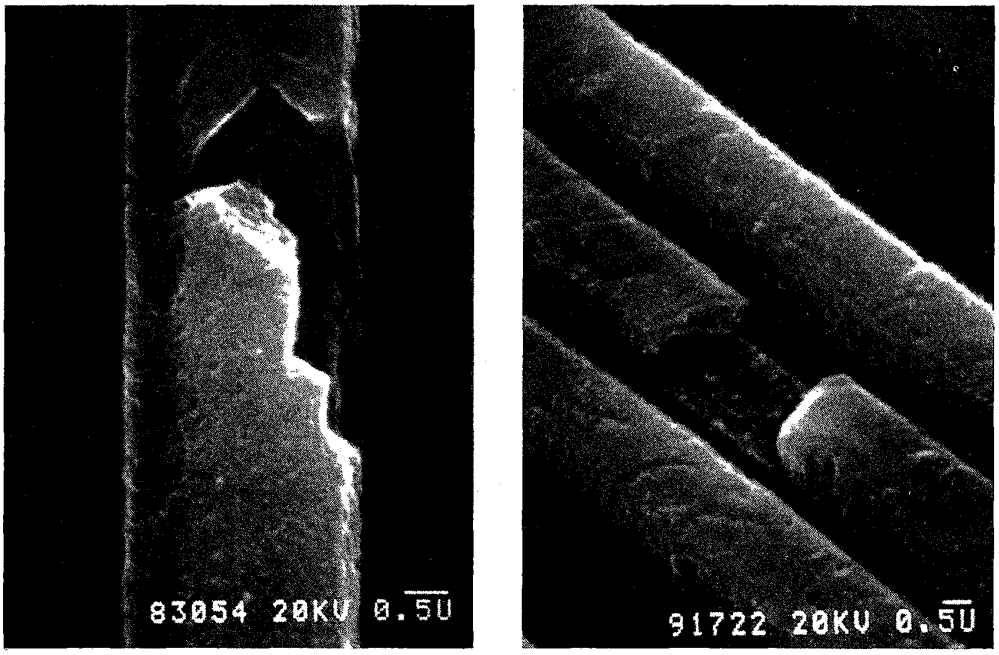


FIGURE 7
Schematic of void formation due to vacancy v flux divergence at grain boundary triple points. A triple point is defined when two or more grain boundaries intersect.

Micron Ex. 1029, p. 694
Micron v. YMTC
IPR2024-00911



no-barrier (*a*) with barrier (*b*)

FIGURE 8
(*a*) SEM micrograph of electromigration void in an Al–Cu–Si metal line without a barrier underlayer. (*b*) SEM micrograph of electromigration void in an Al–Cu–Si metal line with an underlayer barrier refractory metal showing that the barrier is still intact below the void.

between the cathode and anode ends. There are more vacancies in the cathode end than the anode end. The anode end of the interconnect becomes more compressively stressed than the cathode end.

The ion flux F_m, in the presence of the stress gradient, is given by[24]

$$F_m = ND/kT[Z^* q\mathscr{E} - \Omega(d\sigma_n/dx)] \tag{12.15}$$

where Ω = atomic volume
σ_n = stress normal to the grain boundary

The diffusion coefficient D is itself a function of σ_n such that it decreases exponentially with increasing stress. When a passivation overcoat is applied, the additional stress in the Al line results in a slower electromigration process. It has been shown that passivated Al lines have lifetimes that are about 10 times longer than nonpassivated lines.[25,26]

In Eq. (12.15), it is possible that the stress gradient term is large enough to equal the electron wind term. In that case then, $F_m = 0$ and

$$d\sigma_n/dx = Z^* q\mathscr{E}/\Omega \tag{12.16}$$

In general, Al films cannot support such a large stress gradient without growing hillocks or whiskers, which then relieve the stress. Hence, F_m rarely is zero. But

Micron Ex. 1029, p. 695
Micron v. YMTC
IPR2024-00911

by utilizing a pinhole-free passivation layer, hillock and whisker formation can be suppressed to minimize the electromigration damage.

From the drift velocity electromigration experiment, it has been shown that in very short strips (< 20 microns in length),[27] the stress gradient can be so high that Eq. (12.16) is satisfied. Notice that $d\sigma_n/dx$ in a strip is the stress at the anode end minus the stress at the cathode end divided by the length of the strip. Rewriting Eq. (12.16) for Δx, we have

$$\Delta x = \Delta\sigma_n \Omega / Z^* q \mathscr{E} \tag{12.17}$$

where Δx is the critical length, commonly referred to as the *Blech length.* For strip lengths less than Δx, electromigration damage cannot occur. Estimation of Δx from experimentally measured Z^* and typical stress values in Al films has resulted in general agreement with Eq. (12.17). An interesting concept in electromigration prevention is to design the interconnect lengths in segments of less than 20 microns at a time; however, this is impractical.

In multilevel interconnect processes, e.g., two to four layers of Al interconnects, Eq. (12.15) shows that the quality and stress values of the interlevel dielectric (ILD) films are important in determining the lifetime of the metal lines. In fact, the thermal expansion coefficients of ILD films and the thermal process history of the entire interconnect module are key parameters in determining the electromigration lifetime.

12.3.2 Electromigration of Lines

The reliability of a metal interconnect is most commonly described by a lifetime experiment on a set of lines to obtain the median time to failure (MTF). The data for the actual time to failure of each line are plotted on a lognormal graph, and the value of T_{50} (time for 50% of the lines to fail) is extracted along with the lognormal shape parameter (σ). The stress experiment involves stressing the lines at high current densities (usually from 10^6 to 3×10^6 A/cm^2) at temperatures ranging from 150 to 250°C. The failure criterion is typically an electrical open for nonbarrier conductors or about a 20% increase in line resistance for barrier metallization. The general MTF expression is

$$\text{MTF} = AJ^{-n} \exp(E_a/kT) \tag{12.18}$$

where A is a material constant based on the microstructure and geometric properties of the conductor, J is the current density, and E_a and kT have the previous definitions.

Equation (12.18) was first proposed by Black[28,29] and is based on a simple theoretical argument about the rate of mass transport occurring as a result of momentum exchange between the thermally activated ions and the electron wind. In that argument, the rate of mass transport, R, was proportional to the product of the momentum transferred in each collision ($q\mathscr{E}/v = q\rho J/v$) with the number of electrons available for collision ($N = J/q$), the effective cross section, and the activated Al ion density. If we assume R is inversely proportional to MTF, it then follows that MTF $\propto J^{-2}$ and $n = 2$ in Eq. (12.18). This argument is, however, not consistent with the ion flux divergence requirement. For Al lines, n has been found experimentally to range from 1 to 3 or even higher.

Micron Ex. 1029, p. 696
Micron v. YMTC
IPR2024-00911

It has been argued that $n = 1$ in the absence of thermal gradients in the conductor and $n = 3$ when thermal gradients are taken into account.[30] A number of researchers have argued that conceptually $n = 2$ is correct.[31,32] Most recently, by taking into account the physical observation of void density and size during electromigration of a conductor line,[33] it was concluded that $n = 2$. In that work, the number of voids and hillocks formed per conductor length by electromigration was shown experimentally to be proportional to the current density. The average growth rate of each void was observed to be proportional to the current density. Since $(\text{MTF})^{-1}$ is the product of (number of voids) $\times$ (void growth rate), it follows that $n = 2$.

Once T_{50} and σ are determined in the accelerated stress experiment, extrapolation to the circuit design condition gives the lifetime of the metal line. Typically, the electromigration goal is 10 years at the worst-case field usage condition for a 0.1% failure distribution ($T_{0.1\%}$) of the metal strips under test.

Equation (12.18) contains the geometry dependence in the parameter A. Simplistically, $A \approx wt$, where w is the width of the line and t is the thickness. Experimentally, it has been shown that the width dependence of MTF is a function of the ratio of the grain size d of the film and the width of the conductor w. As w/d decreases, the MTF will increase due to the *bamboo effect.* Figure 9 illustrates this effect for Al films without a barrier.[34] The minimum MTF generally occurs at about $w/d \approx 2$ to 4. As the grain size increases with respect to the line width, the interconnecting paths of the triple points parallel to the length of the conductor decrease and even vanish. So the probability of voids accumulation along the grain boundaries is reduced, thus retarding the electromigration process. Figure 10 shows a *bamboo* grain conductor versus a nonbamboo line. The final grain structure of the Al film not only is a function of the metal deposition conditions but is also strongly dependent on post-metallization annealing temperatures and the texture of the underlying surfaces.

In Al conductors with an underlayer barrier, the bamboo effect of MTF still holds and is shown in Fig. 11.[35] Although the Ti/Al–Cu/Ti layered film exhibits improved overall line MTF, it still shows a critical line width value at which the MTF is at a minimum. The line width dependence of MTF has significant implications from a design point of view and is usually characterized during the process development phase of any new metallization process.

12.3.3 Contact and Via Electromigration

Electromigration occurs not only in conductor lines but also in contacts and vias. Contact electromigration resembles the classic Al spiking phenomenon, except that the effect is activated by electric current. It is driven by the Si ion migration in the Al conductor in the direction of electron wind.[36] Figure 12 illustrates this process in an Al–Cu–Si conductor. As Si is driven from the anode near the contact by the electrons in the current, Si in the substrate is diffused into the Al to satisfy the Si solubility in the Al. Al atoms diffuse into the substrate to occupy the vacated Si positions. As time goes on, Al spiking can puncture the n^+ (or p^+) junction to cause junction leakage. The effect is particularly noticeable in shallow junction processes. The migrated Si

Micron Ex. 1029, p. 697
Micron v. YMTC
IPR2024-00911

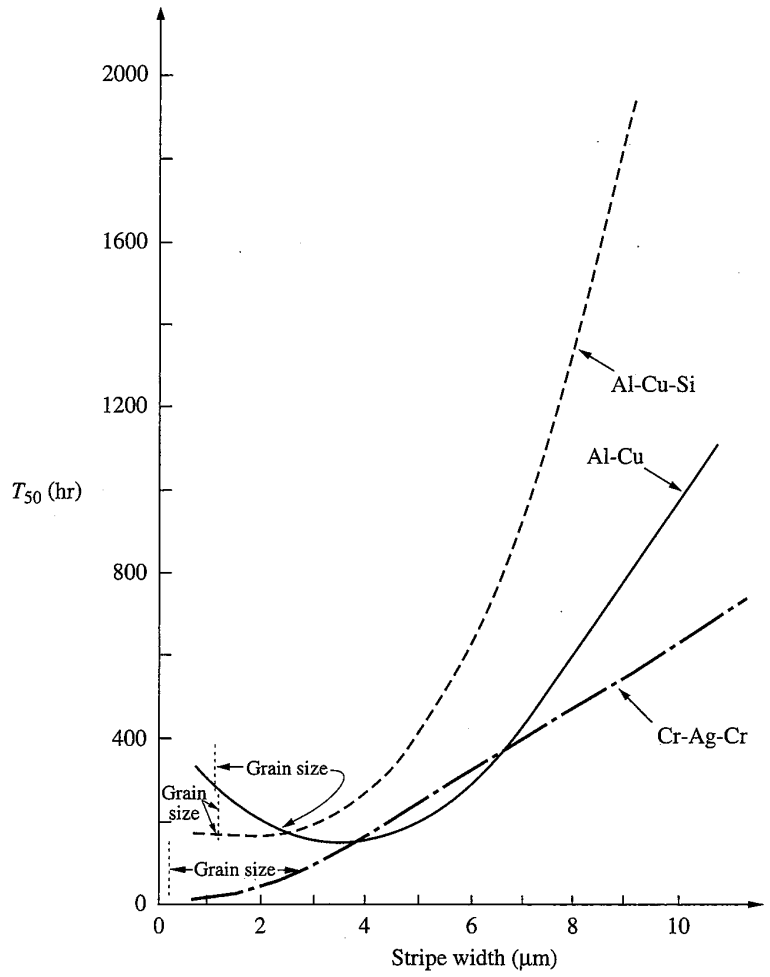


FIGURE 9
EM lifetime (T_{50}) as a function of Al line width for three different Al-based alloys. The *bamboo* effect is the increase in lifetime as the width decreases below a certain critical width. This width is typically about 2 to 4 times the average grain size in the metal film. (*After Scoggen et al., Ref. 34.* ©1975 IEEE.)

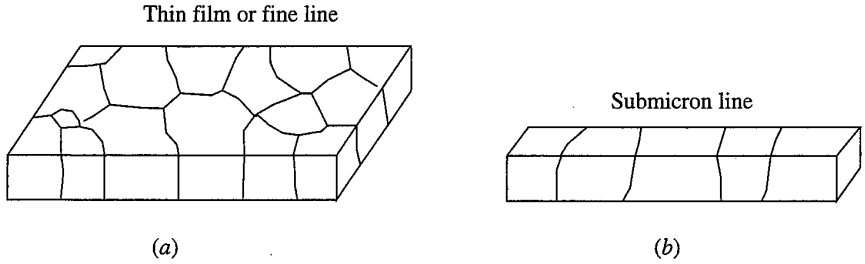


FIGURE 10
Sketch of (*a*) a nonbamboo structure in a wider line dominated by triple points, versus (*b*) a bamboo-like grain structure in a submicron line.

Micron Ex. 1029, p. 698
Micron v. YMTC
IPR2024-00911

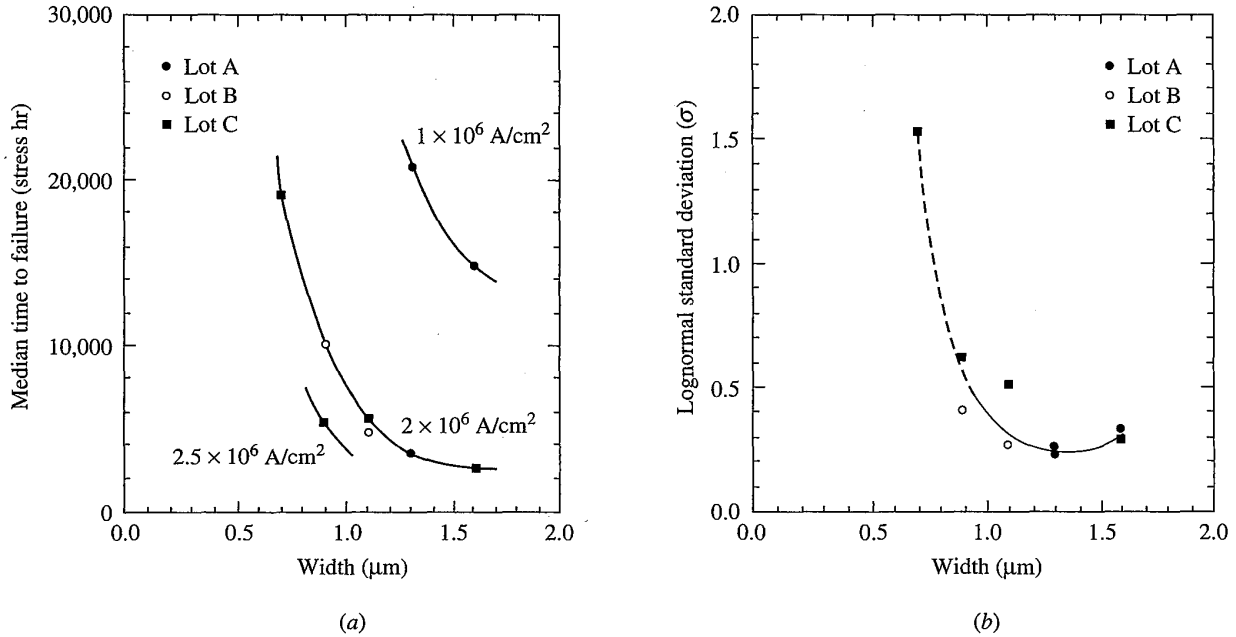


FIGURE 11

(*a*) EM lifetime data for three lots of Ti–Al–Ti metallization tested at three different current densities: 1×10^6, 2×10^6, and 2.5×10^6 A/cm^2. The characteristic *bamboo* effect is observed. (*b*) An increase in the lognormal σ with decreasing line width was measured for all three lots. *(After Rathore et al., Ref. 35.)*

Micron Ex. 1029, p. 699
Micron v. YMTC
IPR2024-00911

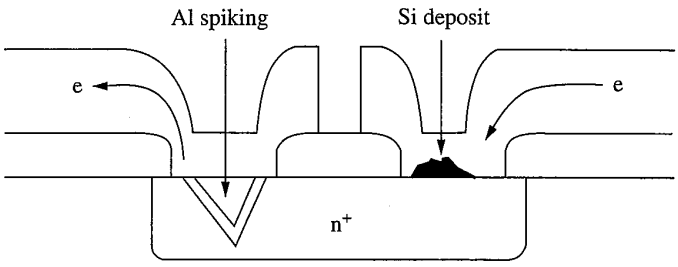


FIGURE 12
Diagram of the contact electromigration process without a barrier metal. Al spiking is shown on the left contact and Si pileup is shown in the right contact. Note the direction of the electron flow.

atoms can also deposit in the contact window in the anode side causing resistive or even open contacts.

The equation for contact electromigration (in the absence of barrier metal) has been shown to follow the empirical expression[36,37]

$$\text{MTF} \propto x_j^2 (W^\alpha / I)^n \exp(E_a / kT) \tag{12.19}$$

where x_j is the junction depth, W is the contact window dimension, I is the current, and n and α are exponents. Unlike the line EM equation in Eq. (12.18), the current I is used in Eq. (12.19). It has been shown that n is a function of Joule heating and approaches 2 for lower current values. Since x_j is less than 1, MTF decreases with the square of the scaling dimensions.

To overcome this problem, a refractory barrier film has been developed that is impermeable to the diffusion of Si or Al. Most of the popular barrier systems are Ti/TiN and TiW barriers. However, even in these films contact EM can still occur if the barrier quality (porosity) is poor and if sufficient current is applied. For all practical purposes in today's field usage conditions, these barriers have solved the contact EM problem.

Note that poor step coverage of the Al film on the contact side walls can also lead to electromigration because the flux divergence is aggravated at those points. This effect is similar to the line electromigration effect discussed in Section 12.3.2 and should be distinguished from the contact electromigration mechanism.

Via electromigration became an issue in multilevel metallization technology. As one metal layer is connected to another metal layer through a via window, the current crowding at the via interface has been shown to exhibit a *liftoff* at the interface.[38,39] This can occur in Al–Al vias or even Al–barrier–Al vias. Figure 13 illustrates this failure mechanism. The driving force is the flux divergence at the interface due to grain size difference, which is aggravated by a very thin interfacial oxide film. This driving force can also be accentuated by the difference in the thickness of the metal layers. However, even with the best cleaning processes, via interfaces can still lift off during long-term current stressing. This failure mechanism should be characterized in a given process because it could be the lifetime-limiting factor and should be considered in the circuit design stage.

Micron Ex. 1029, p. 700
Micron v. YMTC
IPR2024-00911

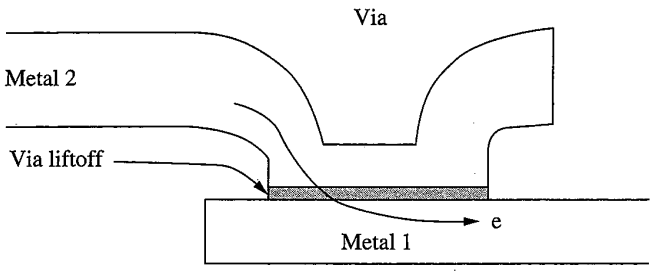


FIGURE 13
Diagram of a via EM failure mechanism due to via liftoff. The separation at the metal 2 and metal 1 interface is denoted by the shaded area.

12.3.4 Methods to Improve Electromigration Lifetime

To improve electromigration lifetime, one must reduce the magnitude of the flux, F_m, in Eq. (12.15) or the divergence of F_m. Examination of the variables that control F_m provides direction for improving electromigration lifetime. From Eq. (12.15) reduction in F_m can be achieved through the following:

1. Reduction in diffusivity of Al, D
2. Reduction in the driving force, $Z^* q\mathscr{E}$
3. Increase in the back pressure, $d\sigma_n/dx$

Reduction in the driving force $Z^* q\mathscr{E}$ is difficult and impractical, since $\mathscr{E}$ for ULSI devices is increasing owing to the circuit performance requirements ($\mathscr{E} = \rho J$). Reduction in the effective charge Z^* is difficult since this is an intrinsic property of the scattering process in Al.

Greater benefit can be achieved through the reduction in the diffusivity, D, of Al. The focus of research has been to reduce the grain-boundary diffusivity of Al by the addition of various solute atoms, most notably Cu. Cu is known to improve electromigration lifetime in Al strips.[40] Other solute atoms such as Mg, Mn, Ti, and Pd are also effective in improving EM lifetime.[41] The mechanism of diffusivity reduction is believed to be the retardation of atomic flux due to the segregation of these impurity atoms at the grain boundaries and their subsequent interaction with the migrating Al ions.[23] We can conceptualize this interaction as a modification of the potential well (Fig. 6) due to the presence of impurity atoms that reduce the number of "activated" Al ions in the grain boundaries. Thus, up to 4% Cu doping in Al has shown continued improvement in EM lifetime as well as in E_a.[42]

An interesting property of Cu is that its solubility in Al is only 0.05 at % at 230°C.[43] Typical doping levels in today's semiconductors are at 0.5%, 2%, or 4%. The additional Cu is then segregated at the grain boundary, and the greater the segregation of Cu, the more improvement in lifetime. However, values of Cu in Al $>$ 8 at % have been shown to reduce EM lifetime.[42] Cu also forms secondary compound, a theta phase particle with Al, $CuAl_2$, which is not necessarily situated in the Al grain but which serves as sources of Cu atoms to the Al grain boundary. This situation is particularly advantageous, because Cu also electromigrates, so a supply of Cu will keep grain-boundary diffusivity at a low level for a longer time.

Increasing the back-pressure term, $d\sigma_n/d\,x$, will reduce F_m. Earlier, from Eq. (12.17), we saw that short strips of conductor improve the back pressure and that for

Micron Ex. 1029, p. 701
Micron v. YMTC
IPR2024-00911

lengths below the Blech length of 20 microns, F_m goes to zero. However, designing interconnects with such short segments, interlaced with vias, is not a practical approach.

Efforts to reduce the flux inhomogeneity have been to modify the Al microstructure. Such microstructure modifications include the following:

1. Single-crystal Al stripes[44]
2. Conductor composed of chains of Al single grain[45]
3. Al films with large grain[46]
4. Ultra bamboo structures[47]

All these have shown improvements in EM lifetime. Ultra bamboo structures have shown improvement as much as 100 times.

Passivation overcoating, or ILD overcoating in multilevel metallization structures, modifies the conductor stress states and improves the lifetime.[23] Although the mechanism is not well understood, it is believed that the overcoating enhances the defect annihilation kinetics,[23] or increases the back pressure, or alters the surface diffusion kinetics. In fact, it has been shown that in narrow conductors, the surface diffusion mechanism contributes significantly to the EM drift velocity.[48]

Transition metal/barrier underlayering is becoming a preferred scheme in the ULSI process. These schemes have also shown marked conductor lifetime improvements ($\sim$ 10 times), e.g., TiN,[49] Cr,[50] and Ti–W.[51] The mechanism is believed to be the extra shunting provided by the underlayer, which allows voids to heal through localized joule heating. The barrier property improves the contact EM lifetime.[50] More exotic metal structures, such as Al metal interlayered with transition metal layers, have also demonstrated significant lifetime improvements ($\sim$ 50 to 100 times).[52] These sandwich structures of transition metals (Hf, Ti, Ta, or Cr) not only provide the shunting property like the single barrier layer just discussed but also modify the microstructure to cause a more bamboo-like conductor and prevent the linkage of voids between the top and the bottom Al–Cu layer.[53]

12.3.5 Electromigration Summary

Electromigration is a failure mechanism that has been widely studied in the last 30 years. The lifetime of any given interconnect technology needs to be evaluated separately because many processing variables determine the final lifetime value. Significant progress has been made to improve the electromigration lifetime, such as the use of Cu (or other transition metal) doping and barrier/refractory metal layering. However, the ever-increasing current density requirements in ULSI dictate that an even greater level of EM performance must be attained. Thus, research is continuing to reduce further the grain-boundary diffusivity in Al as well as the structural inhomogeneity. It appears that Al-doped thin films will still be the main interconnect at the 0.35-micron level, but below this level alternative thin films that have a better EM life, e.g., Cu or Au, may be considered. It remains to be seen how far below the 0.35 micron technology that Al-doped thin films, in combination with refractory layering, can continue to meet the high current density and reliability needs of ULSI.

Micron Ex. 1029, p. 702
Micron v. YMTC
IPR2024-00911

12.4 STRESS MIGRATION

Stress migration (SM) is the phenomenon of metal voiding in conductor lines that are under tension in the absence of electrical current. The voids grow and in time can totally sever a conductor line to cause circuit failures. This reliability problem is a relatively recent phenomenon compared to electromigration or HCI but becomes a major reliability concern because of the initial lack of understanding of the failure mechanism. Figure 14 illustrates a typical metal void caused by stress migration.[54] It looks rather similar to a void caused by electromigration.

The stress migration problem was first discovered on 64K DRAMs employing Al–Si sputtered metallization for 2.5 to 3.5 micron line width passivated with PECVD SiN from various semiconductor suppliers.[55] As early as 1982, failures had occurred during mild thermal or electrical stress in vendor qualification tests or even in unstressed parts in storage.[56] Failure analysis showed metal opens in bit lines over steps as well as metal voids in flat topography.

Early phenomenological studies showed that the maximum SM failure was at 180°C. Failures were caused by N_2 contamination during sputtering,[57] by large silicon nodules in the Al,[58] and by topographical steps.[54] Metal voids had a strong functional dependence on the compressive stress of the passivation films, with the Al tensile stress varying with the thickness of the passivation layer.[54] A more rapid cooling rate from the passivation furnace produced fewer voids.[54] The stress-induced voids also degraded the electromigration lifetime.[54] Large silicon precipitates provided sources of vacancies.[58] The back-end assembly process with ceramic-glass-sealing temperatures at 450°C increased the metal void size and density.[59]

Experimental studies have shown that the SM lifetime on conductor lines tends to follow the expression

$$\tau \propto (\text{width})^m \times (\text{thickness})^n \tag{12.20}$$

where m and n are experimentally determined constants that vary from 2 to 8.[60,61] The effect of Cu addition was studied, and it was shown that as little as 0.1% Cu added to Al–Si alloys substantially suppressed the open-failure rate.[60]

Stress-induced voids can be classified into two major categories, a wedge-shaped and a slit-shaped void. The wedge-shaped voids form at the edges of conductor lines and tend to peak in lines of 2 to 3 micron width. These can be observed even at the end of wafer processing. Slitlike voids are not easily observable after processing, but nearly all open-conductor failures have been caused by them.

12.4.1 Stress Migration Mechanism

Continuing research is devoted to further understanding the mechanisms of void growth and to fully explaining all the experimental data. However, based on the many studies thus far, some general understanding has been developed on the basic mechanisms of void formation and growth. Because of the importance of stress migration in ULSI, they will be discussed briefly.

Micron Ex. 1029, p. 703
Micron v. YMTC
IPR2024-00911

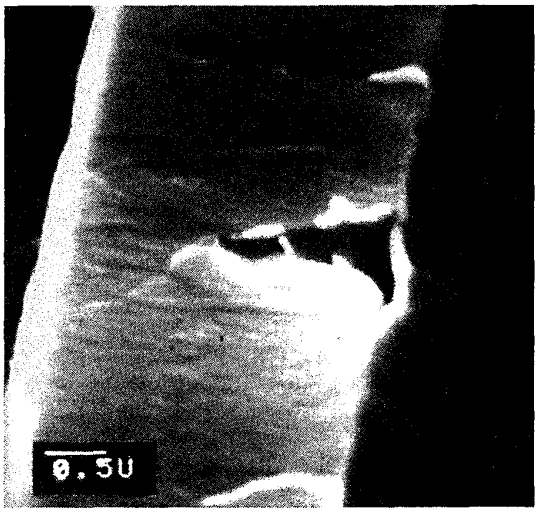


FIGURE 14
SEM micrograph of a stress-induced void in aluminum interconnect. Void was exposed after Ar back-sputter removal of passivation film. (*After Yue et al., Ref. 54.* ©1985 IEEE.)

It is generally believed that void nucleation and growth in Al conductors are a result of stress relaxation in a confined environment. The confined environment is provided by the surrounding passivation. Unpassivated Al lines cannot produce voids that cause failures because the free surfaces act as void sinks and prevent the accumulation of voids. The driving force is the tensile stress in the Al from the thermal expansion mismatch between the Al line and the surrounding passivation films and the Si substrate during the manufacturing process.

The stress in the confined Al line at the end of process, $\sigma_o(T)$, is given by[62]

$$\sigma_o(T) = f_c E(\alpha_{\text{Al}} - \alpha_{\text{Si}})(T_d - T) \tag{12.21}$$

where T is the temperature at the end of processing (such as room temperature), T_d is the passivation deposition temperature, E is the Young's modulus of Al, and α_{Al} and α_{Si} are the thermal expansion coefficients of Al and Si. The term f_c is a constraint factor that depends on the metal-insulator geometry and elasticity of each material. Note that at temperature T_d the stress in Al is zero by definition. An estimate of σ_o (room temperature) is 5×10^9 to 6×10^9 dynes/cm^2, by using $f_c \approx 1$, $E = 7 \times 10^{11}$ dynes/cm^2, $\alpha_{\text{Al}} = 23 \times 10^{-6}$/K, $\alpha_{\text{Si}} = 2.6 \times 10^{-6}$/K. This is comparable to or greater than the yield stress of Al film in confined Al lines.

The Al tries to relieve itself of this large tensile stress. The tensile stress results when the actual lattice spacing is greater than the thermal equilibrium spacing.[63] Since the Al is confined, it cannot relieve the stress by plastic deformation, although local internal plastic deformation could occur.[64] The Al tries to shrink but is prevented by the adhesion with the encapsulating, hard, rigid passivation. It is believed the major relaxation of stress occurs by diffusional or power-law creep, which causes void formation and growth. Thus, voids are created and grown as a by-product of stress relief. Smaller voids agglomerate into larger voids to reduce the total free energy of the system, and finally a void is large enough to sever a metal line.

Experimental data show that the maximum SM failure generally occurs in the temperature range between 150 and 200°C.[57,65] This can be explained by noting that in thermal creep the driving force due to stress decreases as temperature increases,

Micron Ex. 1029, p. 704
Micron v. YMTC
IPR2024-00911

but the vacancy diffusion rate increases exponentially with temperature. The product of these two driving forces gives a maximum somewhere between 150 and 200°C, with the peak temperature dependent on the thermal deposition conditions and the material properties of Al and the surrounding dielectrics.

A generalized power-law creep-rate model for SM is shown to be[65]

$$R(T) = C(T_o - T)^n \exp(-Q/kT) \tag{12.22}$$

where R = creep rate at temperature T
T_o = temperature of metal deposition
Q = activation energy associated with the diffusion process

Equation (12.22) can explain a maximum in R as a function of T, thus agreeing with the experimental data. By fitting Eq. (12.22) to a set of experimental data, Q was found to be 0.58 eV and $n = 2.33$.[65]

It has been observed that the void volume of an Al–Si conductor could be as large as 1 to 3% of the total volume.[54] (See Figure 15.) In explaining where the appreciable void volume comes from, we note that there are many sources of vacancies in a metal conductor. For example, supersaturation of vacancies by as much as 2.5% occurs in the metal conductor as it is pulled from a high-temperature passivation furnace at 400°C.[54] Si precipitation in Al during cooldown creates additional vacancies. Grain boundaries provide ample sources of vacancies. These vacancies can then agglomerate to form voids through such diffusional processes as Nabbarro-Herring creep (intragranular diffusion) or Coble creep (grain-boundary diffusion).[66] Recently it has been suggested that a H reaction[67] or even an Al–SiO_2 reaction[68] can also enhance void formation.

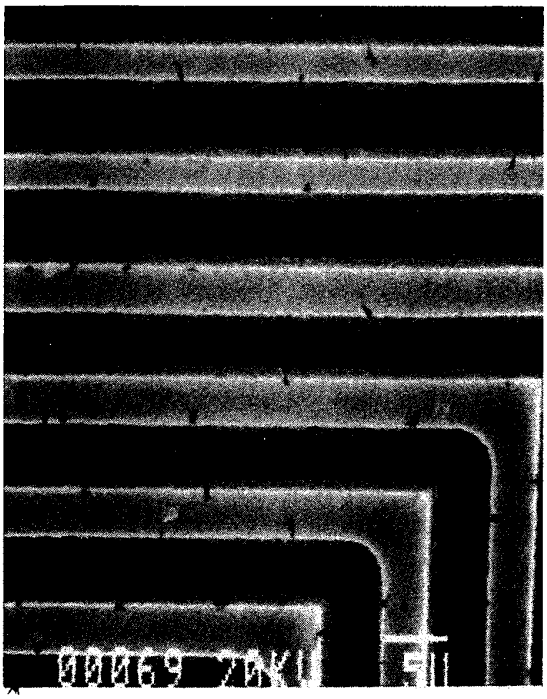


FIGURE 15
SEM micrograph of metal void density in aluminum interconnects, showing predominant location at line edges. (*After Yue et al., Ref. 54.* ©1985 IEEE.)

Micron Ex. 1029, p. 705
Micron v. YMTC
IPR2024-00911

A further complexity to the void formation problem is that the encasing passivation of the conductor is not perfectly hard and could deform under various IC usage environments. It has been reported that localized passivation separation from the Al conductor does lead to void formation.[69] Although this adds an additional element of complication to the stress migration problem, deformation of the passivation should remain a second-order effect.

12.4.2 Dependence of SM on Al Aspect Ratio

It has been shown that the stress migration failure rate is a function of the Al line width and thickness, with an increasing failure rate for narrower and thinner metal lines. This experimental finding predicts a greater stress migration hazard in the ULSI regime. The solid lines in Figs. 16*a* and 16*b* show the SM failure times as a function of line thickness and line width, respectively.[70] This experimental behavior could be explained by taking into account void growth by grain-boundary diffusion.[71] In this diffusion theory, the factor n in Eq. (12.22) is 1. It is shown that

$$\begin{aligned} \tau &= Cw^3/\sigma \qquad \text{for } r < 1 \\ \tau &= Ct^3/\sigma \qquad \text{for } r \geq 1 \end{aligned} \tag{12.23}$$

where r is the ratio of thickness to width and σ is the longitudinal stress acting normal to the grain boundary. X-ray diffraction shows that σ is not a constant but is a function of w and t, decreasing with both w and t.[71] The fit of Eq. (12.23) to experiment is shown by the dotted lines in Figs. 16*a* and 16*b* after taking into account the experimentally determined σ values. Note that there is general agreement between the grain-boundary diffusional theory and experiment. The most accurate way to determine the stress values in a conductor line is by x-ray diffraction. Several researchers have used x-ray diffraction to understand stress relaxation in conductor lines as they are temperature-cycled.[62]

12.4.3 Testing for Stress Migration

Currently, IC manufacturers have no standard way of testing for SM. Typically, test structures of very long conductor lengths at a minimum line width of 100 thousand to 10 million microns are tested. These structures are annealed at temperatures ranging from 150 to 250°C, and line resistances are measured at periodic intervals. Failure criteria are arbitrarily set at some specific resistance increase, such as 20%. The lack of a standard testing procedure and test structure for SM could explain some of the variability reported in the stress migration failure rates. Consequently, there is a need to have a more standard SM testing methodology in the future.

12.4.4 Stress Migration Prevention and Summary

The power-law creep model or diffusional creep model explains many of the experimental behaviors of stress migration. However, a comprehensive model is still

Micron Ex. 1029, p. 706
Micron v. YMTC
IPR2024-00911

FIGURE 16
(*a*) Experimental 10% failure time (●) and calculated lifetime (- - -) of 0.8-micron wide Al lines plotted as a function of line thickness and aspect ratio. (*b*) Experimental 10% failure time (●) and calculated lifetime (- - -) of 1-micron thick Al lines plotted as a function of line width and aspect ratio. *(After Yagi et al., Ref. 70.)*

Micron Ex. 1029, p. 707
Micron v. YMTC
IPR2024-00911

needed. The role of stress and stress relaxation has been identified to be very important in the formation and growth of voids. Based on this understanding, researchers have used alloy dopings, such as Cu, in Al to suppress the grain-boundary diffusion of stress migration. The nature of this suppression is similar to the Cu suppression in electromigration.

More recently, by using a refractory barrier or layered metallization scheme, failures due to the slit-type of voids were totally nullified as a result of the electrical shunting property of the refractory barrier. The refractory barrier is more resistant to stress migration than the Al conductor. Thus, for metal dimensions at the 0.5- to 0.35-micron level, it appears that the stacked Al conductor structure, with Cu doping, should be a sufficient safeguard against SM. For line widths below 0.35 microns, this approach may be inadequate against SM because of the dynamic interaction between SM voids and EM. Voids formed by stress relaxation could aggravate electromigration, and it is expected that this interaction will become worse as line widths become narrower. Further improvement could come from orientation control of Al films in ULSI, as a recent study has shown that void formation is accompanied by the minimization of surface free energy, which is achieved by creating the smallest-surface-energy plane, (111), of a face-centered cubic (fcc) crystal.[72]

12.5 OXIDE BREAKDOWN

Oxide breakdown is an important process reliability subject in the development of MOS and CMOS technologies. In the early days of DRAM devices, oxide breakdowns limited the yield and reliability of circuits. Therefore, understanding how oxides break down became important in MOS processing even before the studies of HCI, EM, or SM.

Oxide breakdown is described by the test methods used to cause the dielectric breakdown. In time-dependent dielectric breakdown (TDDB) a constant voltage is applied across the gate oxide at a given temperature. The leakage current across the gate oxide is monitored, and the time to breakdown is recorded when the current exceeds some value (e.g., 1 μA).

A second way of describing oxide breakdown is charge to breakdown, or Q_{bd}. In a Q_{bd} test, a fixed current in the Fowler-Nordheim (F-N) tunneling region is forced through the oxide and the voltage is monitored. When the voltage across the oxide suddenly decreases, the oxide is considered to have failed. The time to failure is designated by t_{bd}. Q_{bd} is the product of the fixed current and t_{bd}. The Q_{bd} test is important in nonvolatile memories, e.g., EPROMs and flash EPROMs, because the durability of tunnel oxides (typically 80 to 100 Å) used in these devices is characterized by Q_{bd}.

In the TDDB case, a relatively low electric field $\mathscr{E}$ is applied across the gate oxide (about 5 to 8 MV/cm). With TDDB, the times to failure could be in the thousands of hours. In the Q_{bd} case, a larger varying $\mathscr{E}$ field is applied across the thin oxide to achieve a constant tunneling current in the F-N region. The voltage usually is near the breakdown of the oxide, e.g., 9 to 10 MV/cm. With Q_{bd}, the time to failure is

Micron Ex. 1029, p. 708
Micron v. YMTC
IPR2024-00911

short, typically several minutes. Figures 17*a* and *b* compare the two measurement methods for TDDB and Q_{bd}.

12.5.1 Mechanisms of Oxide Breakdown

Oxide breakdown is generally believed to be caused by the positive charge buildup in the oxide near the injecting, cathode interface.[73] In oxide thicknesses greater than 120 Å, the source of these positive charges is traditionally believed to be due to impact ionization deep within the oxide[73] caused by the tunneling electrons as shown in Fig. 18*a*. Initially, the tunneling currents are extremely low. Positive charges drift back toward the cathode and are believed to be trapped at localized "weak spots."[73] These trapped positive charges lower the energy band and lead to further electron injection (Fig. 18*b*). The process leads to further impact ionization and positive charge trapping, resulting in a runaway process.[73]

The final high current I injected at localized spots at *positive charge-trapping* sites produces an I^2R heating sufficient to melt the SiO_2.[74] Before the catastrophic oxide rupture, oxide becomes leaky, and the leakage currents could cause circuits to be nonfunctional.[75]

Recent models of oxide breakdown in very thin oxides (< 100 Å) and at power supply voltages less than 5 volts have suggested that another source of the positive charges could be *hot holes* created at the anode side.[76] These hot holes are created by electrons that have tunneled completely through the thin oxide. Upon reaching the anode side, these electrons transfer their gained kinetic energy to deep valence electron(s) through Auger-type interactions, thus creating hot holes in the valence band at the anode side. These holes subsequently are injected (tunneled) back into the oxide by the electric field and are then trapped by the same type of localized *weak spots* as in the thick oxide model. The trapped positive charges cause band lowering and therefore cause enhanced electron tunneling at the cathode. The process leads to thermal runaway, similar to that in the thick oxide model.

Note that both the thick and thin oxide breakdown models hinge on the concepts of positive charge trapping and thermal runaway. The difference is in the source of the positive charges. It is plausible, for very thin oxides, that the probability is increased for an electron to tunnel across the thin layer without resulting in impact ionization in the oxide. Perhaps the real situation is a combination of the two processes described above, and the transition from a thick oxide breakdown regime to the thin oxide breakdown regime is only a matter of degree.

Weak spots in the gate oxide are introduced during wafer manufacturing. They can be caused by process contaminations (causing localized oxide thinning), impurities from cleaning chemicals or dopants (e.g., mobile ions or metallics), process-induced damage (e.g., plasma or ion implantation), gate oxide thinning at the field oxide bird's beak, and high stress points due to the geometric layout.[75]

It is instructive to review the I-V curve of a capacitor as it is breaking down. Figure 19 shows typical leakage curves of three types of oxide failure as the voltage is increased, such as in a voltage ramp test. Curve *c* represents an intrinsically good gate oxide, as evidenced by the sudden increase in current beyond 10 MV/cm. This

Micron Ex. 1029, p. 709
Micron v. YMTC
IPR2024-00911

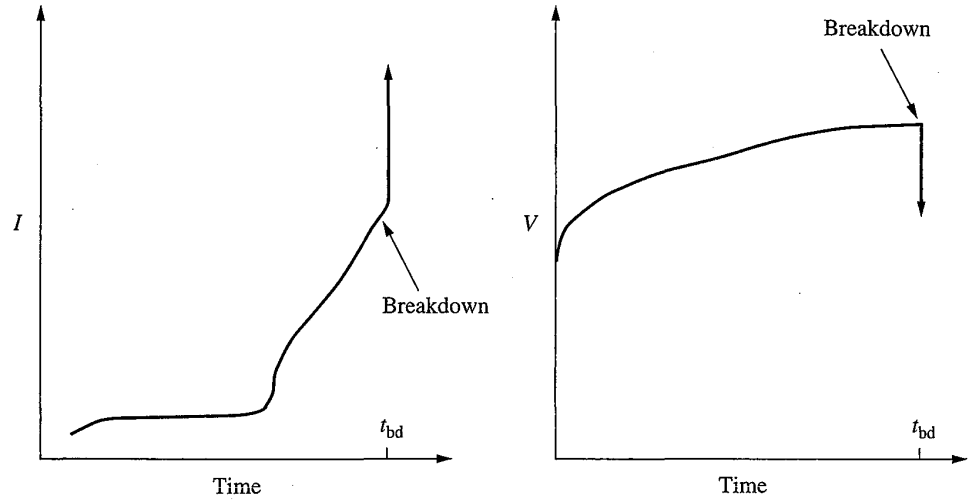


FIGURE 17
(*a*) In TDDB measurement, the leakage current i is monitored as a function of a fixed voltage. (*b*) In Q_{bd} measurement, the voltage is monitored as a fixed F-N tunneling current is applied across the oxide. The time to breakdown is denoted by t_{bd}.

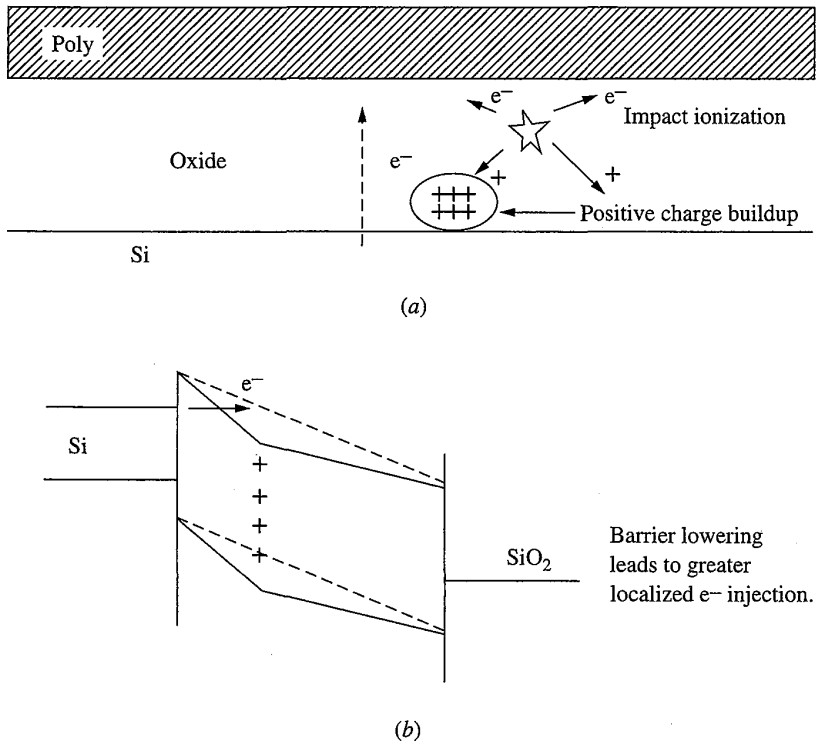


FIGURE 18
(*a*) A model for oxide breakdown mechanism. (*b*) Energy-band diagram showing a lowering of the injection barrier due to positive charge trapping in the oxide.

Micron Ex. 1029, p. 710
Micron v. YMTC
IPR2024-00911

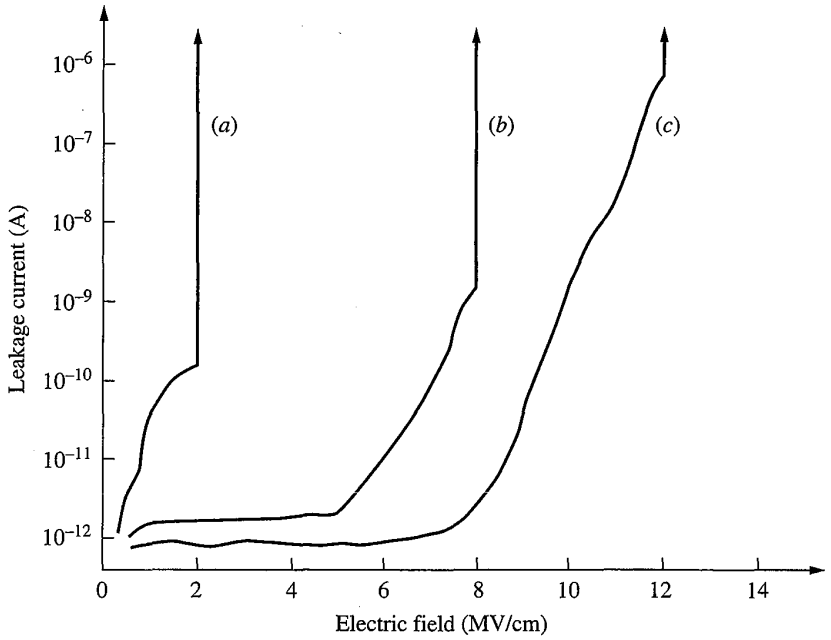


FIGURE 19
I-V curves of three types of oxide failures in a ramped voltage test. Curves (a), (b), and (c) represent Modes A, B, and C types of oxide defects. Mode A defect is usually defined for an electric breakdown value less than 2 MV/cm. Mode B is between 2 and 8 MV/cm. Mode C is greater than 8 MV/cm.

is known commonly as a Mode C type of defect (> 8 MV/cm). Curve *b* represents a Mode B defective oxide, as the rupture is about 7 MV/cm. Notice the leakage in curve *b* occurs earlier. This behavior is often exemplary of Mode B defects. This type of defect (between 2 and 8 MV/cm) poses a long-term reliability problem and such a capacitor could fail within the 10-year life of the circuit. Curve *a* represents a Mode A type of defect and is usually due to a gross oxide defect. Mode A oxide failures are usually tested out before the product is shipped to customers.

12.5.2 Lifetime Prediction of Oxide Breakdown

The lifetime prediction of oxide breakdown is based on empirical modeling of data. There are two popular models to determine the lifetime of oxides: the $\mathscr{E}$ model[77] and the $1/\mathscr{E}$ model.[78] Both appear to fit the experimental data at the high $\mathscr{E}$ fields at which laboratory experiments are obtained. Reliability prediction, however, involves extrapolation to lower electric field conditions. It is not clear at these lower fields (2.5 to 4 MV/cm) which model gives a better estimate of oxide lifetime since it has been extremely difficult to verify the long-term data.

The $\mathscr{E}$ model was obtained from analysis of TDDB failure data.[77,79] This analysis shows that the lifetime of oxide, t_1, stressed at an electric field, $\mathscr{E}_1$, is related to

Micron Ex. 1029, p. 711
Micron v. YMTC
IPR2024-00911

the lifetime of the same oxide, t_2, stressed at electric field, $\mathscr{E}_2$, by the relationship

$$t_1 = t_2 \exp\left[-\beta'(\mathscr{E}_1 - \mathscr{E}_2)\right] \tag{12.24}$$

where β' is the electric field acceleration factor. Since the experimental data are typically represented on a log base 10 scale, β is usually expressed in base 10 decades per MV/cm. Thus,

$$t_1 = t_2 10^{-\beta(\mathscr{E}_1 - \mathscr{E}_2)} \tag{12.25}$$

The relationship of oxide lifetimes at two different temperatures follows the Arrhenius relationship with an activation energy E_a of about 0.3 eV for intrinsic oxide failures. The combined acceleration AF due to voltage and temperature stresses can be calculated, by assuming β and E_a are independent of temperature and electric field, as

$$\text{AF} = \text{AF}_v \times \text{AF}_T \tag{12.26}$$

where AF_v is the voltage acceleration factor and AF_T is the temperature acceleration factor. By using Eq. (12.26), the lifetime of an oxide at any field condition can be extrapolated from any accelerated TDDB tests.

Recent research has shown that in the low cumulative failure distribution there appears to be (1) a dependence of β on temperature, (2) a dependence of E_a on electric field at high temperatures (e.g., 150°C) and at higher fields (> 6 MV/cm),[79] and (3) no dependence of E_a on electric field at lower temperatures (e.g., 60°C) and fields (< 5 MV/cm).[80] This change suggests that the physical oxide breakdown mechanism at lower fields may be different from that at higher fields. The value of β has been measured to be about 5 decades per MV/cm at 60°C and the activation energy to be 0.75 eV at 4 to 5 MV/cm.[80]

The $1/\mathscr{E}$ model is based on the observation that oxide breakdown is a product of the Fowler-Nordheim tunneling currents and the positive charge generation.[81] Time to failure is expressed as

$$t_{\text{bd}} \propto \exp(G/\mathscr{E}_{\text{ox}}) \tag{12.27}$$

where G is about 320 MV/cm and $\mathscr{E}_{\text{ox}}$ is the applied electric field across the oxide. From Eq. (12.27), it can be shown that

$$\beta = -\frac{d}{d\mathscr{E}_{\text{ox}}}[\log_{10}(t_{\text{bd}})] = \frac{G}{2.3\mathscr{E}_{\text{ox}}^2}\left(\frac{\text{decades}}{\text{MV/cm}}\right) \tag{12.28}$$

The oxide lifetime can be modeled by the effective, localized oxide thinning, ΔX_{ox}, where $X_{\text{eff}} = X_{\text{ox}} - \Delta X_{\text{ox}}$, and X_{ox} is the originally deposited oxide thickness. Thus,

$$t_{\text{bd}} = \tau_0 e^{GX_{\text{eff}}/V_{\text{ox}}} \tag{12.29}$$

where τ_o is the intrinsic breakdown time under the applied voltage V_{ox}. In the $1/\mathscr{E}$ model, the oxide lifetime can be predicted once the X_{eff} distribution is determined experimentally.

It is believed that the $\mathscr{E}$ field model is more conservative in predicting oxide lifetime than the $1/\mathscr{E}$ field model.[80]

Micron Ex. 1029, p. 712
Micron v. YMTC
IPR2024-00911

12.5.3 *V*-Ramp and *J*-Ramp Measurement Techniques

In a manufacturing environment, the TDDB test is too time-consuming to use as a process monitor. Instead, a stepped voltage test, the *V*-ramp test, is commonly used as shown in Fig. 20. This stepped voltage test produces a failure distribution that can be translated directly into that obtained by TDDB once a β is known.[82]

Another oxide monitor test used more recently is the current-ramping test, or the J^*t test.[83] In this test, the current is ramped until breakdown, and the integrated value of J^*t gives the total charge density to breakdown. This test is reported to reveal early oxide defects more distinctly than the *V*-ramp test.[83]

12.5.4 Oxide Breakdown Summary

Oxide breakdown is a complex mechanism. Both the $\mathscr{E}$ field and the $1/\mathscr{E}$ field models have been advanced to predict the oxide lifetime based on data from accelerated tests. It has been difficult to verify experimentally which one is more appropriate to project oxide lifetime at actual field stress conditions. As the oxide thickness becomes comparable to the mean free path of the tunneling electrons, it is suggested that new models are required, such as the hot hole model in Section 12.5.1. Furthermore, the activation energy of oxide defects is not well quantified and remains a challenge in estimating early defects. Therefore, it is expected that the understanding of oxide failure mechanisms in the 50 to 100 Å range will continue to be a major challenge.

12.6 EFFECT OF SCALING ON DEVICE RELIABILITY

The effect of device scaling on process reliability can be estimated from earlier discussions. Essentially, scaling leads to an increase in power density in ICs, which accelerates the wearout mechanisms described earlier. In the following discussion, the scaling factor will be defined by k, where $k > 1$. For example, a 20% shrinkage is denoted by $k = 1.20$.

12.6.1 Effect of Scaling on HCI

The HCI lifetime of a transistor is described by Eq. (12.7). The effect of scaling is to increase the substrate current as the channel length decreases. The increase can be described by rewriting Eq. (12.6) as

$$I_{\text{sub}} \propto \exp(-B/\mathscr{E}_m) \tag{12.30}$$

Then,

$$\frac{I_{\text{sub}}(\text{scaled})}{I_{\text{sub}}(\text{unscaled})} = \exp\left[-\frac{B}{\mathscr{E}_m}\left(\frac{1}{k} - 1\right)\right] \tag{12.31}$$

Micron Ex. 1029, p. 713
Micron v. YMTC
IPR2024-00911

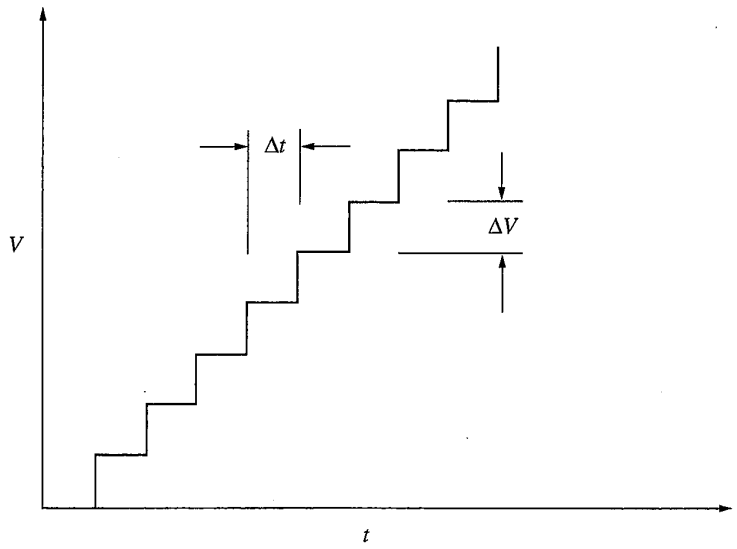


FIGURE 20
Voltage step waveform used in the V-ramp test.

where the increase in electric field is approximated by $k\mathscr{E}_m$. Therefore, the decrease in HCI lifetime is an exponential function of the scaling factor k, or

$$\Delta\tau \propto \exp\left[\frac{mB}{\mathscr{E}_m}\left(\frac{1}{k}-1\right)\right] \tag{12.32}$$

12.6.2 Effect of Scaling on EM

The basic EM lifetime of a conductor line is described by Eq. (12.18). The current density J will be increased as a result of the increased current drive of scaled devices (for a fixed power supply) and the reduced interconnect cross section. To a first-order approximation, the current in the metal line is proportional to the transistor drive current, I_D, where

$$I_D \propto \frac{\epsilon}{t_{ox}}\frac{w}{L} \tag{12.33}$$

where w is the width and L is the channel length of the transistor. In the case of a two-dimensional linear shrink, coupled with gate oxide shrink by the same shrink factor k, this expression becomes

$$I_D \propto \left(\frac{1}{1/k}\right)\left(\frac{1/k}{1/k}\right) \propto k \tag{12.34}$$

Since the current density is $J = I/A$, and $A \approx w \times L \propto k^{-2}$, it follows that

$$J \propto k^3 \tag{12.35}$$

Micron Ex. 1029, p. 714
Micron v. YMTC
IPR2024-00911

Thus, EM lifetime will vary as

$$\text{MTF} \propto k^{-6} \tag{12.36}$$

if we assume $n = 2$ in Eq. (12.18). This sixth-power reduction in lifetime portends an ominous reliability warning for narrow interconnects. However, in reality the so-called bamboo effect shown in Fig. 11 mitigates this effect somewhat.

12.6.3 Effect of Scaling on SM

The effect of scaling on stress migration is estimated from Eq. (12.23), which agrees approximately with experimental data shown in Figs. 16*a* and *b*. If the thickness of the metal remains constant as the line width decreases, SM lifetime is expected to decrease as the inverse cube of the scaling factor, or

$$\Delta\tau \propto k^{-3} \tag{12.37}$$

The prediction shows that SM worsens for narrower interconnect line widths unless alternative prevention techniques are used, e.g., metal layering as discussed earlier.

12.6.4 Effect of Scaling on Oxide Reliability

The effect of scaling on oxide reliability is to increase the electric field across the oxide due to thinner oxides. From Eq. (12.24), the reduced lifetime, τ_s, due to the thinner oxide thickness can be described as

$$\frac{\tau_s}{\tau_u} = \exp[-\beta(k-1)] \tag{12.38}$$

where τ_s is the lifetime of the scaled oxide and τ_u is the lifetime of the unscaled oxide. The effect of scaling is to reduce the oxide reliability exponentially.

12.6.5 Scaling Summary

Scaling is a natural progression of cost improvement. The requirement of putting more functions and memory bits on a single chip dictates the need to go to smaller geometries. So far, this scaling has progressed without a change in the 5-volt power supply system requirement. This has resulted in increased reliability risks toward wearout mechanisms, e.g., HCI, EM, SM, and oxide reliability. However, as system manufacturers change the power supply to 3.3 V or even lower voltages, the reliability risks will be reduced.

12.7 RELATIONS BETWEEN DC AND AC LIFETIMES

Previous discussions have been based on dc stressing methods, primarily because dc measurements have been easier to perform than ac measurements. Since transis-

Micron Ex. 1029, p. 715
Micron v. YMTC
IPR2024-00911

tors, capacitors, and interconnect elements in ICs are operated in dynamic environments, it is important to understand how the dc lifetimes are related to ac lifetimes. It should be noted that researchers have not derived these relationships for HCI, EM, and TDDB in an entirely straightforward manner because various relaxation and switching effects in the frequency domain are encountered during ac stressing.

12.7.1 AC and DC HCI

The relationship between ac and dc HCI lifetimes can be illustrated by considering a CMOS inverter and the waveforms experienced by the n-channel pull-down transistor, shown in Fig. 21. As discussed in Section 12.2, the n-channel HCI degradation is greatest at the peak of the substrate current and is negligible when the I_{sub} is small. Figure 21 shows the short-duration pulses generated at I_{sub} peaks that cause HCI degradation in the ac case. The duration of these pulses is determined by the rise and fall times in V_D and V_G, and the pulses occur in intervals determined by the frequency of the clock.

$I_{sub}(t)$ will be a function of $V_G(t)$ and $V_D(t)$. It is natural to assume that in the ac (or dynamic) case, the net ac degradation is a summation of each small interval of dc stress defined by Eq. (12.9). Then the ac lifetimes of the n-channel transistor can be expressed as

$$\int_0^{t_{ac}} \left[\frac{I_{sub}}{W}\right]^m dt = B \tag{12.39}$$

where I_{sub}, B, and m have the same meaning as before and are obtained from dc HCI measurements. The term τ_{ac} is the ac lifetime, and W is the width of the transistor. Equation (12.39) is the so-called quasi-static model. Similar expressions can be derived for the p-channel transistor. The same arguments can be applied to other transistors in the circuits with their own $I_{sub}(t)$ or $I_{gate}(t)$ values.

It is now generally believed that the quasi-static model can basically predict the ac HCI lifetime based on dc HCI lifetime data for digital logic circuits.[84] Equation (12.39) has been verified for CMOS inverters and 101-stage ring oscillators over very long time periods and at 30 MHz.[85–87] For circuit designs susceptible to HCI degradation, circuit-reliability simulators, e.g., BERT (Berkeley Reliability Tools), can be used to simulate the degradation of circuit performance based on dc HCI degradations.[88,89]

There remain some subtle dynamic HCI effects that are subjects of on-going research. These include the following:

1. Hole and electron relaxation and neutralization effects from dynamic stressing[90]
2. Mechanical stress effects on HCI[91]
3. Enhanced AC degradation in the channel hot electron (CHE) region.[92]

The HCI degradation of n-channel and p-channel devices must be interpreted in light of degradation in the corresponding circuit blocks. When an IC parameter drifts beyond its specification limit, the IC is considered to have failed. Therefore, MOSFET parameter drifts must be correlated to the IC drift. For example, in DRAM, the hold

Micron Ex. 1029, p. 716
Micron v. YMTC
IPR2024-00911

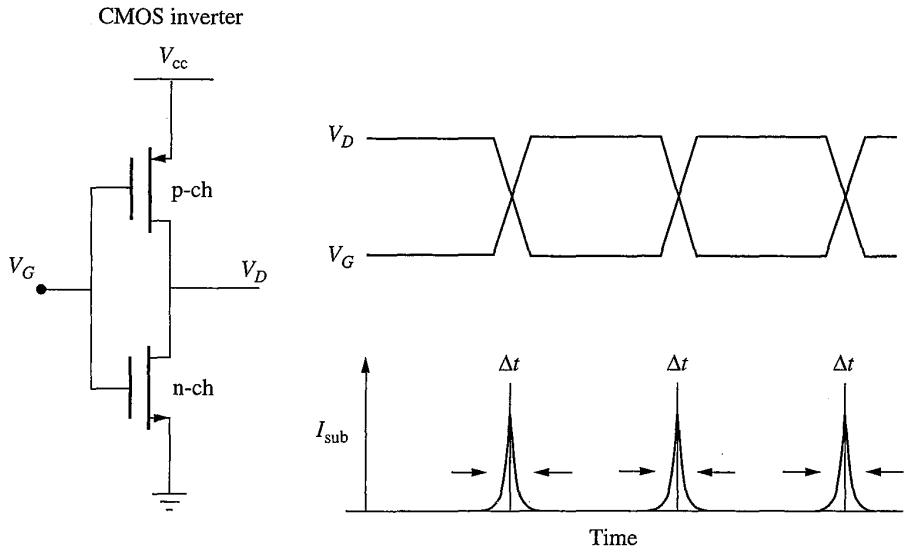


FIGURE 21
Example of an ac clock waveform in a CMOS inverter and corresponding substrate currents (I_{sub}) experienced by the n-channel pull-down transistor.

time, t_{hold}, and the minimum supply voltage of operation, $V_{DD,min}$, are typical parameters used to monitor the HCI drift. Bootstrapped nodes in DRAM are known to be vulnerable to HCI damage. Typically, a DRAM fails as a result of a sense amplifier mismatch, which is caused by slowing down of clocks, and leaky nodes, caused by HCI degradation.[1]

12.7.2 AC versus DC Electromigration

Conductor lines in ICs are stressed under either pulsed dc- or ac-current waveforms at various frequencies. The basic challenge in electromigration is to find the conductor lifetime under these stress conditions based on dc experiments. Hence, the question to address is this: Can the basic EM lifetime model based on dc experiments, described by Eq. (12.18), still adequately model the pulsed ac condition? If it can, then how should the current density J be defined in Eq. (12.18)?

To answer these questions, researchers have studied EM with square wave pulses of various duty cycles and frequencies. It is instructive to review the parameters of a general square wave pulse shown in Fig. 22. J_1 is the positive direction peak current; J_2 is the negative direction peak current; r is the duty ratio; τ is the period; and f is the frequency. Note that (1) for $r = 1$, the waveform becomes the simple dc continuous stress; (2) for $r < 1$ and $J_2 = 0$, the waveform is a unidirectional, pulsed dc (pdc) waveform; (3) for $r < 1$, $J_2 < 0$, the waveform is an ac bipolar, and (4) for $r = +0.5$ and $J_2 = -J_1$, the waveform becomes a perfectly symmetrical ac pulse.

In the case of a unidirectional dc pulse, it has been shown that at low frequencies ($f < 1$ kHz), the lifetime can be described by the on-time of the pulse.[93,94] The

Micron Ex. 1029, p. 717
Micron v. YMTC
IPR2024-00911

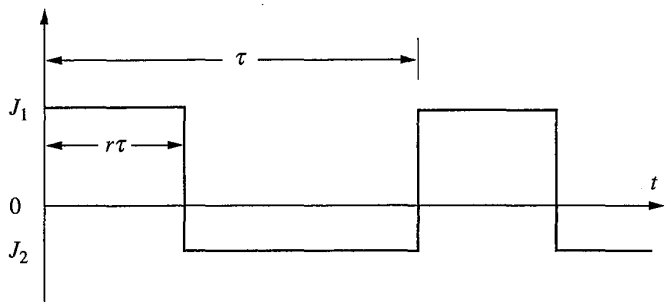


FIGURE 22
A general pulsed square waveform used in EM tests. J_1 is the positive peak current and J_2 is the negative current. The period is τ, and r is the duty cycle.

effective total time of stress is the sum of the times of each individual pulse. Since the on-time as well as the off-time of the pulse in this frequency range is longer than the thermal time constant of the Al ions, the vacancies can follow the on-off cycles of the electron wind. This phenomenon is described by

$$t_{p,\mathrm{dc}} = \frac{1}{r} t_{\mathrm{dc}} \qquad f < 1 \text{ kHz} \tag{12.40}$$

Equation (12.40) is also the prediction of the root mean square (RMS) current density model.

At higher frequencies, it is expected that the electron wind will change direction at a rate faster than the response time of the individual atoms (or vacancies), so that the effective current density experienced by the ions or vacancies should be described by the time-average current density. This theory has generally been proven up to the low MHz range.[95–97] The time-average EM model can be described by

$$\mathrm{MTF} = \frac{A}{(J_{\mathrm{avg}})^n} \exp\left(\frac{E_a}{kT}\right) \tag{12.41}$$

where J_{avg} is the time-average current density of an arbitrary waveform. For the waveform shown in Fig. 22, with $n = 2$, Eq. (12.41) becomes

$$t_{p,\mathrm{dc}} = \frac{1}{r^2} t_{\mathrm{dc}} \tag{12.42}$$

This is the familiar $1/r^2$ time-average current model.

The shortcoming of the time-average current model is that there is a singularity for perfectly symmetrical wave pulses, i.e., $J_{\mathrm{avg}} = 0$. The physical implication of this is that the EM damage created by the electron wind in one direction cancels that of the other direction. For high frequency, this model is dangerous to use in the design rule, since it allows very high $\pm J_1$ peaks, which can cause high joule heating and temperature gradients. To overcome this, the vacancy relaxation model[98] and the average current recovery model[99] have been proposed. Furthermore, there

Micron Ex. 1029, p. 718
Micron v. YMTC
IPR2024-00911

is experimental evidence that the exponent for the duty factor r may be different from $n = 2$ in a generalized wave form such as shown in Fig.22.[99,100] Therefore, the time-average current model, although adequate for unidirectional pulsed dc waveforms, should be used carefully for ac electromigration design rules.

12.7.3 Dynamic Testing of Oxides

Historically, it has been postulated that ac and dc oxide lifetimes can be related by the duty factor. By taking into account the total on-time, the lifetime of a unipolar, pulsed dc stress, $t_{bd}(\text{pdc})$, can be related to the dc lifetime, $t_{bd}(\text{dc})$, by the expression

$$t_{bd}(\text{pdc}) = \frac{1}{r} t_{bd}(\text{dc}) \tag{12.43}$$

where r is the duty factor of the on-time and $t_{bd}(\text{dc})$ is the TDDB lifetime of the oxide discussed in Section 12.5.

Research has shown that Eq. (12.43) is correct at low frequencies, e.g., $<$ 200 Hz.[101] However, at higher frequencies, e.g., $>$ 10 kHz, $t_{bd}(\text{dc})$ is shown to be about four times longer than the lifetime predicted by Eq. (12.43) for 80 to 110 Å oxides.[101] This enhancement is attributed to the reduced hole trapping at the weak spots at the cathode side of the oxide.[101,102] Recall that hole trapping is responsible for oxide breakdown (Section 12.5.2). Thus, a reduction in hole trapping can lead to a longer lifetime. The reduced hole trapping is believed to arise from hole-trap relaxation during the off-time and has been found to have a time constant of 0.3 msec.[101] The charge to breakdown, Q_{bd}, is similarly increased during unipolar dc stress.

The lifetime under bipolar stressing is enhanced even more than in the unipolar case, that is, by about 40 to 100 times over that of the dc lifetime at frequencies ranging from 10 kHz to 4 MHz.[103] This further enhancement is believed to be due to the neutralizing effect of electron injection at both electrodes with the trapped holes near both electrodes.[103]

Experiments show that the electric field acceleration factor β, in the bipolar case, is the same as the dc stress value.[103] This fact allows the determination of the ac lifetime of oxides at low fields from a minimum amount of bipolar stress data at high fields.

12.8 SOME RECENT ULSI RELIABILITY CONCERNS

Newer materials and processing conditions have been developed at the submicron level. Some of these materials and processing conditions have caused reliability problems not observed previously. Two recent reliability problems will be discussed: (1) the effect of tungsten plug on electromigration of interconnect structures and (2) the effect of plasma damage on thin gate oxide reliability.

Micron Ex. 1029, p. 719
Micron v. YMTC
IPR2024-00911

12.8.1 Tungsten Plug

The traditional approach to Al contact and via construction has led to poor Al step coverage in high-aspect-ratio contacts and vias. See Fig. 23 for an illustration. Two problems are due to poor Al step coverage: (1) Electromigration lifetime is reduced as a result of the thinned Al at the contact (via) walls, and (2) subsequent topography over the contact (via) structure is worsened. The first drawback limits the maximum current density allowed in a design. The second limits the compactness of layout rules on top of the contact (via), e.g., the use of stacked contacts. To overcome this problem of poor Al step coverage, tungsten plug processes have been developed to make the overlying topography smooth.[104] However, the interface between the W plug and the Al introduces a source of maximum Al ion flux divergence and causes shorter electromigration lifetimes than the traditional Al/Al via construction with good Al step coverage.[105] Figure 24 illustrates the problem. The upper conductor has a finite source of Al and, in the presence of a maximum flux divergence, Al atoms are quickly depleted near this W interface. The electromigration of Al can be delayed by adding 0.5 to 2% Cu alloys.[35,106] However, even Cu will electromigrate downstream,[106] so that eventually Al migration starts near the interface as the Al is depleted of Cu. It has been reported that via electromigration is the limiting factor (rather than line electromigration) in W plug interconnect technology.[106]

The challenge to researchers is to develop W plug interconnect schemes that will be reliable enough to handle the current density requirements of power-hungry logic circuits. An alternative via filling technique that uses very high Al sputtering temperatures ($> 500°C$) is also being developed. At this temperature, the Al flows into and completely fills the via. The advantage of this approach is that it removes the flux divergence problem at the W plug interface. But the reliability of these vias has not been demonstrated, and it is not clear how high an aspect ratio of vias this process can fill.

12.8.2 Plasma Damage

As gate oxide thicknesses become thinner (< 150 Å), they are more prone to plasma damage even under normal conditions of multilevel processing.[107] The symptoms of this damage range from outright oxide shorts to transistor threshold shifts, interface damage, and degraded transistor HCI performance.[108–110] These kinds of plasma damage have recently been reported in triple-level metal processing.[108,111] The cause has been attributed to the spatial nonuniformity in the plasma potential.[112,113] The nonuniform plasma causes an imbalance of ion and electron currents in the vicinity of the wafer. Currents are injected through the thin oxides as the wafer tries to balance out the inhomogeneous potential across the wafer surface.

In multilevel-metal schemes, many process steps potentially can cause plasma damage. These steps include interlayer dielectric (ILD) deposition and etching, contact and via etch, plasma resist stripping (or ashing), tungsten plug formation, aluminum plasma etch, and bond-pad etching. The behavior of plasma-damaged

Micron Ex. 1029, p. 720
Micron v. YMTC
IPR2024-00911

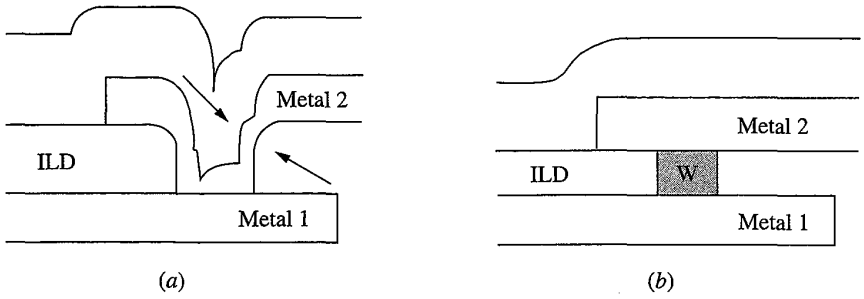


FIGURE 23
(*a*) Diagram of a traditional Al-to-Al via showing Al thinning, resulting in poor step coverage at the side wall (indicated by arrows); (*b*) W plug via showing practically no step coverage problem and good overlying metal topography.

transistors and capacitors at the end of the process can be simulated experimentally by injecting F-N currents into transistors or capacitors.[108,110] By studying the behavior of these injected structures, researchers have found that threshold shifts and degraded HCI performance are predominantly due to bulk trapping of electrons and holes. Damaged oxides cannot be annealed out totally, and the damage can reappear upon subsequent stressing of transistors. Transistors with large antenna ratios exhibit worse plasma damage, where the antenna ratio is defined as the total area of the conductor surface over the field oxide connected to a given transistor divided by the thin gate oxide area. By protecting the transistor gates with parallel diodes, the damage can be reduced.[108,111] Minimizing the plasma damage problem is expected to be a major challenge for the deep submicron technologies.

12.9 MATHEMATICS OF FAILURE DISTRIBUTION

The mathematics of failure distribution is important to process reliability because, to predict reliability from a limited number of samples, it is necessary to fit the experimental data to some failure distribution. It is the failure distribution that describes the failure characteristics of a specific failure mechanism. From the failure distribution,

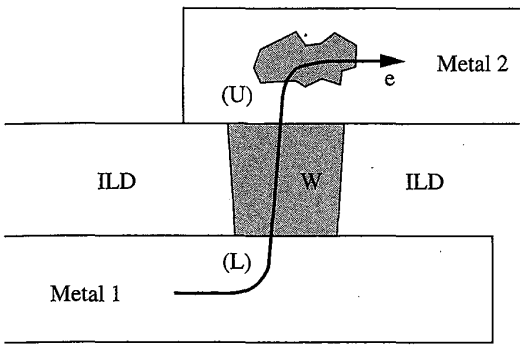


FIGURE 24
W plug electromigration. In the direction of electron flow, the upper (U) W/Al interface experiences a finite Al source. Voiding (shaded area) occurs rapidly as Al is swept downstream by the large flux divergence at the interface *u*.

Micron Ex. 1029, p. 721
Micron v. YMTC
IPR2024-00911

one can then project the failure rate. If accelerated testing is used, e.g., in temperature, voltage, current, or humidity, then appropriate acceleration models must be used to project the reliability to field conditions.

In the foregoing discussions on HCI, EM, SM, and TDDB, failure data are commonly fitted to either the Weibull or lognormal failure distributions. Generally both models fit most sets of failure data equally well. The major difference between them is in the extrapolation beyond the range of the sample data. When projecting from a measured 5% cumulative failure data to, say, 0.01% of the failure population, the projection based on the lognormal model is usually more optimistic than the projection based on the Weibull distribution fit.[114] The question then arises as to which model should be used.

The Weibull model is based on the *extreme value distribution* that applies when many small defect sites compete with each other to be the one that causes the earliest time of failure.[114] Oxide failures fit this concept reasonably well, so Weibull models are commonly used for TDDB data. On the other hand, the lognormal distribution is based on a model for degradation processes,[114] which explains why it has been very successful in modeling failures due to chemical reactions, molecular diffusion, or migration, such as EM and SM. The general equations of failure distribution and the basic equations of Weibull and lognormal models will be summarized in the following sections.

12.9.1 Failure Rate Concepts

The cumulative distribution function (CDF) of failure is related to the probability density function (PDF) of failure by the relationship[114,115]

$$F(t) = \int_{-\infty}^{t} f(t')\,dt' \tag{12.44}$$

where $F(t)$ is the CDF and $f(t)$ is the PDF. The instantaneous failure rate is defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{12.45}$$

Both the CDF, $[F(t)]$, and the PDF, $[f(t)]$, are commonly used terms in quantifying the reliability of a failure mechanism. For example, in an EM test of 20 samples, if 8 have failed after 1000 hours of stressing, then the corresponding CDF is 40% at 1000 hours. The instantaneous failure rate $\lambda(t)$ can be found by applying Eq. (12.45). The quantity $\lambda(t)$ is the failure rate of the survivors to time t in the very next instant following t. We define the units of failure rate to be 1 Failure Unit $\equiv$ 1 FIT $\equiv$ 1 failure in 10^9 device-hours.

12.9.2 Weibull Distribution

The CDF of the Weibull model is defined[114,115]

$$F(t) = 1 - \exp[-(t/c)^m] \qquad (\text{for } t > 0 \text{ only}) \tag{12.46}$$

Micron Ex. 1029, p. 722
Micron v. YMTC
IPR2024-00911

and the PDF is defined as

$$f(t) = F'(t) = \frac{m}{t}\left(\frac{t}{c}\right)^m \exp\left[-\left(\frac{t}{c}\right)^m\right] \tag{12.47}$$

where m is the shape parameter, c is the characteristic lifetime parameter, and both m and c are positive.

The instantaneous failure rate can be written as

$$\lambda(t) = \frac{m}{t}\left(\frac{t}{c}\right)^m \tag{12.48}$$

From Eq. (12.46), the data of cumulative failure rate can be analyzed by using the expression

$$\ln\left\{\ln\left[\frac{1}{1-F(t)}\right]\right\} = m\ln t - m\ln c \tag{12.49}$$

Therefore, by plotting the left-hand side of Eq. (12.49) against $\ln t$, the parameters m and c can be obtained. Standard Weibull plotting paper is available, as shown in Fig. 25; the ordinate is marked in both cumulative failure and in $\ln\ln[1/(1-F)]$, and the abscissa is in $\ln t$.

For $m = 1$, $\lambda(t)$ becomes a constant ($\lambda = 1/c$), so the exponential distribution is a special case of the Weibull distribution when $m = 1$. Note the exponential distribution is defined to have $\lambda(t) =$ constant. For $m < 1$, $\lambda(t)$ decreases with time; and the Weibull distribution may be used to represent the early failure period (e.g., extrinsic TDDB failures or early device failures). For $m > 1$, $\lambda(t)$ increases with time; and the Weibull distribution may be used to represent the wearout period of a process or device (e.g., intrinsic TDDB failures).[114,115]

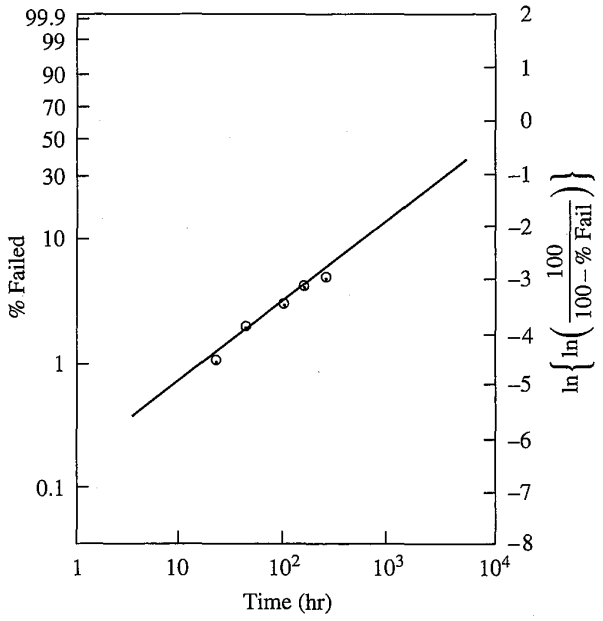


FIGURE 25
A typical Weibull plot. (*After* S.M. Sze, ed., *VLSI Technology*, 2nd ed., McGraw-Hill, 1988, reproduced with permission of McGraw-Hill, Inc.)

Micron Ex. 1029, p. 723
Micron v. YMTC
IPR2024-00911

12.9.3 Lognormal Distribution Function

The PDF for the lognormal distribution is given by[114,115]

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left\{ -\left(\frac{1}{2\sigma^2}\right)\left[\ln\left(\frac{t}{T_{50}}\right)\right]^2 \right\} \tag{12.50}$$

where the shape parameter σ is the standard deviation of the logarithmic failure times, and T_{50} is the median time parameter in which 50% of the population has failed. The CDF, $[F(t)]$, is the integral of PDF from 0 to time t as prescribed in Eq. (12.44). Therefore,

$$F(t) = \Phi\left[\frac{\ln(t/T_{50})}{\sigma}\right] \tag{12.51}$$

where $\Phi(z)$ is the standard normal CDF function found in statistics books.[116] Equation (12.51) can be rewritten as

$$\ln t_f = \ln T_{50} + \sigma\Phi^{-1}[F(t_f)] \tag{12.52}$$

showing that in the lognormal distribution, the logarithm of failure times can be plotted as a linear function of $\Phi^{-1}(F) = z$, which is the inverse transformation of the CDF of the percentage of failures. Any standard linear regression program can be used to extract the slope, σ, and the intercept, $\ln T_{50}$. Standardized lognormal plotting paper is available, as shown in Fig. 26.

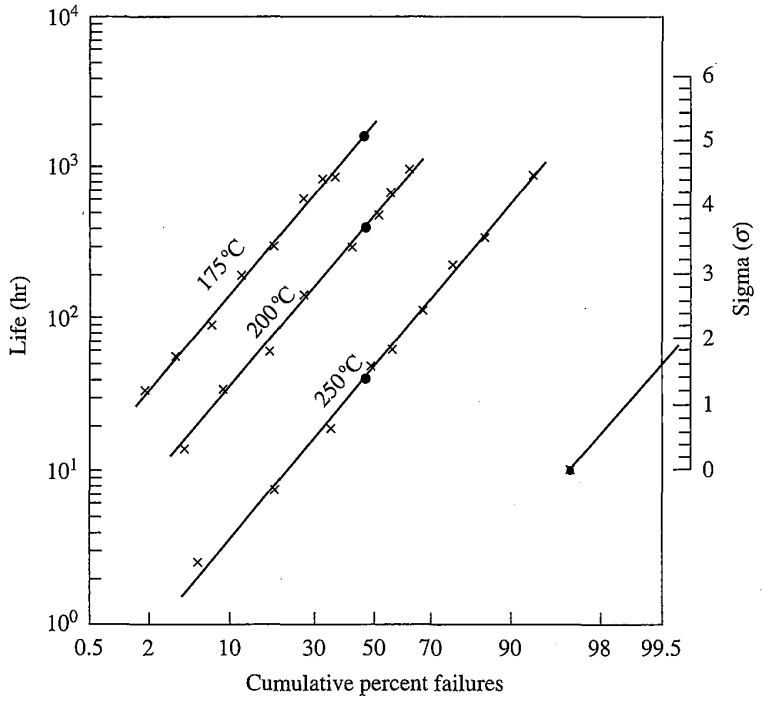


FIGURE 26
A typical lognormal plot. *(After Sze, Ref. 115.)*

Micron Ex. 1029, p. 724
Micron v. YMTC
IPR2024-00911

Thus, in any lognormal analysis of failure rate data, the key parameters to obtain from the analysis are T_{50} and σ. Once they are extracted from the experimental data, the entire lognormal distribution function is determined and any time to failure at any percentage of cumulative failure distribution, t_f (%), can be obtained.

Commonly used lognormal properties are as follows:

$$T_{50} = t\exp[-\sigma\Phi^{-1}F(t)] \tag{12.53}$$

$$\sigma = \frac{\ln(t/T_{50})}{\Phi^{-1}[F(t)]} \cong \ln\left(\frac{T_{50}}{t_{16}}\right) \tag{12.54}$$

$$t = T_{50}\exp[\sigma\Phi^{-1}F(t)] \tag{12.55}$$

Note that the lognormal distribution function is closely related to the normal distribution function. If one takes the familiar normal distribution function $f(x)$ and substitutes into it $x = \ln(t)$, then Eq. (12.50) immediately follows. The lognormal distribution is simply the normal distribution of the logarithm in time.

The lognormal instantaneous failure rate has a wide variety of appearances depending on the shape parameter σ and is similar in appearance to the Weibull failure rate [Eq. (12.48)]. For $\sigma > 2$, the failure rate rapidly falls in time and can be used to model the early failures (similar to small values of m for the Weibull). For $\sigma < 1$, the failure rate increases in time, corresponding to the wearout failure rates (similar to large values of m in the Weibull).

12.10 SUMMARY AND FUTURE TRENDS

This chapter has emphasized process reliability subjects that are critical to new process technology development. Hot carrier injection (HCI), electromigration (EM), stress migration (SM), and oxide reliability are wearout failure mechanisms, yet they can be easily influenced by defect density and new process materials. Thus, in the development of any new process technology, the performance of these basic wearout mechanisms must be carefully studied. As shown in the chapter, scaling significantly degrades reliability. The traditional concept of a 10-year dc lifetime on test structures is no longer adequate with the push toward finer dimensions, higher current density and power, and more complex, multilevel processing. To calculate ac lifetimes accurately, dynamic degradation models and some circuit-level reliability simulators have been developed. The challenge remains for reliability engineers to define how much more performance can be squeezed out of a transistor, how much more current density can be pumped through an interconnect line or contact, or how much thinner gate oxide can be for a specific application without jeopardizing the 10-year field life goal. Other reliability parameters that are layout- and design-sensitive, e.g., latch-up, electrostatic discharge (ESD), and soft error, should also be studied during the early stage of technology development. The successful development of new technologies must include the identification and characterization of all these failure mechanisms and the subsequent process optimization to ensure the overall process reliability.

Micron Ex. 1029, p. 725
Micron v. YMTC
IPR2024-00911

The task of the reliability engineer is still not completed when a technology is implemented into volume manufacturing. The engineer must still work with manufacturing engineers to maintain the same highly reliable process by ensuring that process drifts that could potentially damage reliability are monitored and corrected. Recent innovations in the IC industry in wafer level reliability (WLR)[117,118] and building in reliability (BIR)[119] are attempts to create these reliability monitoring systems. The idea in WLR is to monitor specific process reliability mechanisms inside the fab so as to detect quickly if any damaging process drifts have occurred. Or, as in BIR, the concept is to control the critical input parameters of all key process modules, at the earliest possible stage, to ensure that the process does not drift.

Finally, the element of manufacturing cost must be considered. Even after understanding the reliability of the process, difficult trade-offs are sometimes needed to reduce the cost of complex process modules at the expense of a lower, but adequate, reliability.

REFERENCES

1. A. G. Sabnis, "VLSI Reliability," in *VLSI Electronic Microstructure Science,* **22**, Academic Press, New York, 1990, Chapter 6, "Instabilities in I_c."
2. C. Hu, "Advanced MOS Device Physics," in *VLSI Electronic Microstructure Science,* **18**, Academic Press, New York, 1989, Chapter 3.
3. T. H. Ning, P. W. Cook, R. H. Dennard, C. M. Osburn, S. E. Schuster, and H. N. Yu, "1 μm MOSFET VLSI Technology, Part IV: Hot-Electron Design Constraints," *IEEE Trans. Electron Dev.,* **ED-26**, 346 (1979).
4. Y. W. Sing and B. Sudlow, "Modeling and VLSI Design Constraints of Substrate Current," *Tech. Dig. IEDM,* 732 (1980).
5. J. Y. Choi, P. K. Ko, C. Hu, and W. F. Scott, "Hot-Carrier-Induced Degradation of Metal-Oxide-Semiconductor Field-Effect Transistors: Oxide Charge versus Interface Traps," *J. Appl. Phys.,* **65**, 354 (1989).
6. S. K. Lai, "Interface Trap Generation in Silicon Dioxide When Electrons Are Captured by Trapped Holes," *J. Appl. Phys.,* **54**, 2540 (1983).
7. I. C. Chen, S. Holland, and C. Hu, "Electron-Trap Generation by Recombination of Electrons and Holes in SiO_2," *J. Appl. Phys.,* **61**, 4544 (1987).
8. E. Takeda, A. Shimizu, and T. Hagiwara, "Role of Hot-Hole Injection in Hot-Carrier Effects and the Small Degraded Channel Region in MOSFETs," *IEEE Electron Dev. Lett.,* **EDL-4**, 329 (1983).
9. C. Hu, S. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terill, "Hot-Electron Induced MOSFET Degradation—Model, Monitor and Improvement," *IEEE Trans. Electron Dev.,* **ED-32**, 375 (1985).
10. P. Heremans, J. Witters, G. Groesneken, and H. E. Maes, "Analysis of the Charge Pumping Technique and Its Application for the Evaluation of MOSFET Degradation," *IEEE Trans. Electron Dev.,* **ED-36**, 1318 (1989).
11. M. Koyanagi, A. Lewis, R. Martin, T. Huang, and J. Chen, "Investigation and Reduction of Hot-Electron Punch Through (HEIP) Effect in Submicron PMOSFETs," *Tech. Dig. IEDM,* 722 (1986).
12. T. Tsuchiya, Y. Okazaki, M. Miyake, and T. Kobayashi, "New Hot-Carrier Degradation Mode and Lifetime Prediction Method in Quarter-Micron PMOSFET," *IEEE Trans. Electron Dev.,* **ED-39**, 404 (1992).

Micron Ex. 1029, p. 726
Micron v. YMTC
IPR2024-00911

13. Q. Wang, M. Brox, W. H. Kraustschneider, and W. Weber, "Explanation and Model for the Logarithmic Time Dependence of p-MOSFET Degradation," *IEEE Electron Dev. Lett.,* **EDL-12**, 218 (1991).
14. T. C. Ong, P. K. Ko, and C. Hu, "Hot-Carrier Current Modelling and Device Degradation in Surface Channel PMOSFET," *IEEE Trans. Electron Dev.,* **ED-37**, 1658 (1990).
15. K. N. Quader, P. Fang, J. T. Yue, P. K. Ko, and C. Hu, "Simulation of CMOS Circuit Degradation Due to Hot-Carrier Effects," *Reliability Physics, 30th Annual Proceedings,* 1982, p. 16.
16. P. Fang, J. T. Yue, and D. Wollesen, "A Method to Project Hot-Carrier-Induced Punch Through Voltage Reduction for Deep Submicron LDD PMOS FETs at Room and Elevated Temperatures," *Reliability Physics, 30th Annual Proceedings,* 1982, p. 131.
17. H. Wever and W. Seith, "Neue Ergebnisse bei der Elecktrolyse Fester Metallischer Phasen," *Z. Elecktrochem.,* **59**, 942 (1955).
18. H. B. Huntington and A. R. Grone, "Current Induced Marker Motion in Gold Wires," *J. Phys. Chem. Solids,* **20**, 76 (1961).
19. I. A. Blech and H. Sello, "The Failure of Thin Aluminum Current-Carrying Strips on Oxidized Silicon," *Physics of Failures in Electronics* **5**, USAF-RADC Series, 496 (1966).
20. H. B. Huntington, "Electromigration in Metals," in A. S. Nowick and J. J. Burton, Eds., *Diffusion in Solids: Recent Development,* Academic Press, New York, 1975, p. 303.
21. J. Bass, "The Formation and Motion Energies of Vacancies in Aluminum," *Philos. Mag.,* **15**, 717 (1967).
22. J. R. Black, "Physics of Electromigration," *Reliability Physics, 12th Annual Proceedings,* 1974, p. 142.
23. T. Kwok and P. S. Ho, "Electromigration in Metallic Thin Films," in D. Grupta and P. S. Ho, Eds., *Diffusion Phenomena in Thin Films and Microelectronic Materials,* Noyes Publications, Park Ridge, NJ, 1988, p. 369.
24. K. N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science,* Macmillan, New York, 1993, Chapter 14.
25. J. R. Lloyd and P. M. Smith, "The Effect of Passivation Thickness on the Electromigration Lifetime of Al/Cu Thin Film Conductor," *J. Vac. Sci. Technol.,* **A1**, 455 (1983).
26. L. Yau, C. Hong, and D. Crook, "Passivation Material and Thickness Effects on the MTTF of Al-Si Metallization," *Reliability Physics, 23rd Annual Proceedings,* 1985, p. 115.
27. I. A. Blech, "Electromigration in Thin Aluminum Films on Titanium Nitride," *J. Appl. Phys.,* **47**, 1203 (1976).
28. J. R. Black, "A Brief Survey and Some Recent Results," *IEEE Trans. Electron Dev.,* **ED-16**, 338 (1969).
29. J. R. Black, "Electromigration Failure Modes in Aluminum Metallization for Semiconductor Devices," *Proc. IEEE,* **57**, 1587 (1969).
30. D. S. Chhabra, N. G. Ainslie, and D. Jepsen, "Theory of Failure in Thin Film Conductors," *Abstracts of the Electrochem. Soc.,* Spring Mtg. (1967).
31. J. R. Lloyd, "Electromigration Failure," *J. Appl. Phys.,* **69**, 7601 (1991).
32. M. Shatzkes, J. R. Lloyd, "A Model for Conductor Failure Considering Diffusion Concurrently with Electromigration Resulting in a Current Exponent of 2," *J. Appl. Phys.,* **59**, 3890 (1986).
33. K. Hinode, T. Furusawa, and Y. Homma, "Dependence of Electromigration Lifetime on the Square of Current Density," *Reliability Physics, 31st Annual Proceedings,* 1993, p. 319.
34. G. A. Scoggen, B. N. Agarawala, P. P. Peressini, and A. Brouillard, "Width Dependence of Electromigration Life," *Reliability Physics, 13th Annual Proceedings,* 1975, p. 151.

Micron Ex. 1029, p. 727
Micron v. YMTC
IPR2024-00911

35. H. S. Rathore, R. G. Filippi, R. A. Wachnik, J. J. Estabil, and T. Kwok, "Electromigration and Current-Carrying Implications for Aluminum-Based Metallurgy with Tungsten Stud-Via Interconnections," *Proc. SPIE,* **1805**, 251 (1992).
36. P. A. Gargini, C. Tseng, and M. H. Woods, "Elimination of Silicon Electromigration in Contacts by the Use of an Interposed Barrier Metal," *Reliability Physics, 20th Annual Proceedings,* 1982, p. 66.
37. S. Vaidya and A. K. Sinha, "Electromigration Induced Leakage at Shallow Junction Contacts Metallized with Aluminum/Poly Silicon," *Reliability Physics, 20th Annual Proceedings,* 1982, p. 50.
38. N. D. Bui, V. H. Pham, J. T. Yue, and D. L. Wollesen, "A Via Failure Mode in Electromigration of Multilevel Interconnect," *Proceedings of the 7th International VLSI Multilevel Interconnection Conference,* 1990, p. 142.
39. T. Yamaha, M. Naitou, and T. Hotta, "Three Kinds of Via Electromigration Failure Modes in Multilevel Interconnections," *Reliability Physics, 30th Annual Proceedings,* 1992, p. 349.
40. I. Ames, F. M. d'Heurle, and R. Horstman, "Reduction of Electromigration in Aluminum Films by Copper Doping," *IBM J. Res. Devel.,* **4**, 461 (1970).
41. A. Gangulee and F. M. d'Heurle, "Effect of Alloy Additions on Electromigration Failures in Thin Aluminum Films," *Appl. Phys. Lett.,* **19**, 75 (1971).
42. A. J. Learn, "Electromigration Effects in Aluminum Alloy Metallization," *J. Electron. Matls.,* **3**, 531 (1974).
43. M. Hansen, in M. Hansen and K. Anderko, Eds., *Constitution of Binary Alloys,* McGraw-Hill, New York, 1958.
44. F. M. d'Heurle and I. Ames, "Electromigration in Single-Crystal Aluminum Film Conductors," *Appl. Phys. Lett.,* **16**, 80 (1970).
45. C. H. Herzig and W. Wiemann, "Experimental Determination of the Electrostatic Driving Force in the Electromigration of Tin in Gold and Gold-Tin Alloys," *Phys. Status Solidi,* **A26**, 459 (1974).
46. J. M. Pierce and M. E. Thomas, "Electromigration in Aluminum Conductors Which Are Chains of Single Crystals," *Appl. Phys. Lett.,* **39**, 165 (1981).
47. C. V. Thompson and H. Kahn, "Effects of Microstructure on Interconnect and Via Reliability: Multimodal Failure Statistics," *J. Electron. Matls.,* **22**, 581 (1993).
48. C. K. Hu, N. J. Mazzeo, and C. Stanis, "Electromigration in Two-Level Bamboo Grain Structure Al(Cu)/W Interconnections," *Mater. Chem. Phys.,* **35**, 95 (1993).
49. B. Grabe and H. U. Schreiber, "Lifetime and Drift Velocity Analysis for Electromigration in Sputtered Al Films, Multilayers and Alloys," *Solid State Electron.,* **26**, 1026 (1983).
50. E. Levine and J. Kitcher, "Electromigration Induced Damage and Structure Change in Cr-Al/Cu and Al/Cu Interconnection Lines," *Reliability Physics, 22nd Annual Proceedings,* 1984, p. 242.
51. L. J. Fried, J. Haves, J. S. Lechaton, J. S. Logan, G. Paal, and P. A. Totta, "A VLSI Bipolar Metallization Design with Three-Level Wiring and Area Array Solder Connections," *IBM J. Res. Devel.,* **26**, 362 (1982).
52. J. K. Howard, J. F. White, and P. S. Ho, "Intermetallic Compounds of Al and Transition Metals: Effect of Electromigration in 1–2 μm Wide Lines," *J. Appl. Phys.,* **49**, 4083 (1978).
53. T. Kwok, P. S. Ho, and H. C. W. Huang, "Summary Abstract: Electromigration Studies of Al-Intermetallic Structures," *J. Vac. Sci. Technol.* A, vol 2, 241 (1984).
54. J. T. Yue, W. P. Funsten, and R. V. Taylor, "Stress Induced Voids in Aluminum Interconnects during IC Processing," *Reliability Physics, 23rd Annual Proceedings,* 1985, p. 126.

Micron Ex. 1029, p. 728
Micron v. YMTC
IPR2024-00911

55. J. Curry, G. Fitzgibbon, Y. Guan, R. Muollo, G. Nelson, and A. Thomas, "New Failure Mechanisms in Sputtered Aluminum-Silicon Films," *Reliability Physics, 22nd Annual Proceedings,* 1984, p. 6.
56. P. A. Totta, "Stress Induced Phenomena in Metallization: U.S. Perspective," in C. Y. Li, P. Totta, and P. Ho, Eds., *Stress Induced Phenomena in Metallization,* American Vacuum Society Series, **13** (AIP Conf. Proceedings No. 263), American Institute of Physics, Ithaca, NY, 1991, p. 1.
57. J. Klema, R. Pyle, and E. Domangue, "Reliability Implication of Nitrogen Contamination during Deposition of Sputtered Aluminum/Silicon Metal Films," *Reliability Physics, 22nd Annual Proceedings,* 1984, p. 1.
58. T. Turner and K. Wendel, "The Influence of Stress on Aluminum Conductor Life," *Reliability Physics, 23rd Annual Proceedings,* 1985, p. 142.
59. M. R. Lin and J. T. Yue, "Impact of Ceramic Packaging Anneal on the Reliability of Al Interconnects," *Reliability Physics, 24th Annual Proceedings,* 1985, p. 164.
60. S. Mayumi, T. Umemoto, M. Shishino, H. Nanatsue, S. Ueda, and M. Inoue, "The Effect of Cu Addition to the Al-Si Interconnects on Stress Induced Open-Circuit Failures," *Reliability Physics, 25th Annual Proceedings,* 1987, p. 15.
61. K. Hinode, N. Wada, T. Nishida, and K. Mukai, "Stress-Induced Grain Boundary Fractures in Al-Si Interconnects," *J. Vac. Sci. Technol.,* **85**, 518 (1987).
62. A. Tezaki, T. Mineta, and H. Egawa, "Measurement of Three Dimensional Stress and Modeling of Stress Induced Migration Failure in Aluminum Interconnects," *Reliability Physics, 28th Annual Proceedings,* 1990, p. 221.
63. K. N. Tu, J. W. Mayer, and L. C. Feldman, *Electronic Thin Film Science,* Macmillan, New York, 1993, Chapter 4.
64. P. A. Flinn, "Study by Wafer Curvature and X-Ray Diffraction Techniques," in *Stress Induced Phenomena in Metallization,* American Vacuum Society Series, **13**, 73 (1991).
65. J. W. McPherson and C. F. Dunn, "A Model for Stress-Induced Metal Notching and Voiding in Very Large-Scale-Integrated Al-Si (1%) Metallization," *J. Vac. Sci. Technol.,* **B5**, 1321 (1987).
66. I. Le May, *Principles of Mechanical Metallurgy,* Elsevier, New York, 1981.
67. P. A. Flinn and C. Chiang, "X-Ray Diffraction Determination of the Effects of Various Passivations on Stress in Metal Films and Patterned Lines," *J. Appl. Phys.,* **67**, 2927 (1990).
68. W. T. Tseng, "Interface Reaction Model for Process Voiding in Aluminum Conductor Lines," *Appl. Phys. Lett.,* **59**, 680 (1991).
69. Y. Sugano, S. Minegishi, H. Sumi, and M. Itabashi, "In Situ Observation and Formation Mechanism of Aluminum Voiding," *Reliability Physics, 26th Annual Proceedings,* 1988, p. 34.
70. H. Yagi, H. Niwa, T. Hosada, M. Inoue, H. Tsuchikawa, and M. Kato, "Analytical Calculation and Direct Measurement of Stress in an Aluminum Interconnect of Very Large Scale Integration," in C. Y. Li, P. Totta, and P. Ho, Eds. *Stress Induced Phenomena in Metallization,* American Vacuum Society Series, **13**, (AIP Conf. Proceedings No. 263), American Institute of Physics, Ithaca, NY, 1991, p. 44.
71. M. Kato, H. Niwa, H. Yagi, and H. Tsuchikawa, "Diffusional Relaxation and Void Growth in an Aluminum Interconnect of Very Large Scale Integration," *J. Appl. Phys.,* **68**, 334 (1990).
72. H. Keneko, M. Hasunuma, A. Sawabe, T. Kawanoue, Y. Kohanawa, S. Komatsu, and M. Miyauchi, "A Newly Developed Model for Stress Induced Slit-Like Voiding," *Reliability Physics, 28th Annual Proceedings,* 1990, p. 194.
73. I. C. Chen, S. Holland, and C. Hu, "Electrical Breakdown in Thin Gate and Tunneling Oxides," *IEEE Trans. Electron. Dev.,* **32**, 413 (1985).

Micron Ex. 1029, p. 729
Micron v. YMTC
IPR2024-00911

74. D. R. Wolters and J. T. Van der Schoot, "Dielectric Breakdown in MOS Devices Part III: The Damage Leading to Breakdown," *Philips J. Res.*, **40**, 164 (1985).
75. A. G. Sabnis, "VLSI Reliability," in *VLSI Electronic Microstructure Science,* **22**, Academic Press, New York, 1990, Chapter 5.
76. K. Schuegraf and C. Hu, "Hole Injection Oxide Breakdown Model for Very Low Voltage Lifetime Extrapolations," *Reliability Physics, 31st Annual Proceedings,* 1993, p. 7.
77. D. Crook, "Method of Determining Reliability Screens for Time Dependent Dielectric Breakdown," *Reliability Physics, 17th Annual Proceedings,* 1979, p. 1.
78. I. C. Chen, S. Holland, and C. Hu, "A Quantitative Physical Model for Time-Dependent Breakdown in SiO_2," *Reliability Physics, 23rd Annual Proceedings,* 1985, p. 24.
79. J. W. McPherson and D. A. Baglee, "Acceleration Factors for Thin Gate Oxide Stressing," *Reliability Physics, 23rd Annual Proceedings,* 1985, p. 1.
80. K. C. Boyko and D. L. Gerlach, "Time Dependent Dielectric Breakdown of 210 Å Oxides," *Reliability Physics, 27th Annual Proceedings,* 1989, p. 1.
81. J. C. Lee, I. C. Chen, and C. Hu, "Modeling and Characterization of Gate Oxide Reliability," *IEEE Trans. Electron Dev.,* **35**, 2268 (1988).
82. A. Berman, "Time-Zero Dielectric Reliability Test by a Reamp Method," *Reliability Physics, 19th Annual Proceedings,* 1981, p. 204.
83. D. Crook, "Detecting Oxide Quality Problems Using JT Testing," *Reliability Physics, 29th Annual Proceedings,* 1991, p. 337.
84. W. Weber, M. Brox, T. Kunemune, H. Muhlhoff, and D. Schmitt-Landsiedel, "Dynamic Degradation in MOSFET's—Part II: Application in Circuit Environment," *IEEE Trans. Electron. Dev.,* **ED-38**, 1859 (1991).
85. P. M. Lee, P. K. Ko, and C. Hu, "Relating CMOS Inverter Lifetime to DC Hot-Carrier Lifetime of NMOSFET's," *IEEE Electron. Dev. Lett.,* **EDL-11**, 39 (1990).
86. K. N. Quader, P. Fang, J. T. Yue, P. K. Ko, and C. Hu, "Simulation of CMOS Circuit Degradation Due to Hot-Carrier Effects," *Reliability Physics, 30th Annual Proceedings,* 1982, p. 16.
87. P. Fang, R. Rakkhit, and J. T. Yue, "An Investigation of Hot Carrier Effects in Submicron CMOS Integrated Circuits," *Microelectron. Reliab.,* **33**, 1713 (1993).
88. "BERT-Berkeley Reliability Tools," University of California-Berkeley Memorandum No.UCB/ERL M91/107 (1991).
89. K. N. Quader, P. Fang, J. T. Yue, P. K. Ko, and C. Hu, "Hot-Carrier Design Rules for Translating Device Degradation to CMOS Digital Circuit Degradation," *IEEE Trans. Electron. Dev.,* **41**, 681 (1994).
90. M. Brox and W. Weber, "Dynamic Degradation in MOSFET's—Part I: The Physical Effects," *IEEE Trans. Electron. Dev.,* **ED-38**, 1852 (1991).
91. A. Hamada and E. Takada, "AC Hot-Carrier Effect under Mechanical Stress," *1992 Symposium on VLSI Technology—Digest of Technical Papers,* 1992, p. 19.
92. K. Mistry and B. Doyle, "The Role of Electron Trap Creation in Enhanced Hot-Carrier Degradation during AC Stress," *IEEE Electron. Dev. Lett.,* **EDL-11**, 267 (1990).
93. F. M. d'Heurle, "Electromigration and Failure in Electronics: An Introduction," *Proc. IEEE,* **59**, 1409 (1971).
94. J. S. Suehle and H. A. Schafft, "The Electromigration Damage Response Time and Implications for DC and Pulsed Characterizations," *Reliability Physics, 27th Annual Proceedings,* 1989, p. 229.
95. J. A. Maiz, "Characterization of Electromigration Under Bidirectional (BC) and Pulsed Unidirectional (PDC) Currents," *Reliability Physics, 27th Annual Proceedings,* 1989, p. 220.
96. J. M. Towner and E. P. van der Van, "Aluminum Electromigration under Pulsed DC Conditions," *Reliability Physics, 21st Annual Proceedings,* 1983, p. 36.

Micron Ex. 1029, p. 730
Micron v. YMTC
IPR2024-00911

97. J. J. Clement, "Vacancy Supersaturation Model for Electromigration Failure under DC and Pulsed DC Stress," *J. Appl. Phys.*, **71**, 4264 (1992).
98. B. K. Liew, N. W. Cheung, and C. Hu, "Projecting Interconnect Electromigration Lifetime for Arbitrary Current Waveforms," *IEEE Trans. Electron. Dev.*, **37**, 1343 (1990).
99. L. M. Ting, J. S. May, W. R. Hunter, and J. W. McPherson, "AC Electromigration Characterization and Modeling of Multilayered Interconnects," *Reliability Physics, 31st Annual Proceedings,* 1993, p. 311.
100. K. Hatanaka, T. Noguchi, and K. Maneguchi, "A Generalized Lifetime Model for Electromigration under Pulsed DC/AC Stress Conditions," *Proc. 9th Symposium on VLSI Technology—Digest of Technical Papers,* 1989, p. 19.
101. Y. Fong, I. C. Chen, S. Holland, J. Lee, and C. Yu, "Dynamic Stressing of Thin Oxides," *Tech. Dig. IEDM,* 664 (1986).
102. M. S. Liang, S. Haddad, W. Cox, and S. Cagnina, "Degradation of Very Thin Gate Oxide MOS Devices Under Dynamic High Field/Current Stress," *Tech. Dig. IEDM,* 394 (1986).
103. E. Rosenbaum and C. Hu, "High-Frequency Time-Dependent Breakdown of SiO_2," *IEEE Electron. Dev. Lett.*, **12**, 267 (1991).
104. Workshops on "Tungsten and Other Refractory Metals for ULSI Applications," *Materials Research Society Conference Proceedings,* 1986–1990.
105. J. Tao, K. K. Young, N. W. Cheung, and C. Hu, "Comparison of Electromigration Reliability of Tungsten and Aluminum Vias under DC and Time-Varying Current Stressing," *Reliability Physics, 30th Annual Proceedings,* 1992, p. 338.
106. J. J. Estabil, H. S. Rathore, and E. N. Levine, "Electromigration Improvements with Titanium Underlay and Overlay in Al(Cu) Metallurgy," *Proceedings of the 8th International VLSI Multilevel Interconnection Conference,* 1991, p. 242.
107. F. Shone, K. Wu, J. Shaw, E. Hokelet, S. Mittal, and A. Haranahalli, "Gate Oxide Charging and Its Elimination for Metal Antenna Capacitor and Transistor in VLSI CMOS Double Layer Metal Technology," *Proc. 9th Symposium on VLSI Technology—Digest of Technical Papers,* 1989, p. 73.
108. Y. H. Lee, L. Yau, R. Chau, E. Hansen, B. Sabi, S. Hui, P. Moon, and G. Vandentop, "Correlation of Plasma Process Induced Charging with Fowler-Nordheim Stress in p- and n-Channel Transistors," *Tech. Dig. IEDM,* 65 (1992).
109. H. Shin, C. C. King, T. Horiuchi, and C. Hu, "Thin Oxide Charging Current during Plasma Etching of Aluminum," *IEEE Electron. Dev. Letts.*, **13**, 404 (1991).
110. H. Shin, C. C. King, and C. Hu, "Thin Oxide Damage by Plasma Etching and Ashing Processess," *Reliability Physics, 30th Annual Proceedings,* 1992, p. 37.
111. R. Rakkhit, F. P. Heiler, P. Fang, and C. Sander, "Process Induced Oxide Damage and Its Implications to Device Reliability of Submicron Transistors," *Reliability Physics, 31st Annual Proceedings,* 1993, p. 293.
112. S. Fang and J. McVittie, "Thin Oxide Damage from Gate Charging during Plasma Processing," *IEEE Electron. Dev. Letts.*, **13**, 288 (1992).
113. S. Fang and J. McVittie, "A Mechanism for Gate Oxide Damage in Nonuniform Plasma," *Reliability Physics, 31st Annual Proceedings,* 1993, p. 13.
114. P. A. Tobias and D. Trindade, *Applied Reliability,* Van Nostrand Reinhold, New York, 1986.
115. S. M. Sze, ed., *VLSI Technology,* 2nd ed., McGraw-Hill, 1988.
116. P. A. Tobias and D. Trindade, *Applied Reliability,* Van Nostrand Reinhold, New York, 1986, Chapter 5, Table 5.1.
117. T. A. Dillon, W. A. Miller, D. G. Pierce, and E. S. Snyder, "Wafer Level Reliability," *Microelectronic Manufacturing and Reliability, SPIE,* **1802**, 144 (1992).

Micron Ex. 1029, p. 731
Micron v. YMTC
IPR2024-00911

118. JEDEC task group on Wafer Level Reliability, JEDEC Committee JC14.2, *JEDEC Standard Docs.* **33** (1993) and *JEDEC Standard Docs.* **35** (1992).
119. H. A. Schafft, D. A. Baglee, and P. E. Kennedy, "Building-In Reliability: Making It Work," *Reliability Physics, 29th Annual Proceedings,* 1991, p. 1.

PROBLEMS

1. **Hot Carrier Injection—Part I:** A group of NMOS transistors is stressed under HCI tests. The transistors are stressed under 3 V_{ds} values (V_{ds} = 7 V, 6.6 V, and 6.2 V) at the maximum I_{sub} conditions. The average percentages of degradation in time of the three stressed groups of transistors are given below:

$\Delta I_{dsat}/I_{dsat}$	1.4 hr	4 hr	10 hr	40 hr
6.2 v	0.43%	0.5%	0.68%	1.0%
6.6 v	0.62%	0.8%	1.0%	1.4%
7.0 v	1.15%	1.3%	1.9%	2.8%

Find the lifetime (in hours) to reach 10 percent I_{dsat} degradation for each V_{ds} stress group. Use Eq. (12.8). (*Hint*: Use the power law regression analysis, $\Delta p/p = At^b$).

2. **Hot Carrier Injection—Part II:** In Problem 1, the I_{sub} values for each of the V_{ds} groups were 1, 2.4, and 6 μA/μm for the V_{ds} values of 6.2, 6.6, and 7 V, respectively. (*a*) Using Eq. (12.9), project for the HCI lifetime of this group of transistors at V_{ds} = 5.5 V, if the I_{sub} value is 0.2 μA/μm at that bias. (*Hint*: Use power law regression analysis, $\tau = AI_{sub}^m$). (*b*) What is the slope *m*? (*c*) Is this lifetime acceptable from a customer point of view?

3. **EM—Part I:** In an electromigration experiment, 13 samples were stressed to failure under the condition of 5×10^6 A/cm^2 at an oven temperature of 175°C. The failure times were found at times = 1, 1.2, 1.8, 2, 2.05, 2.2, 2.3, 2.3, 2.8, 3, 3.2, 4.7, and 5.2 hr. Use the lognormal statistics to estimate the T_{50} and σ of this experiment. [*Hint*: Use Eq. (12.54) and a lognormal graphing plot such as Fig. 12.26; assume the percent cumulative failure rate = $(i - 0.3)/(N + 0.4)$, where i is the ith failure in a total population of N parts.]

4. **EM—Part II:** Typical electromigration studies of a new metal interconnect system involve the measurement of kinetic parameters n and E_a. The study is often performed with a "three-cell" stress experiment. Suppose one of the cells was at the condition stipulated in Problem 3, with the T_{50} solution of Problem 3; the other two conditions and T_{50} results are given below:

	225°C	175°C
4.1×10^6 A/cm^2	9.6 hr	
5×10^6 A/cm^2	2.6 hr	T_{50} (Prob. 3)

Using the fact that the effective temperatures (T_{eff}) of the three cells were 325, 295, and 259°C as a result of joule heating, find the best set of n and E_a that fits the experimental

Micron Ex. 1029, p. 732
Micron v. YMTC
IPR2024-00911

data. (*Hint*: solve for E_a first, then use that E_a to normalize the data to the same temperature before solving for n.)

5. Use the T_{50} found in Problem 3 to find the lifetime at 0.1% of the lognormal failure distribution at a field condition of 100°C junction temperature and a current density of 1×10^6 A/cm^2. [*Hint*: The T_{eff} is 259°C. Assume $T(0.1\%) = 0.177\ T(50\%)$.]

6. **Oxide Reliability:** In a TDDB experiment of 100 Å oxide, the time required to get 1% of the capacitors to fail at 9.5 MV/cm of electric field stress was 3.15×10^3 sec. The time to get 1% of another similar set of capacitors to fail at 8.5 MV/cm was 3.15×10^5 sec. Suppose it is desired to analyze these data with the $\mathscr{E}$ field model [Eq. (12.25)]. (*a*) Find the field acceleration factor beta. (*b*) Find the 1% lifetime of the 100 Å oxide at 5.5 volts.

7. Suppose it is desired to analyze the TDDB data in Problem 6 with the $1/\mathscr{E}$ model. (*a*) Find the 1% lifetime of the 100 Å oxide using Eq. (12.27). (*b*) Which model predicts a shorter lifetime, the $\mathscr{E}$ or $1/\mathscr{E}$ model?

8. Often in lognormal analysis it is useful to know the relationship between the 0.1% cumulative failure rate ($T_{0.1}$) and the median failure rate (T_{50}). In Problem 5, it was assumed that $T_{0.1} = 0.177\ T_{50}$. In general, $T_{0.1}$ is related to T_{50} through the σ of the lognormal distribution. Use Eq. (12.55) and the property of inverse function of Standard Normal CDF to show the relationship of $T_{0.1}$ and T_{50} as a function of σ.

Micron Ex. 1029, p. 733
Micron v. YMTC
IPR2024-00911

APPENDIX A

Properties of Si at 300 K

Properties	Si
Atoms/cm^3	5.0×10^{22}
Atomic weight	28.09
Breakdown field (V/cm)	$\sim 3 \times 10^5$
Crystal structure	Diamond
Density (g/cm^3)	2.33
Dielectric constant	11.9
Effective density of states in conduction band, N_C (cm^{-3})	2.8×10^{19}
Effective density of states in valence band, N_V (cm^{-3})	1.04×10^{19}
Effective mass, m^*/m_0:	
Electrons	$m_l^* = 0.91$, $m_t^* = 0.19$
Holes	$m_{lh}^* = 0.15$, $m_{hh}^* = 0.53$
Electron affinity (V)	4.05
Energy gap (eV)	1.12
Heat capacity (J/mol-K)	20.07
Index of refraction	3.42
Intrinsic carrier concentration (cm^{-3})	1.02×10^{10}
Intrinsic Debye length (μm)	24
Intrinsic resistivity (Ω-cm)	3.16×10^5
Lattice constant (Å)	5.435
Linear coefficient of thermal expansion, $\Delta L/L\Delta T$ (°C^{-1})	2.59×10^{-6}
Melting point (°C)	1412
Minority-carrier lifetime (s)	2.5×10^{-3}

Micron Ex. 1029, p. 734
Micron v. YMTC
IPR2024-00911

Properties	Si
Mobility (drift) (cm^2/V-s):	
μ_n (electrons)	1450
μ_p (holes)	505
Optical phonon energy (eV)	0.063
Phonon mean free path λ_0 (Å):	
Electrons	76
Holes	55
Poisson's ratio, (100) orientation	0.28
Specific heat (J/g-°C)	0.7
Thermal conductivity (W/cm-°C)	1.5
Thermal diffusivity (cm^2/s)	0.9
Vapor pressure (Pa)	1 at 1650°C, 10^{-6} at 900°C
Young's modulus (GPa)	130

Micron Ex. 1029, p. 735
Micron v. YMTC
IPR2024-00911

APPENDIX B

List of Symbols

Symbol	Description	Units
a	Lattice constant	Å
B	Magnetic induction	Wb/m^2
c	Speed of light in a vacuum	cm/s
C	Capacitance	F
D	Electric displacement	C/cm^2
D	Diffusion coefficient	cm^2/s
E	Energy	eV
E_f	Fermi energy level	eV
E_g	Energy bandgap	eV
$\mathscr{E}$	Electric field	V/cm
$\mathscr{E}_m$	Maximum field	V/cm
f	Frequency	Hz
h	Planck's constant	J-s
$h\nu$	Photon energy	eV
I	Current	A
J	Current density	A/cm^2
k	Boltzmann constant	J/K
kT	Thermal energy	eV
L	Length	cm or μm
m_0	Electron rest mass	kg
m^*	Effective mass	kg
n	Refractive index	
n	Density of free electrons	cm^{-3}
n_i	Intrinsic density	cm^{-3}
N	Doping concentration	cm^{-3}
N_A	Acceptor impurity density	cm^{-3}

Micron Ex. 1029, p. 736
Micron v. YMTC
IPR2024-00911

Symbol	Description	Units
N_D	Donor impurity density	cm^{-3}
p	Density of free holes	cm^{-3}
P	Pressure	N/m^2
q	Magnitude of electronic charge	C
Q_{it}	Interface-trap density	charges/cm^2
R	Resistance	Ω
t	Time	s
T	Absolute temperature	K
v	Carrier velocity	cm/s
v_s	Saturation velocity	cm/s
V	Voltage	V
V_{bi}	Built-in potential	V
V_B	Breakdown voltage	V
W	Thickness	cm or μm
x	x direction	
∇T	Temperature gradient	K/cm
ϵ_0	Permittivity in vacuum	F/cm
ϵ_s	Semiconductor permittivity	F/cm
ϵ_i	Insulator permittivity	F/cm
ϵ_s/ϵ_0 or ϵ_i/ϵ_0	Dielectric constant	
τ	Lifetime or decay time	s
θ	Angle	rad
λ	Wavelength	μm or Å
ν	Frequency of light	Hz
μ_0	Permeability in vacuum	H/cm
μ_n	Electron mobility	cm^2/V-s
μ_p	Hole mobility	cm^2/V-s
ω	Angular frequency ($2\pi f$ or $2\pi\nu$)	rad/s
Ω	Ohm	Ω

Micron Ex. 1029, p. 737
Micron v. YMTC
IPR2024-00911

APPENDIX C

International System of Units

Quantity	Unit	Abbreviation	Units
Length	meter	m	
Mass	kilogram	kg	
Time	second	s	
Temperature	kelvin	K	
Current	ampere	A	
Frequency	hertz	Hz	1/s
Force	newton	N	$kg\text{-}m/s^2$
Pressure	pascal	Pa	N/m^2
Energy	joule	J	N-m
Power	watt	W	J/s
Electric charge	coulomb	C	A-s
Potential	volt	V	J/C
Conductance	siemens	S	A/V
Resistance	ohm	Ω	V/A
Capacitance	farad	F	C/V
Magnetic flux	weber	Wb	V-s
Magnetic induction	tesla	T	Wb/m^2
Inductance	henry	H	Wb/A

Micron Ex. 1029, p. 738
Micron v. YMTC
IPR2024-00911

APPENDIX D

Physical Constants

Quantity	Symbol	Value
Ångstrom unit	Å	$1\ Å = 10^{-1}$ nm $= 10^{-4}$ μm $= 10^{-8}$ cm $= 10^{-10}$ m
Avogadro constant	N_{Avo}	6.02204×10^{23} mol^{-1}
Bohr radius	a_B	0.52917 Å
Boltzmann constant	k	1.38066×10^{-23} J/K (R/N_{Avo})
Elementary charge	q	1.60218×10^{-19} C
Electron rest mass	m_0	0.91095×10^{-30} kg
Electron volt	eV	$1\ eV = 1.60218 \times 10^{-19}$ J $= 23.053$ kcal/mol
Gas constant	R	1.98719 cal mol^{-1}K^{-1}
Permeability in vacuum	μ_0	1.25663×10^{-8} H/cm(4×10^{-9})
Permittivity in vacuum	ϵ_0	8.85418×10^{-14} F/cm($1/\mu_0 c^2$)
Planck constant	h	6.62617×10^{-34} J-s
Reduced Planck constant	$\hbar$	1.05458×10^{-34} J-s ($h/2\pi$)
Proton rest mass	M_p	1.67264×10^{-27} kg
Speed of light in vacuum	c	2.99792×10^{10} cm/s
Standard atmosphere		1.01325×10^5 N/m^2
Thermal voltage at 300 K	kT/q	0.0259 V
Wavelength of 1-eV quantum	λ	1.23977 μm

Micron Ex. 1029, p. 739
Micron v. YMTC
IPR2024-00911

INDEX

Absorption edge, 312–313
Acoustical noise, 21
Activation, 187–192
Adatoms, 117
Air cleanliness, 5–8
Air quality, 21
Airy disc, 276
Alignment, 288–290
Alloy sputtering, 383
Aluminum etching, 359–361
Aluminum plug, 410–412
American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE), 21
Angle-lap and stain, 136
Anisotropy, 338–341
Annealing, 153–155, 229
  defect, 177–179
  dopant activation, 187–192
  rapid thermal, 176–192
  RTA of ion-implanted junctions, 154, 179–187
AR/H_2 plasma cleaning, 94, 98–99
Area defects, 121
As-built condition, 21
Aspect ratio-dependent etching (ARDE), 342, 357
Assembly:
  challenges in, 530
  chip bonding, 541–547
  chip interconnection, 541
  wafer preparation, 540–541
  wire bonding, 547–552
Atmospheric pressure chemical vapor deposition (APCVD), 126, 128, 164, 206, 215–225
Atmospheric pressure CVD reactors, 209, 212
Atomic force microscope (AFM), 74, 226
Auger electron spectroscopy (AES), 136, 138
Autodoping, 120–121, 207, 509–510
Automation, 34–38, 55, 524, 594–595
  computer integrated manufacturing (CIM), 35–36
  cost analysis, 38
  mechanization, 35, 36
  SECS equipment interface, 36–37
  shop-floor-control software, 35
Axial fan unit, 26, 27

Ball-grid array (BGA) package, 578–580
Ballroom concept of design, 15
Bamboo effect, 668, 670
Barrel reactor, 344
Baseline-management method, 30–31, 54
Baseline specification, 638–639
Batch-furnace thermal processes, 144, 150–151
Beam irradiation time, 310
Bias-sputtered quartz, 421–422
BiCMOS technology, 502–510
  buried layers, 509
  defined, 502
  epitaxy and autodoping, 509–510
  further enhancement, 507–509
  process design approaches to, 503
  standard buried-collector (SBC), 495, 496, 504
  triple diffused (3D), 495, 503–504
  twin-well, 474, 475–477, 504–507
Bipolar technology, 495–502
  base regions, 501–502
  high-performance transistor, 499
  poly emitter structure, 501
  scaling, 499–500
  self-aligned structures and fabrication, 497–499
  standard process, 495–497
  super-high-performance SiGe HBT, 502
  vertical impurity profile, 500–501
Bird's beak encroachment, 133, 478, 520
Boats, 211

Micron Ex. 1029, p. 740
Micron v. YMTC
IPR2024-00911

Bonding pads, 448
Borderless contacts, 446–447, 454
Boundary layer, 208
Bridging, 401–402
Building in reliability (BIR), 697
Built-in quality and reliability, 646
Buried contact, 515

Cassetteless wet-benches, 39
Catadioptric projection, 282, 283
Centrifugal fan unit, 26
Centrifugal spray cleaning, 92
Ceramic-package fabrication, 552–554
Channel electron mobility, 78–79, 82
Chemical/mechanical polishing (CMP) process, 64, 261, 433–444
  difficulties in, 439, 442–444
  implications of, 435, 437
  mechanism for, 436, 438–439
  polishing pads, 442
  requirements for, 434–435
  uniformity, 440–442
Chemical supply system, 44–47
Chemical vapor deposition (CVD), 131, 206–208, 386–394
  Al, 391–392
  characteristics and applications of, 219
  Cu, 391
  defined, 206
  equipment, 209–213
  excimer laser-assisted, 128
  gases used in, 214
  limited-reaction-processing, 128
  metal-organic, 256–257
  and metallization, 386–394
  photo-chemical, 126, 128
  photon-induced, 247, 249–251
  plasma-enhanced, 126, 128, 160, 206–207, 212
  rapid thermal processing, 128, 163–164
  safety issues, 213–215
  $TaSi_2$, 392–393
  Ti, 389
  TiN, 388–389
  $TiSi_2$, 394
  ultraclean low-pressure, 126, 128
  ultrahigh vacuum, 126, 128–129, 160, 164
  uses of, 215–216
  W, 386–388
  WSi_2, 393–394
Chemically amplified resists, 292
Chilled water system, 27
Chip bonding, 541–547
  epoxy, 545–547
  eutectic, 544–545
Chip interconnection, 541
Clean machines, 12
Cleanliness class, 19–20, 524
Cleanroom design, 8–24
  acoustical noise, 21
  air quality, 21
  automation, 34–38
  ballroom concept, 12, 15
  basic concepts applicable to, 14–19
  chemical supply system, 44–47
  cleanliness class, 19–20, 524
  cleanroom concept, 14–19
  construction project approach and schedule, 30–31
  cost analysis, 31–32
  development of, 12–14
  DI water, 38–44
  electromagnetic interference (EMI), 22–23
  electrostatic discharge (ESD), 23
  Federal Standard 4, 5–8, 19, 209
  grounding, 22
  installation, 24–32
  layout, 9–12
  minienvironment concept, 12, 13
  operations, 32–34
  outgassing, 21, 50
  performance criteria, 19–23
  pressurization, 21
  process gas piping system, 47–52
  process utilities requirements, 8–9
  sizing, technical data for, 23–24
  space management, 52–54
  temperature and humidity control, 20–21
  tunnel concept, 12
  vibration isolation, 22
Cleanroom installation:
  chemical supply system, 44–47
  cost analysis, 31–32
  DI water, 38–44
  process gas piping system, 47–52

Micron Ex. 1029, p. 741
Micron v. YMTC
IPR2024-00911

schedule for, 30–31
space management, 52–54, 55
subsystems, 24–30
chilled water system, 27
cleanroom partition system, 29–30
exhaust systems, 25
filter ceiling system, 28–29
hot water system, 27
make-up air system, 25
raised floor system, 30
recirculation air system, 25–27
Cleanroom layout, 9–12
Cleanroom operations, 32–34
gowning procedures, 14, 33, 54
maintenance, 33–34
Cleanroom partition system, 29–30
Cleanroom pressurization, 21
Cleanroom technology:
automation, 34–38, 55, 524, 594–595
cleanroom classification, 4–8
cleanroom design concept, 8–24
cleanroom installation, 24–32
cleanroom operations, 32–34
overview of, 1–4
related facility systems, 38–54
trends in, 55
Closed-system chemical cleaning, 92
Clustered plasma processing, 351–352
CMOS technology, 489–495
dual polysilicon gates, 494–495
fabrication process sequence, 489–492
latch-up, 493–494
overview of, 489
Collimation, 385–386
Compressed Gas Association (CGA) cylinder valve, 52
Computer integrated manufacturing (CIM), 35–36
Concurrent quality and reliability (Q&R) parameters, 643
Conditioning, 442
Conduction, 564
Conformal step coverage, 221
Contact angle, 88–89
Contact electromigration, 448–450, 668–671
Contact printing, 272
Contact resistance, 376–377, 397, 398
Contamination control, 1–2, 156, 364, 620
Continuous improvement, 646–652
problem-solving methodology, 648–651
supplier program, 651–652
Contrast, 291
Control, 3
Convection, 564–565
Conventional Si epitaxy, 107, 115–125
advantages of, 116
defect generation and minimization, 121–122
deposition equipment, 123–125
doping and autodoping, 118–121, 163–164
growth chemistry, 116–118
origins of, 115
pattern shift, distortion, and washout, 122–123
Cost-effectiveness, 3, 30–31
Cracking resistivity, 567, 568
Cross-talk noise, 560, 561
Crystallographic defects, 139
Cumulative distribution function (CDF) of failure, 693
CVD plugs, 407–410
Cycle time, 641

Damascene process, 444–445
Defect(s), 121–122, 177–179
density, 1–2, 68, 310
minimization, 125–126
Density, 520–521
Denuded substrates, 122
Depletion width, 519
Deposition pressure, 227
Deposition processes, 205–263
chemical vapor deposition (CVD), 131, 206–208, 386–394
collimation, 385–386
electro- and electroless plating, 394–395
electron cyclotron resonance chemical vapor deposition, 98, 251–253, 261
interlevel dielectric deposition (ILD), 413–424
oxide, 164–167
photon-induced chemical vapor deposition (PHCVD), 247, 249–251

Micron Ex. 1029, p. 742
Micron v. YMTC
IPR2024-00911

Deposition processes, *continued*
physical vapor deposition (PVD), 208–209, 379–385
plasma-assisted, 240–247
polycrystalline silicon, 169–171, 226–240
reaction mechanisms, 205–209
silicon nitrides, 168–169, 225–226, 259–260
silicon oxides, 215–218
trends in, 262–264
ultrahigh-vacuum chemical vapor deposition, 126, 128–129, 160, 164, 253–256
variables, 218–221
Design for reliability concept, 656
Designed-in manufacturability, 609
Device and contact degradation, 424
DI water, 38–44
Diameter Index Safety System (DISS) cylinder valve, 52
Dielectric layers, 215
Diffraction-limited optics, 276
Diffusion-controlled kinetics, 193–195
Directional enhancement, 385–386
Direct writing, 311
Distortion, 122–123
Doping, 118–121, 163–164, 231–232
Double-implant lightly doped drain (DI-LDD), 481–482
Downstream plasma etchers, 345
Drain engineering, 480–482
double-implant lightly doped drain (DI-LDD), 481–482
large-angle-tilt implanted drain (LATID), 480–481
lightly doped drain (LDD), 480
Dry cleaning technology, 93–100
AR/H_2 plasma cleaning, 94, 98–99
HF/H_2O vapor cleaning, 94, 96–98
thermal cleaning, 94, 99–100
ultraviolet-ozone cleaning (UVOC), 94, 95–96
Dry-coding principle, 27
Dry etching, 329
(*See also* Plasma etching; Reactive ion etching)
Dry filter ceiling system, 28
Dual damascene process, 445
Dynamic random-access memory (DRAM) processing technology, 2, 133–134, 510–513

Effective emissivity, 146
Electromagnetic interference (EMI), 22–23
Electromigration (EM), 657, 663–673
ac versus dc, 688–690
contact, 448–450, 668, 671
defined, 663
lifetime, methods to improve, 672–673
of lines, 667–668, 669, 670
scaling, 685–686
stress on, 665–667
vias, 671
Electron beam exposure, 272
Electron confinement, 346–347
Electron cyclotron resonance chemical vapor deposition (ECR CVD), 98, 251–253, 261
Electron cyclotron resonance plasma etchers (ECRs), 348–349
Electron lithography, 295–311
applications, 309–311
commercially available systems, 296, 297
electron optics, 296–300
electron resists, 305–309
projection printing, 305
proximity printing, 304–305
raster scan and vector scan, 300–302
resolution in, 295
variable shaping, 303–304
Electron optics, 296–300
Electron PVD, 305–309
Electron resists, 305–309
Electron spectroscopy for chemical analysis (ESCA), 136, 138
Electron wind force, 663
Electronegativity, 67
Electrostatic discharge (ESD), 23
Embryo, 109
Emissivity, 146–149, 173
End-of-range (EOR) damage, 177
End-point detection, 198
Epitaxial films:
chemical and electrical properties, 136–138
physical properties, 135–136

Micron Ex. 1029, p. 743
Micron v. YMTC
IPR2024-00911

Epitaxial lateral overgrowth (ELO), 132, 133
Epitaxy, 105–139
  basic growth kinetics of, 109–111
  for BiCMOS, 509–510
  conventional Si, 107, 115–125
  defined, 105–107
  doping, 118–121, 163–164
  films, characterization of, 135–138
  fundamental aspects of, 107–115
  growth models for, 111–115
  growth sequence, 107–109
  low-temperature epitaxy (LTE), 125–131
  molecular beam epitaxy (MBE), 126, 127, 129–130, 160, 208–209
  pattern-dependent, 342
  rapid thermal processing for, 159–164
  selective epitaxial growth (SEG) of Si, 107, 131–134, 161–163
  surface preparation, 110
  thin-film deposition, 164
  trends in, 138–139
  vapor-phase, 107
  wafer cleaning and, 85–92
Epitaxy-over-trench (EOT) process, 133
Epoxy chip bonding, 545–547
Etch rate ratio (ERR), 341
Etchback planarization, 224, 420, 427–433
Etching, 329–367
  challenges in, 366–367
  immersion, 365
  laser interferometry, 362–363
  optical emission spectroscopy, 362
  oxide, 357–359, 431–433
  overview of, 329–330
  plasma, 329, 330–343
  plasma-induced damage, 363–364, 366
  plasma-process-induced particle contamination, 364
  reactive plasma techniques and equipment, 294, 329, 343–353
  wet chemical, 329, 364–366
Eutectic chip bonding, 544–545
Evaporation technology, 208
Excimer laser-assisted CVD, 128
Exhaust systems, 25
Extrinsic emissivity, 146, 148

Failure distribution, 692–696
Faraday cage, 22–23
Federal Standard, 4, 5–8, 19, 209
Filter ceiling system, 28–29
Filter fan unit, 26
Flexibility, 3, 14, 26, 54
Flip-chip package, 577–578
Fluid filter ceiling system, 28

Gas-source MBE (GSMBE), 130
Gate oxide quality, 62, 78, 85–92
Glass-sealed refractory packaging, 535
Glazing, 442
Global planarization, 425, 428, 431–433
Glow discharge, 330, 333–334
GOLDEP piping, 49
Gowning procedures, 14, 33, 54
Groove and stain, 136
Grounding, 22
Growth hillock, 121
Gull-wing leads, 533

Halo implant, 481–482
Helicon wave plasma etchers, 351, 352
Hermetic packages, 531, 532
Heteroepitaxy, 105, 122
Heterogeneous reactions, 206
Heterojunction bipolar transistor (HBT), 502
HF/H_2O vapor cleaning, 94, 96–98
High-density plasma (HDP), 422–423
High-resolution electron energy loss spectroscopy (HREELS), 86, 95
Homoepitaxy, 105
Homogeneous reactions, 206
Horizontal-tube PECVD reactors, 242
Hot carrier injection (HCI), 657–663
  ac and dc, 687–688
  defined, 657
  equations of, 660–662
  NMOS and PMOS transistor, 658–660
  scaling and, 684–685
Hot holes, 680
Hot-wall reactors, 242
Hot water system, 27
Humidity control, 20–21, 25
Hybrid-excitation CVD, 264
Hydrocarbon contamination, 21

Micron Ex. 1029, p. 744
Micron v. YMTC
IPR2024-00911

IMEC (Interuniversity Microelectronic Center) cleaning, 93
Immersion cleaning, 92
Immersion etching, 365
Implantation-diffusion interaction matrix, 179
Implantation through polysilicon (ITP), 234
Inductively coupled plasma etchers, 349, 351
Innerlead bonding, 574–575
Interface trap generation, 658, 660
Interlevel dielectric (ILD) deposition, 413–424
  bias-sputtered quartz, 421–422
  device and contract degradation by, 424
  high-density plasma (HDP) enhanced SiO_2, 422–423
  plasma-enhanced SiO_2, 413–419
  spin-on-glass (SOG), 419–421
Intrabay automation, 36, 55
Intrinsic emissivity, 146
Ion implantation, 153–154
Ion lithography, 320–322
Isolation, 477–479
  LOCOS process, 131, 133, 225, 477–479
  poly-buffered LOCOS, 479
  shallow trench, 479

J-leads, 533
J-ramp test, 684
Junction(s), 154, 179–185
  capacitance, 372
  depth, 405, 455, 482–484, 519
  leakage, 403, 405–406
  raised source/drain, 454–455

Kooi effect, 478, 479

Lagging parameters, 643
Laminar airflow, 13–14, 109
Laminar-flow cleanroom concept, 4
Large-angle-tilt implanted drain (LATID), 480–481
Laser ablation, 264
Laser beam scanning, 272
Laser interferometry, 362–363
Laser PHCVD, 251
Lateral diffusion spread, 519
Lightly-doped drain (LDD), 480
Limited reaction processing (LRP), 159
Limited-reaction-processing CVD (LRP-CVD), 128
Line defects, 125
Line-width dimension, 519
Lithography, 270–324
  defined, 270
  depth of focus, 426–427
  electron, 295–311
  ion, 320–322
  multilayer, 361
  optical, 272–295
  trends, 323–324
  x-ray, 312–320
Local interconnects (LIs), 403, 404, 485–487
Local planarization, 425
LOCOS process, 131, 133, 225, 477–479
  poly-buffered, 479
Lognormal distribution function, 695–696
Low-dose implants, 181
Low-pressure CVD, 164, 206
Low-temperature epitaxy (LTE) of Si, 125–131
  chemical aspects, 125–126
  device applications, 130–131
  growth techniques, 127–130
  surface preparation, 127
LPCVD polysilicon films, 226–240, 357
  deposition variables, 227–229
  doping, 231–232
  oxidation, 236–240
  oxygen-doped polysilicon, 236
  properties of, 232–236
  structures, 229–231
LPCVD silicon nitrides, 225–226
LPCVD silicon oxides, 215–225

Magnetic-enhanced reactive ion plasma etchers (MERIEs), 346–347
Magnetron, 381
Maintenance (*see* Cleanroom operations)
Make-up air system, 25
Manufacturing:
  accuracy and consistency in, 605, 606
  automation, 34–38, 55, 524, 594–595

Micron Ex. 1029, p. 745
Micron v. YMTC
IPR2024-00911

certification procedure and control, 617–618
continuous improvement, 646–652
cost structure of, 158–159, 588, 589–590, 593, 697
critical success factors, 587–588
cycle time and inventory control, 641–642
defined, 587
delivery time, 588
equipment availability and utilization, 639–640
equipment checks, 619–623
infrastructure considerations, 608–609
material requirements, 595–602
metrology and analytic capability, 605–608, 612
parametric parameters, 628–630
organization for, 588–589, 591–592
process characterization/verification, 623–626
process flow, 588, 589
process layout considerations, 593–594
process and product baseline, 638–639
product characterization, 626–634
product qualification, 634–637
quality, 587
service, 588
specification generation, 614, 615
specification systems and operation procedures, 602–605
start-up technology, 609–639
sustaining production, 642–646
trends, 652–653
utility requirements and control, 614, 616–619
volume ramp-up considerations, 639–646
Mask fabrication, 309–311
Masking tolerances, 520
Mass-transport-controlled growth process, 108, 113
Mass-transport-limited deposition, 207, 228
Matching network, 332, 333
Mechanization, 35, 36
Megasonic cleaning process, 64–65, 66, 92
Memory effect, 120
Metal contamination, 66–71
Metal corrosion, 450–451
Metal-organic CVD (MOCVD), 256–257
Metallization, 192–196, 371–456
chemical vapor deposition (CVD), 386–394
collimation, 385–386
contact electromigration, 448–450
contact resistance, 376–377, 397, 398
cost reduction issues, 455–456
defined, 371
demand for, 451–452
electro- and electroless plating, 394–395
low-dielectric-constant materials, 452–454
metal corrosion, 450–451
multilevel, 377–379, 412–448
new conductor materials, 452
new contact structures, 454
parasitics: in circuit performance, 373–376
in device performance, 371–373
physical vapor deposition (PVD), 379–385
plug processes, 406–412
raised source/drain structures, 454–455
reliability, 379, 448–451
RTA of silicides, 198–199
silicide growth kinetics, 199–201
silicide process, 395–406
Microloading, 342, 343, 366
Micromasking, 360
Microscopic uniformity, 342–343
Minienvironment concept of design, 12, 13
Modulation transfer function (MTF), 276
Moisture-induced cracking resistivity, 567, 568
Molecular beam epitaxy (MBE), 126, 127, 129–130, 160, 208–209
MOS memory technology, 510–518
DRAM processing technology, 510–513
representative process flow, 510–513

Micron Ex. 1029, p. 746
Micron v. YMTC
IPR2024-00911

MOS memory technology, *continued*
soft-error rate (SER) consideration, 517–518
SRAM processing technology, 513–517
Multilayer lithography, 361
Multilevel interconnects, 488–489
Multilevel metallization, 377–379, 412–448
advantages of, 378
bonding pad limitations, 448
interlevel dielectric (ILD) deposition, 413–424
planarization, 424–447
problems in, 412
processing complexity of various structures, 456
Multi-poly technology, 403

N-well process, 474
Native oxide growth, 73–74, 79–85
Negative resists, 290, 305–306
Nitride etching, 357–359
Noise, 21, 560–562
Nucleation process, 109, 193
Nucleus, 109

Off-axis alignment, 288
Off-axis illumination, 286–288
Open process cooling water system, 27
Optical emission spectroscopy, 362
Optical lithography, 272–295
contact and proximity printing, 272–275
enhancement, 284–288
optical resists, 290–293
overlay accuracy, 288–290
process technologies, 293–295
projection printing, 275–284
Optical printers, 273
Optical resists, 290–293
Organic contamination, 71–74
Organic materials etching, 361
Outgassing, 21, 50
Overlay accuracy, 288–290
Oxidation (*see* Rapid thermal oxidation)
Oxide breakdown, 679–684
defined, 679
dynamic testing and, 690
lifetime prediction of, 682–683
mechanisms of, 680–682
scaling on, 684
V-ramp and *J*-ramp measurement techniques, 684
Oxide deposition, 164–167
Oxide etching, 357–358, 431–433
Oxygen-doped polysilicon, 236
Oxynitrides, 246
Ozone concentration, 73–74

P-well process, 474
Packaging:
a-particle protection, 573
ball-grid array (BGA) package, 578–580
cleaning and assembly environment, 570–572
cost of, 531
design features: electrical considerations, 558–562
mechanical considerations, 566–569
thermal considerations, 562–565
device trends and, 535–537
fabrication technologies: ceramic-package, 552–554
glass-sealed refractory, 555
plastic molding, 555–558
flip-chip package, 577–578
hermetic versus plastic, 531–532, 533
importance of, 530
for logic and microprocessor ULSI devices, 538–539
mechanical characteristics of materials, 544
for memory ULSI devices, 537–538
packaging-level integrations, 539
reliability testing, 571
sealing and corrosion, 572
tape carrier package (TCP), 573–577
through-hole (TH) and surface-mount (SM) types, 532–535
trends in, 580–582
Parallel-plate PECVD reactors, 242
Parasitic capacitors:
in circuit performance, 373–376
in device performance, 371–373
Partial planarization, 425
Particle adhesion, 63–66
Pattern-dependent etching, 342

Micron Ex. 1029, p. 747
Micron v. YMTC
IPR2024-00911

Pattern shift, 122–123
Phase-shifting mask, 284, 286, 287
Photochemical CVD (photo-CVD), 126, 128
Photon-induced chemical vapor deposition (PHCVD), 247, 249–251
Photoresist(s), 71, 290–293
  commercially available, 292, 293
Physical vapor deposition (PVD), 208–209, 379–385
  electron, 305–309
  sputtering, 209, 210, 379, 380–385
  surface-imaging processes, 293–295
  x-ray, 318, 320
Planarization, 224, 261, 424–447
  categories of, 425
  chemical/mechanical polishing (CMP), 433–444
  damascene and dual damascene, 444–445
  depth of focus, 426–427
  dummy features, generating, 428–430
  etchback, 224, 420, 427–433
  global, 425, 428, 431–433
  local, 425
  stacked and borderless contacts, 446–447
Planarization block mask (PBM), 428
Planarization etching, 361–362
Plasma-assisted deposition, 240–247
  film properties, 244–247
  horizontal-tube reactors, 242
  methods, 241–244
  parallel-plate reactors, 242
  single-wafer reactors, 243–244
  variables, 244
Plasma-enhanced CVD (PECVD), 126, 128, 160, 206–207, 212
Plasma-enhanced SiO_2, 413–419
Plasma etching, 329, 330–343
  ac plasma discharge, 331
  aluminum and tungsten metal etching, 359–361
  anisotropy, 338–341
  damage caused by, 363–364
  end point systems, 362–364
  etch chemistries, 354
  etch mechanisms, 334–336
  fluorocarbon chemistry, 336–338
  glow discharge chemical phenomena, 333–334
  helicon wave etchers, 351, 352
  low-pressure gas discharge, 330–334
  matching network, 332, 333
  microscopic uniformity, 342–343
  organic materials etching, 361–362
  oxide and nitride etching, 357–358
  particle contamination, 364
  planarization etching, 361–362
  polysilicon and polycide gate etching, 357, 359
  RF discharge, properties of, 331–333
  selectivity, 341–342
  silicon trench etching, 355–357
Plastic packages, 531–532, 533, 555–557
Plating, 394–395
Plugs, 406–412
  aluminum, 410–412
  blanket CVD W, 407–410
  CVD Al, 412
  CVD Cu, 412
  selective CVD W, 407
Point defects, 121, 153
Poisoned vias, 260
Poly-buffered LOCOS, 479
Poly-metal interlevel dielectric (PMD) films, 215–216
Polycides, 226, 258, 357, 386
  gate structure, 395–396
Polycrystalline silicon deposition, 169–171, 226–240
Polyoxides, 236–240
Polysilicon films:
  applications of, 257–262
  deposition of, 226–240
  etching, 357, 365
Positive charge-trapping sites, 680
Positive resists, 290–291, 305–306
Preoxide process, 84
Pressure National Code (PNC) levels, 21
Preventive maintenance, 32
Probability density function (PDF) of failure, 693
Probe testing, 631–634, 635
Process characterization/verification, 623–626
Process exhaust systems, 25

Micron Ex. 1029, p. 748
Micron v. YMTC
IPR2024-00911

Process gas piping system, 47–52
Process integration:
  BiCMOS technology, 502–510
  bipolar technology, 495–502
  CMOS technology, 489–495
  compactness, 520–521
  design rules, 519–520
  drain engineering, 479–482
  evolution of, 472–473
  isolation, 477–479
  local interconnects, 485–487
  MOS memory technology, 510–518
  multilevel interconnects, 488–489
  process parameter monitoring, 522
  process window and sensitivity, 521–522
  salicidation, 484–485
  self-aligned structures and contact, 487–488
  shallow junctions, 482–484
  step heights, 521
  trends in, 523–525
  in ULSI fabrication technology, 519–522
  well structures, 474–477
Process modules, 9–10
Process parameter monitoring, 522
Process speed, 641
Process window, 521–522
Product characterization, 626–634
Product qualification, 634–637
Project approach, 3, 30–31
Project organization, 3
Projection printing, 275–284
  electron, 305
  reflection and catadioptric projection, 282, 283
  refraction projection, 282, 284
Proximity effect, 307
Proximity printing, 274–275, 304–305, 313–314

Radiation, 565
Raised floor system, 30
Rapid thermal annealing (RTA), 176–192
  defect annealing, 177–179
  dopant activation, 187–192
  of ion-implanted junctions, 154, 179–187
  of silicides, 192–196
Rapid thermal nitridation (RTN), 174
Rapid thermal oxidation (RTO), 155, 171–176
  oxide growth comparisons, 171–174
  processed oxides, properties of, 174–176
Rapid thermal processing (RTP), 159–198
  advantages of, 196
  alternative approaches to, 150, 167
  annealing, 176–192
  and conventional furnaces compared, 150–151
  defined, 145–146
  epitaxy, 159–164
  to form silicides, 397
  metallization, 192–196
  oxidation, 171–176
  thermophysics in, 146–149
  thin-film deposition, 164–171
  trends in, 196, 198
Rapid thermal processing CVD (RTPCVD), 128, 163–164
Raster scan, 309–311
Rate-limiting step, 108
RCA clean process, 60–61, 127
Reactive ion etchers (RIEs), 345–346
Reactive ion etching (RIE), 294, 329, 343–353
  barrel reactor, 344
  charge buildup, 364
  clustered plasma processing, 351–352
  commercially available plasma sources, 350
  downstream plasma etchers, 345
  electron cyclotron resonance plasma etchers (ECRs), 348–349
  helicon wave plasma etchers, 351, 352
  inductively coupled plasma etchers (ICPs), 349, 351
  magnetic-enhanced etchers (MERIES), 346–347
  reactive ion etchers (RIEs), 345–346
Reactive sputtering, 383–384
Recirculation air system, 25–27
Reduced-pressure reactors, 209
Reflection projection, 282, 283
Reflow, 222
Refraction projection, 282, 284
Registration error, 289

Micron Ex. 1029, p. 749
Micron v. YMTC
IPR2024-00911

Reliability
built-in, 646
components of, 656–657
design for, 656
dynamic testing of oxides, 690
electromigration (EM), 668–673
failure distribution, mathematics of, 692–696
failure rate concepts, 693
hot carrier injection (HCI), 657–663
lognormal distribution function, 695–696
oxide breakdown, 679–684
plasma damage, 691–692
recent problems in, 690–692
relations between dc and ac lifetimes, 686–690
scaling, 684–686
stress migration (SM), 674–679
tests of, 636
trends in, 696–697
tungsten plug, 691
wafer-level, 630–631
Weibull distribution, 693–694
Resists (*see* Photoresist(s))
Retrograde well, 477
Reverse osmosis units, 40
Ring field, 282
Roughness (*see* Surface microroughness)
Rutherford backscattering spectroscopy (RBS), 136, 138

Salicidation process, 484–485
Salicide (*see* Self-aligned silicide)
Salicide structure, 234
Scaling, 684–686
effect of on electromigration (EM), 685–686
effect of on hot carrier integration (HCI), 684–685
effect of on oxide reliability, 686
effect of on stress migration (SM), 686
Scanning tunneling microscope (STM), 74
Scrubbers, 25
Secondary-ion mass spectroscopy (SIMS), 136, 137–138
Selective epitaxial growth (SEG) of Si, 107, 131–134
applications,133–134
fundamental aspects, 132–133
Selective-epitaxy-base transistor (SEBT), 134
Selective-epitaxy emitter-window (SEEW) process, 134
Selectivity, 341–342
Self-aligned silicide (salicide), 192–193, 226, 258, 376, 397–402
layout area, effects on, 399, 400
processing issues, 399, 401–402
transistor performance, effects on, 398–399
Self-alignment, 487–488
Semi-insulating polysilicon (SIPOS), 236
Semiconductor Equipment Communications Standard (SECS) equipment interface, 36–37
Sensitivity, 521–522
Sequential-sampling option, 7–8
Shallow trench, 479
Shaped-beam method, 303
Sheath, 331
Shiraki clean process, 91–92, 127
Shop-floor-control software, 35
Si beam cleaning, 127
Signal delay, 558–559
Signal reflection, 559–560
Silicide structures, 258, 395–406
junction leakage, 403, 405–406
local interconnects, 403, 404
polycide gates, 395–396
shallow source/drain junctions, 403–406
(*See also* Self-aligned silicide)
Silicon dioxide etching, 365–366
Silicon nitride deposition, 168–169, 225–226, 259–260
Silicon nitride etching, 365
Silicon oxides
deposition processes, 215–218
deposition variables, 218–221
film properties of, 224–225, 248
low-pressure CVD, 215–216
step coverage and reflow, 221–224
uses, 258–259
Silicon oxynitride, 226
Silicon trench etching, 355–357
Simultaneous switching noise, 560–562

Micron Ex. 1029, p. 750
Micron v. YMTC
IPR2024-00911

Single-wafer PECVD reactors, 243–244
Smoothing, 425
Soft-error rate (SER), 517–518
Solid-phase epitaxy (SPE), 127
Source inspection, 620, 621
Space management, 52–54, 55
Spin-on-glass (SOG), 419–421
Spray etching, 365
Sputtering, 209, 210, 379, 380–385
  alloy, 383
  collimation, 385–386
  film properties, 382–383
  from quartz, 421–422
  reactive, 383–384
  step coverage, 384–385
Stacked contacts, 446, 454
Standard buried-collector (SBC) BiCMOS, 495, 496, 504
Standard cleaning 1 (SC-1), 61, 64, 67, 77
Standard cleaning 2 (SC-2), 61, 67
Standard mechanical interface (SMIF) pods, 12
Static random access memory (SRAM) processing technology, 513–517
Statistical process control (SPC), 522
Stefan–Boltzmann law, 146
Step height, 521
Stigmator coils, 300
Stress, 139–140, 665–667
Stress migration (SM), 674–679
  defined, 674
  dependence of on AI aspect ratio, 677
  mechanism, 674–677
  prevention, 677–679
  scaling, 686
  testing for, 677
Submicrometer technology, 3
Surface-controlled growth process, 108
Surface-imaging resists, 293–295
Surface microroughness, 74–79
Surface morphology, 139
Surface-reaction-rate-limited deposition, 207, 228, 237
Synchrotron radiation, 316, 318, 319

Tape-automated bonding, 574–575
Tape carrier package (TCP):
  package fabrication, 575–577
  tape-automated bonding, 574–575
  tape carrier, 573–574
TCE mismatch, 567, 569
Temperature control, 20–21
Thermal cleaning, 94, 99–100
Thermal processes:
  ambient control budget, 155
  annealing, 176–192
  batch furnaces, 144, 150–151
  contamination budget, 156
  cost issues and manufacturing budget crisis, 158–159
  device electrical budget, 156–158
  metallization, 192–196
  overview of, 144–146
  rapid thermal processes, 146–151
  requirements for, 151–159
  thermal budget, 152–155
Thermal stress, 566–567
Thickness tolerances, 519–520
Through-the-lens (TTL) alignment, 288–290
Time-dependent dielectric breakdown (TDDB), 657
Time-zero dielectric breakdown (TZDB), 175
Transmission electron microscopy (TEM), 229
Trenching, 357, 359
Triboelectricity, 23
Triple-diffused bipolar transistors (3D), 495, 503–504
Tungsten etching, 359–361
Tunnel concept of design, 12
Turbulent airflow, 14, 109
Twin-well process module, 474, 475–477, 504–507
Type 1 SEG-Si, 132–133
Type 2 SEG-Si, 132, 133

U descriptor, 6
Ultraclean low-pressure CVD, 126, 128
Ultrahigh-vacuum chemical vapor deposition (UHV/CVD), 126, 128–129, 160, 164, 253–256
Ultraviolet-ozone cleaning (UVOC), 94, 95–96

V-ramp test, 684
V-type (vacancy) defects, 177
Vapor-phase epitaxy (VPE), 107
Vector scan, 309–311

Micron Ex. 1029, p. 751
Micron v. YMTC
IPR2024-00911

Vibration isolation, 22
Volume defects, 121

Wafer acceptance test (WAT), 522
Wafer cleaning:
  contamination, sources and effects of, 61
  dry cleaning technology, 93–100
  and epitaxy, 85–92
  evolution of, 60–61
  last step for gate oxidation/epitaxy, 85–92, 122
  megasonic process, 64–65, 66, 92
  metal contamination, 66–71
  native oxide growth, 73–74, 79–85
  organic contamination, 71–74
  particles, 63–66
  RCA wet clean process, 60–61
  removing particles and contamination, mechanisms for, 62–92
  surface microroughness, 74–79
  trends in, 101
  wet cleaning technology, 92–93
Wafer-level reliability (WLR) 630–631, 697
Wafer preparation, 540–541
Washout, 122–123
Weibull distribution, 693–694
Well structures, 474–477
  retrograde well, 477
  twin-well, 474, 475–477, 504–507
Wet chemical etching, 329, 364–366
  mechanisms and techniques, 365
  polysilicon and silicon nitride etching, 365
  silicon dioxide etching, 365–366
Wet cleaning:
  advanced processes, 95–96
  centrifugal spray cleaning, 95
  closed-system chemical cleaning, 95
  immersion technique, 95
  megasonic cleaning process,64–65, 66, 95
  problems associated with, 93–94
Wet-cooling principle, 27
Wien's displacement law, 146
Wire bonding, 547–552
  ULSI design rules, 549–552
Work in process (WIP) inventory, 641

X-ray lithography, 312–320
  proximity printing, 313–314
  synchrotron radiation, 316, 318, 319
  x-ray mask, 314–315
  x-ray projection, 318
  x-ray resists, 318, 320
  x-ray sources, 315–316, 317
X-ray mask, 314–315
X-ray PVD, 318, 320

Yield:
  gate oxide quality and, 62
  manufacturing systems and, 607
  role of contamination control in, 1, 2
  systems affecting, 38

Micron Ex. 1029, p. 752
Micron v. YMTC
IPR2024-00911

ACKNOWLEDGMENTS

Chapter 1

Figure 10, p. 20, from Suzuki, "Supercleanroom" (in Japanese), *Hitachi Hyoron* (Commentary), **68** (9), 737, 1986, reprinted with permission from Hitachi, Ltd.

Figure 14, p. 37, reprinted with permission from Murata Machinery, Ltd.

Figure 16, p. 40; Fig. 19, p. 43; Table 9, p. 44, reprinted with permission from Christ AG. Switzerland.

Figure 18, p. 42, reprinted with permission from Asahi Chemical Industry Co., Ltd.

Figure 24, p. 51, from M. Nakamura, T. Ohmi, and K. Kawada, "All Metal and Oxygen Passivation Tubing Technology for Ultra Clean Gas Delivery System," *Institute of Environmental Sciences Technical Meeting,* San Diego, 1991, p. 605, reprinted with permission from the Institute of Environmental Sciences.

Table 10, p. 45, reprinted with permission from Merck-Kanto Advanced Chemicals Ltd.

Table 11, p. 48, and Table 12, p. 51, reprinted with permission from BOC Gases, Murray Hill, NJ.

Chapter 4

Figure 6, p. 152; Fig. 7, p. 154; Fig. 38, p. 187, © 1990 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.

Figure 9, p. 157; Table 2, p. 195, reprinted with permission from C. W. Osburn.

Figure 11, p. 160; Fig. 12, p. 161; Fig. 15, p. 165; Fig. 16, p. 166; Fig. 17, p. 166; Fig. 28, p. 178; Fig. 31, p. 181; Fig. 33, p. 183, reprinted with permission from Academic Press, Inc.

Figure 18, p. 167; Fig. 32, p. 182; Fig. 44, p. 194, reprinted with permission from Elsevier Science Publishers BV.

Figure 19, p. 168; Fig. 20, p. 169; Fig. 23, p. 172, reprinted with permission from Materials Research Society.

Figure 21, p. 170, reprinted with permission from the American Institute of Physics.

Figure 22, p. 170, © 1991 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.

Figure 26, p. 175; Fig. 27, p. 176, © 1992 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.

Figure 34, p. 185; Fig. 45, p. 196, reprinted with permission from The Electrochemical Society.

Figure 36, p. 186, reprinted with permission from *Journal of Electronic Materials,* Vol. 18 (1989), p. 143, a publication of The Minerals, Metals & Materials Society, Warrendale, PA 15086.

Chapter 6

Table 1, p. 273, reprinted with permission from the April 1993 issue of *Solid State Technology,* © 1993 PennWell Publishing Co.

Figure 4, p. 277, reprinted with permission from the April 1977 issue of *Solid State Technology,* © 1977 PennWell Publishing Co.

Micron Ex. 1029, p. 753
Micron v. YMTC
IPR2024-00911

Figure 5, p. 278, reprinted with permission from the August 1977 issue of *Solid State Technology,* © 1977 PennWell Publishing Co.
Figure 6, p. 279, reprinted with permission from the August 1980 issue of *Solid State Technology,* © 1980 PennWell Publishing Co.
Figure 7, p. 280; Fig. 8, p. 281; Fig. 9, p. 283; Fig. 12, p. 287; Fig. 13, p. 287; Fig. 14, p. 289; Table 2, p. 290; Fig. 22, p. 302, reprinted with permission from Society of Photo-Optical Instrumentation Engineers, Bellingham, WA.
Figure 10, p. 284; Fig. 11, p. 285, © 1982 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 15, p. 291, reprinted with permission from *Solid State Technology,* © 1974 PennWell Publishing Co.
Figure 16, p. 292, © 1975 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 24*a,* p. 304; Fig. 24*b,* p. 304; Fig. 25, p. 306; Fig. 26, p. 309; Fig. 27, p. 312; Fig. 29*b,* p. 314; Fig. 30, p. 319; Fig. 31, p. 320; Fig. 32, p. 322, reprinted with permission from the American Institute of Physics.
Table 8, p. 317, reprinted with permission from the February 1986 issue of *Solid State Technology,* © 1986 PennWell Publishing Co.

Chapter 7

Figure 3, p. 335, © 1992 by International Business Machines Corporation, reprinted with permission.
Figure 4, p. 336; Fig. 6, p. 339; Fig. 7, p. 340; Fig. 19, p. 351; Fig. 20, p. 352, reprinted with permission from the American Institute of Physics.
Figure 5, p. 337; Fig. 16, p. 347; Fig. 18, p. 349; Fig. 22, p. 356; Fig. 23, p. 358, reprinted by permission of the publisher, The Electrochemical Society, Inc.
Figure 8, p. 340, reprinted from D. L. Flamm and V. M. Donnelly, "The Design of Plasma Etchants," *Plasma Chem. Plasma Process.,* **1,** 317 (1981), with permission from Plenum Publishing Group.
Table 2, p. 350, reprinted with permission of *Semiconductor International* magazine, 1993, © by Cahners Publishing Co.
Table 3, p. 354, reprinted with permission of *Semiconductor International* magazine, 1992, © by Cahners Publishing Co.
Figure 24, p. 359, reprinted with permission from *Japanese Journal of Applied Physics.*

Chapter 9

Figure 1, p. 473, © 1991 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 2, p. 475; Fig. 3, p. 476; Fig. 5, p. 481; Fig. 16, p. 498, © 1980 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 6, p. 482, © 1992 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 7, p. 484; Fig. 8, p. 486; Fig. 23, p. 509; Fig. 29, p. 516, © 1987 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 9, p. 487; Fig. 10, p. 488, © 1988 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 12, p. 491; Fig. 27, p. 515; Fig. 28, p. 516, © 1989 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.

Micron Ex. 1029, p. 754
Micron v. YMTC
IPR2024-00911

Figure 14, p. 496; Fig. 17, p. 500, reprinted with permission of Lattice Press.
Figure 15, p. 497, © 1986 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 24, p. 511, © 1993 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 30, p. 518, © 1985, 1989 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.

Chapter 12

Figure 1, p. 659; Fig. 2, p. 659; Fig. 3, p. 661, reprinted with permission from Academic Press, Inc.
Figure 9, p. 669, © 1975 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 11, p. 670, reprinted with permission from H. S. Rathore.
Figure 14, p. 675; Fig. 15, p. 676, © 1985 IEEE. Reprinted with permission from The Institute of Electrical and Electronics Engineers, Inc.
Figure 16, p. 678, reprinted with permission from the American Institute of Physics and H. Yagi.
Figure 25, p. 694; Fig. 26, p. 695, reprinted with permission from McGraw-Hill, Inc.

Micron Ex. 1029, p. 755
Micron v. YMTC
IPR2024-00911

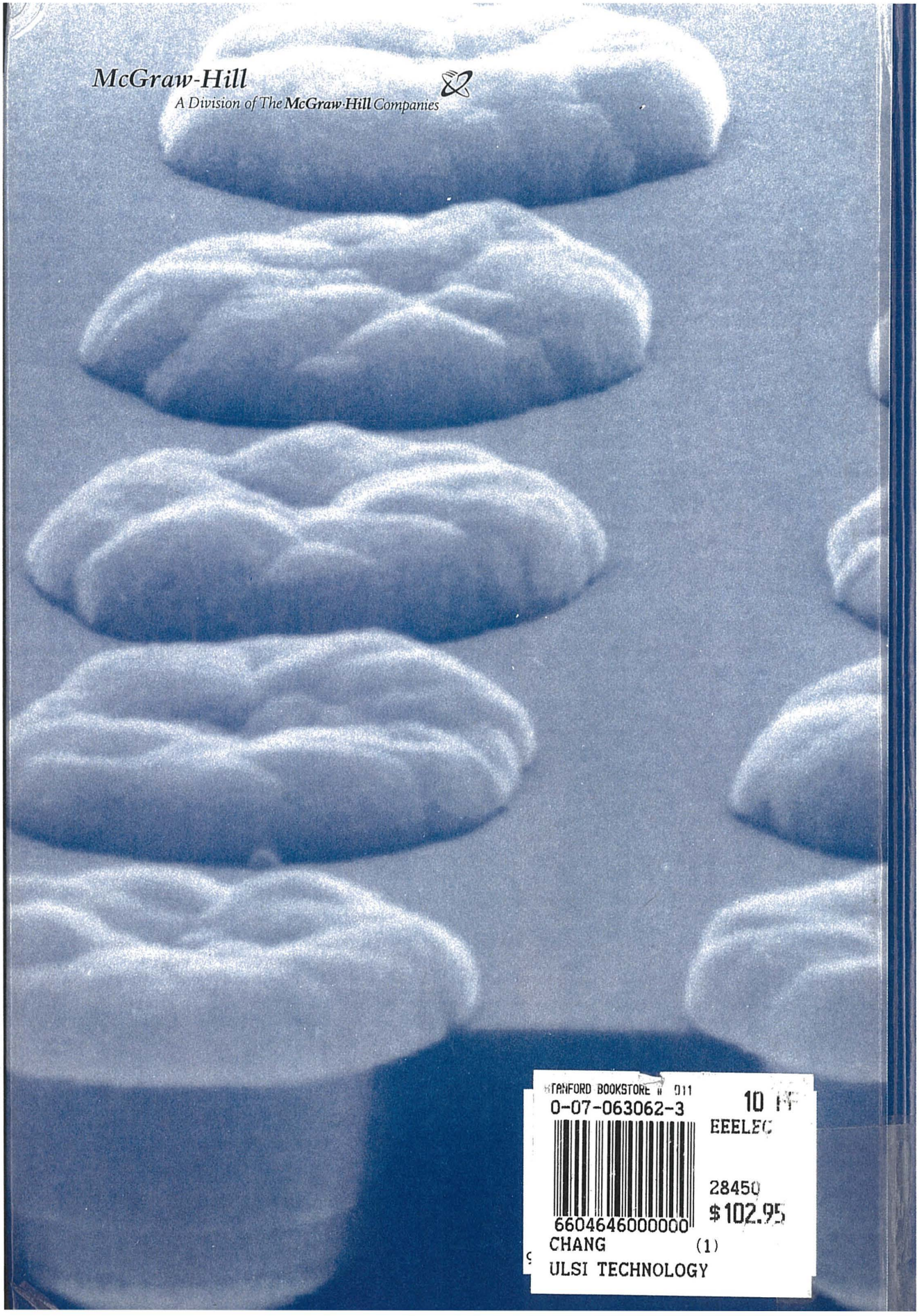
TK7874.76/U17/1996
Chang, C.Y.
ULSI technology / edited
by C.Y. Chang, S.M. Sze. btgh
c.1 men n/a

DATE | ISSUED TO

TK7874.76/U17/1996
Chang, C.Y.
ULSI technology / edited
by C.Y. Chang, S.M. Sze. btgh
c.1 men n/a

DEMCO

Micron Ex. 1029, p. 758
Micron v. YMTC
IPR2024-00911



STANFORD BOOKSTORE # 011
0-07-063062-3
10 NF
EEELEC
28450
$102.95
6604646000000
CHANG (1)
ULSI TECHNOLOGY

Micron Ex. 1029, p. 759
Micron v. YMTC
IPR2024-00911