Kirk-Othmer Encyclopedia of Chemical Technology.

INTEGRATED CIRCUITS

Since its inception in the 1960s, the electronics industry has driven phenomenological changes in the economies of the world. The fundamental cornerstone of the development of the electronics age has been the concurrent decrease in cost associated with designing and manufacturing an increasingly sophisticated and miniaturized unit of integrated circuitry. The basic building block of an integrated circuit (IC) is a device or component which is a specific combination of conducting, semiconducting, and nonconducting layers that performs a single electrical function on an IC chip. Many devices are connected (or integrated) into a complete integrated circuit, or chip. Thousands of chips are formed simultaneously on a wafer. In general, a number of wafers are processed at the same time, although some equipment designs operate on one wafer at a time (single-wafer processing).

The nomenclature of integrated circuits has changed as the complexity of ICs has increased: small-scale integration (SSI) has evolved to medium-scale integration (MSI), to large-scale integration (LSI), and to the mature very-large scale integration (VLSI), which has 10^5 or more devices per chip. The next generation of ICs are classified as ultra-large scale integration (ULSI).

A new generation of IC technology has developed roughly every three years. The Rule of Two holds that approximately for every two generations (six years), the device feature size decreases by two, and other properties such as logic gate speed, chip area, power dissipation, and maximum input/output (I/O) pins increase by two (1). As of this writing, IC technology has advanced to 0.5-μm circuit geometries for volume production, and is reaching into the deep submicrometer (0.35 μm and smaller) dimensions. Forecasts predict an ambitious scaling of device dimensions of 0.25 μm by 1998, and 0.18 μm by 2001 (2). Moreover, the number of chips per wafer is increasing, as well as the size of the chips and wafers. Silicon wafers are now fabricated in sizes up to 200 mm in diameter, and undergoing development of 300–400 mm sizes (3, 4). The increase in the number of components per chip for dynamic random access memory (DRAM) ICs over time is shown in Figure 1 (5). The decrease in feature size and increase in chip area are shown in Figure 2**a;** the changes in cost and speed per device are shown in Figure 2**b.**

Trends in the industry include transferring more of the functions that are found on the supporting printed circuit boards onto the wafers themselves, to reduce the amount of chip packaging and required interconnections (see Electrical connectors; Packaging, semiconductors and electronic materials). This development is known as wafer-scale integration (WSI) where the goal is to design a complete computer on a wafer (see Computer technology). These developments have been supported by concurrent advances in computer-aided design (CAD) tools, both in software and hardware, which have been used to develop integrated CAD design systems that perform IC layout design, simulation, and testing.

Silicon [7440-21-3], Si, technology predominates in the semiconductor industry (see Silicon and silicon alloys). Gallium arsenide [1303-00-0], GaAs, is considered a possible substitute for silicon substrates, based on its potential for high speed applications where it can operate at high (1.9 GHz) frequencies using low power consumptions and high sensitivity (see Electronic materials). One reason that GaAs technology has not fulfilled its promise is that silicon technology has dramatically improved in the interim, particularly with improvements in speed, and has reduced the cost-effectiveness of pursuing GaAs development. Expense has limited usage

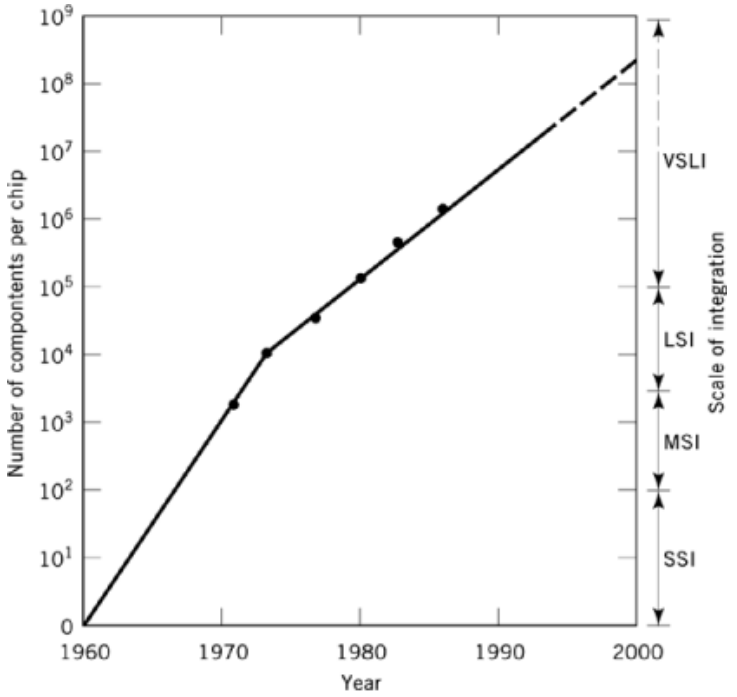

Fig. 1. Change in number of components per chip where ( ) represents projected growth. See text.

of GaAs to microwave devices, primarily for military use (see Microwave technology). However, nonmilitary applications for GaAs devices have been growing, particularly in wireless products such as cellular phones (6).

1. Basics of Silicon Technology

VLSI technology is based on the unique attributes of silicon which have allowed the rapid evolution of integrated circuits. As a semiconductor source silicon has an adequate bandgap in its electronic make-up for the movement of electrons, it forms a stable insulating oxide, is abundant, and is inexpensive to make. In constrast, the use of GaAs has been limited to specialty applications because of its costly source and fragile nature (see Gallium and gallium compounds).

The property that allows silicon to function in a number of capacities is electronic configuration. Silicon has four electrons in its outer shell. Its crystalline structure allows other elements to reside next to silicon and share electron orbitals with silicon, altering its electrical properties. These other elements are called dopants, and are introduced into the silicon structure through doping processes. The most common dopant is boron [7440-42-8], B, which has three electrons in its outer shell. Silicon doped with the electron-deficient boron has an overall positive charge, and is called *p*-type silicon. A second common dopant is phosphorus [7723-14-0], P, having five electrons in its outer shell. Silicon doped with P has an overall negative charge, and is called *n*-type silicon. Arsenic [7740-38-2], As, is another commonly used dopant.

The fabrication of an integrated circuit involves the sequential formation of alternating layers of insulators, semiconductors, and conductors on a silicon wafer. These layers are assembled to form transistor devices that are interconnected to produce particular electrical functions. The layers can be formed by deposition of new material, oxidation of material present on the surface, implantation of additional constituents into surface

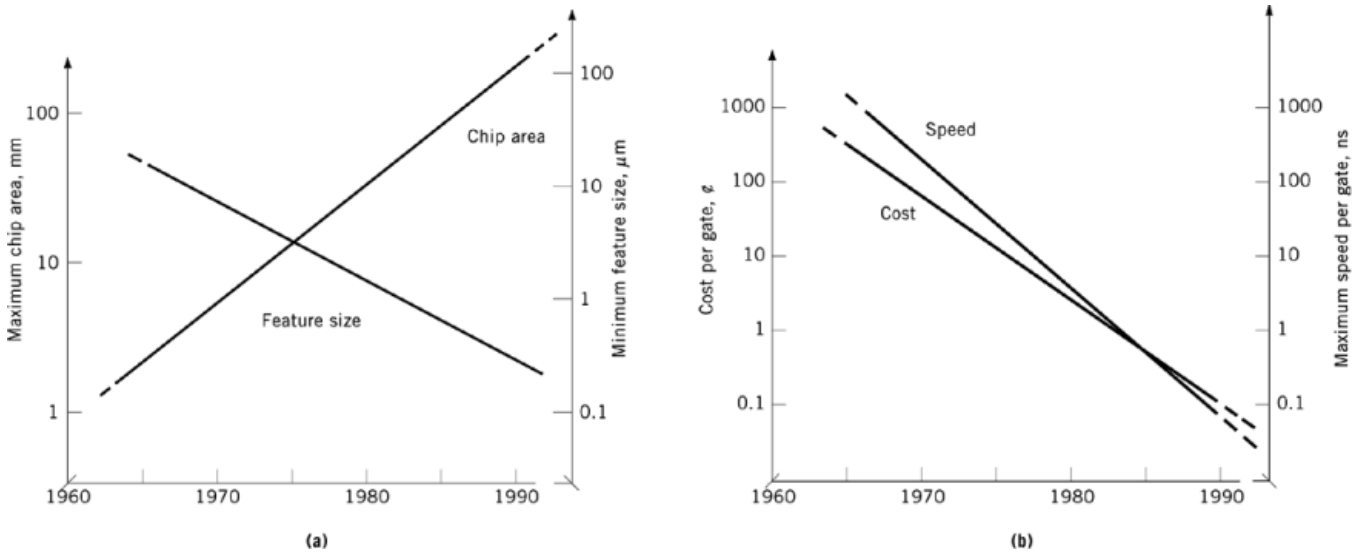

Fig. 2. Improvement over time: (**a**) chip area and minimum feature size; (**b**) maximum speed and cost per logic gate.Courtesy of Custom VLSI Microelectronics.

features, or epitaxial growth of silicon. In order to interconnect the layers, isolate devices from each other, and form integrated circuitry, these layers must be selectively patterned. The patterning is accomplished by photolithography and etching processes (see Lithographic resists).

There are two kinds of integrated circuits (ICs): analogue, or linear ICs, and digital, or logic ICs. Analogue ICs produce, amplify, or respond to various voltages, and are used for any kinds of amplifiers, timers, oscillators, etc. Digital ICs respond to or produce signals that have only two voltage levels. These are used for microprocessors, memories, and microcomputers. It is possible to combine digital and analogue devices on one chip.

Digital IC families are further divided by design and function. The principal IC technologies include *p*- and *n*-channel metal-oxide semiconductors (PMOS and NMOS, respectively), complementary metal-oxide semiconductors (CMOS), bipolar, and integrated-injection-logic (I^2L) devices. Of these, CMOS designs are by far the most popular, having an estimated 73% of the worldwide market in 1994 (7). Leading-edge microprocessors, application specific integrated circuits (ASICs), and DRAM ICs larger than 1 megabyte (Mb) are almost entirely fabricated with CMOS technology. Bipolar devices are the choice technology for high speed applications.

There are several reasons for the widespread use and development of CMOS devices, including low power density, relatively good noise immunity and soft error protection, design simplicity, and the capability to include lower power analogue and digital circuitry on the same chip. The most attractive feature has been the ability to scale CMOS technology to smaller dimensions. Processes exist that produce 0.5-μm dimensions under manufacturing conditions. Development is underway to effect 0.18-μm regimes (2). Another significant development in CMOS technology has been to reduce the power supply voltage from 5.5 to 3.3 V. There is expectation to step down further to 2.5 V as geometries decrease (1, 8). Developments in equipment and processes are necessary to attain these goals.

Typical CMOS devices use both NMOS and PMOS transistors to form logic devices. A simple NMOS transistor is shown in Figure 3**a.** Two channels of *n*-doped silicon are formed in *p*-silicon to form a source and drain. An NMOS transistor is designed to permit a negative charge to move from the source to the drain in response to a positive charge in the gate. When the charge on the gate is large enough such that the

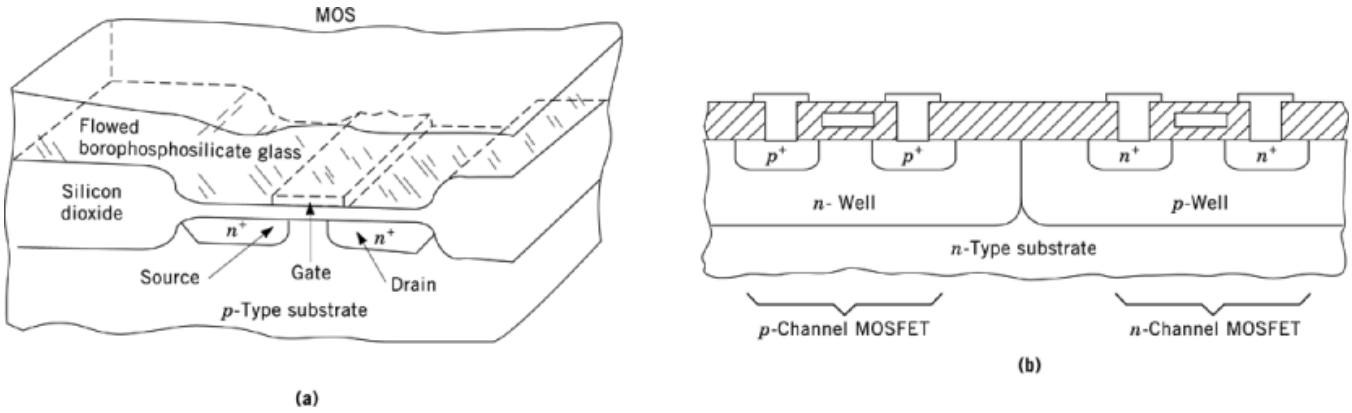

Fig. 3. Cross sections of electronics devices used in ICs. (**a**) NMOS transistor; (**b**) a twin-tub CMOS device on an n-type substrate.Courtesy of Custom VLSI Microelectronics.

source-to-gate voltage is higher than a threshold voltage V_t, electrons create a conducting path between drain and source causing current to flow. A common CMOS design combines NMOS and PMOS constructions in twin-well (twin-tub) structures, as shown in Figure 3**b.**

2. Crystal Growth and Wafer Preparation

The single-crystal silicon that is used in IC technology starts with a polycrystalline material called electronic-grade silicon (EGS). Its purity is determined by resistivity measurements made on the test ingot, or by low-ir absorption measurements. EGS is made by starting with a relatively pure form of sand, SiO_2. The sand is melted to form metallurgical-grade silicon (MGS), which reacts with HCl gas to form trichlorosilane [10025-78-2], $SiHCl_3$, gas. The $SiHCl_3$ reacts with H_2 in a chemical vapor deposition (CVD) process to form EGS.

$$2\ SiHCl_3\ (g) + 2\ H_2\ (g) \longrightarrow 2\ Si\ (s) + 6\ HCl\ (g)$$

An alternative process involves pyrolysis of silane [7803-62-5].

$$SiH_4\ (g) + \text{heat} \longrightarrow Si\ (s) + 2\ H_2\ (g)$$

This latter process is lower in cost and has fewer harmful by-products (9).

The polycrystalline EGS is converted to single-crystal silicon via the Czokralski (CZ) crystal growing process, based on the solidification of silicon atoms from the liquid phase at a moving interface. Volume production of 200-mm diameter crystals is standard. Development of crystals having diameters of up to 400 mm has been predicted (3).

The process of growing a pure crystal is sensitive to a host of process parameters that impact the incorporation of impurities in the crystal, the quality of the crystal structure, and the mechanical properties of the crystal rod. For example, the crystal-pulling mechanism controls the pull rate of the crystallization, which affects the incorporation of impurities in the crystal, and the crystal rotation, which affects the crystal structure.

Two common impurities that must be controlled in silicon crystal growth are oxygen and carbon. Oxygen affects the formation of donor regions, the yield strength of the crystal, and the level of defect generation.

Oxygen, as silicon tetraoxide [12359-25-0], SiO_4, can act as a donor and change the resistivity of the silicon. Carbon also contributes to the formation of defects, and comes from graphite parts of the melt furnace.

Ingots of EGS are evaluated for resistivity, crystal perfection, and mechanical and physical properties, such as size and mass. The ingots are sliced into wafers using at least 10 machining and polishing procedures. These wafers are sliced sequentially from the ingot, and evaluated for the correct surface orientation, thickness, taper, and bow. As a final procedure, the wafers are chemically cleaned to remove surface contaminants prior to use.

Defects in the silicon crystals affect the electrical, optical, and mechanical properties. Possible defects include point defects, which affect the kinetics of diffusion and oxidation; line defects, which can affect diffusion; planar defects; and volume defects, where impurities have precipitated and can act as sites for dislocation generation. Integrated circuits are typically constructed on wafers having ⟨100⟩ crystal orientation. Wafers 200-mm in diameter are being fabricated primarily for high volume, large-area circuits such as 16 Mb DRAMs, although 150 mm and smalled-sized wafers form the majority of product substrates. It is possible to fabricate almost twice as many 16 Mb chips on a 200-mm wafer as on a 150-mm wafer; the higher number of chips offsets the additional costs of the more expensive equipment and reduced throughput found in 200-mm processing (3). Important factors in the production of 200-mm wafers include a greater sensitivity to flatness, thermal stress, uniformity, and surface microroughness. Particle contamination control has become essential at the wafer fabrication stage to prevent damage in subsequent process steps. IC manufacturers are requiring very low particle counts from wafer suppliers, on the order of 50 or less particles per square centimeter, 0.1–0.15 μm in diameter (3).

3. Fabrication Processes

3.1. Deposition Processes

3.1.1. Silicon Epitaxy

A critical step in IC fabrication is the epitaxial deposition of silicon on an integrated circuit. Epitaxy is defined as a process whereby a thin crystalline film is grown on a crystalline substrate. Silicon epitaxy is used in bipolar ICs to create a high resistivity layer on a low resistivity substrate. Most epitaxial depositions are done either by chemical vapor deposition (CVD) or by molecular beam epitaxy (MBE) (see Thin films). CVD is the mainstream process.

The CVD process consists of placing silicon wafers into a reactor chamber where process gases are introduced and heated to a high temperature. This induces a series of chemical reactions that result in the deposition of the desired epitaxial layer on the wafer substrates. Any of four gases may be used for silicon epitaxy: silicon tetrachloride [10026-04-7], $SiCl_4$, dichlorosilane [4109-96-0], SiH_2Cl_2, trichlorosilane, $SiHCl_3$, and silane, SiH_4. The most widely used and studied is silicon tetrachloride. The reaction that occurs is a hydrogen reduction.

$$SiCl_4\ (g) + 2\ H_2\ (g) \longrightarrow Si\ (g) + 4\ HCl\ (g)$$

This reaction is not a simple one. There are a number of intermediate chlorosilanes generated by competing reactions (10). The process is sensitive both to the thermodynamics and kinetics of the chemical reactions, and to the fluid mechanics (qv) of the gas flow in the reactor. The overall procedure involves purging the reactor with hydrogen gas, raising the temperature of the reactor, cleaning the wafers with a brief HCl etch, and replacing the HCl with the silicon source gas. A complete process cycle can take up to an hour.

CVD reactors can have one of several configurations. Each has particular advantages and disadvantages. Reactors that support wafers horizontally have difficulty controlling the deposition uniformity over all the

Table 1. Physical Properties of SiO_2 and Si

Property	SiO_2	Si
density, g/cm^3	2.27	2.33
dielectric constant	3.9	11.7
mp, °C	1700[a]	1415
breakdown field, V/μm	600[a]	30[a]
specific heat, J/(g·K)[b]	1.0	0.7
thermal conductivity, W/(m·K)	1.4	150
thermal diffusivity, cm^2/s	0.006	0.9
linear coefficient of expansion, 10^{-6} K	0.5	2.5
etch rate, in buffered[c] HF at 27°C, μm/min	0.10[a]	0[a]
band gap, eV	9	1.11
resistivity, Ω/cm	$>10^{16}$[d]	10^{-3} to 10^5[e]

[a]Approximate value.
[b]To convert J to cal, divide by 4.184.
[c]in 34.6% NH_4F, 6.8% HF, 58.6% H_2O buffer at 27°C
[d]Insulator.
[e]Semiconductor.

exposed wafers. Reactors having vertical wafer support produce uniform deposition, but are mechanically complex. Barrel reactors are not suited for extended operation at temperatures greater than 1200°C.

It is often necessary to introduce dopant atoms into the epitaxial (epi) layers. Typically, the dopant sources are hydrides (qv) of the impurity atoms. Common dopants are boron hydride, ie, diborane(6) [19287-45-7], B_2H_6, for *p*-type doping, and arsine [7784-42-1], AsH_3, and phosphorus hydrides for *n*-type doping (11). For example:

$$2\ AsH_3\ (g) \longrightarrow 2\ As\ (s) + 3\ H_2\ (g) \qquad As\ (s) \longrightarrow 2\ As^+\ (s) + 2e^-$$

Autodoping occurs when dopants are unintentionally released from a substrate through diffusion and evaporation, and subsequently reincorporated during the deposition layer. Epitaxial layers are typically doped at concentrations of $10^{14} - 10^{17}$ atoms/cm^3. The higher levels of doping are used in bipolar technology where the epi layer forms the transistor base. The epitaxial layer can be up to several hundred micrometers, and as thin as 0.05–0.5 μm. Uniformities of $\pm 5\%$ are common.

Molecular beam epitaxy is a non-CVD epitaxial process that deposits silicon through evaporation. MBE is becoming more common as commercial equipment becomes available. In essence, silicon is heated to moderate temperature by an electron beam in a high vacuum ($10^{-6} - 10^{-8}$ Pa ($10^{-8} - 10^{-10}$ Torr)) condition such that the volatile species travels at a relatively high velocity to the substrate wafer. The growth rate is $0.01 - 0.3\ \mu$/min which starts to be competitive with CVD deposition rates.

The MBE process has the disadvantage of being expensive and having low throughput (the number of processed wafers per unit time). The principal advantage of MBE over CVD is that the former uses temperatures in the range of 400–800°C, reduces diffusion from substrate layers and consequent autodoping and this permits more accurate control of doping levels.

3.1.2. Oxidation of Silicon

Silicon dioxide [7631-86-9], SiO_2, is a basic component of IC fabrication. SiO_2 layers are commonly used as selective masks against the implantation or diffusion of dopants into silicon. SiO_2 is also used to isolate one device from another. It is a component of MOS devices, and provides electrical isolation of multilevel metallization structures (12). A comparison of Si and SiO_2 properties is shown in Table 1.

Silicon dioxide layers can be formed using any of several techniques, including thermal oxidation of silicon, wet anodization, CVD, or plasma oxidation. Thermal oxidation is the dominant procedure used in IC fabrication. The oxidation process selected depends on the thickness and properties of the desired oxide layer. Thin oxides are formed in dry oxygen, whereas thick ($>0.5\ \mu m$) oxide layers are formed in a water vapor atmosphere (13).

$$Si\ (s) + O_2 \longrightarrow SiO_2\ (s)$$

$$Si\ (s) + 2\ H_2O \longrightarrow SiO_2\ (s) + 2\ H_2\ (g)$$

The oxidation of surface silicon consumes the surface silicon, and the Si / SiO_2 interface moves down into the bulk silicon as the silicon is converted to the oxide. Oxygen atoms are transported from the gas phase through the oxide layer to the Si / SiO_2 interface where reaction with silicon takes place. Because oxygen atoms are incorporated into the silicon crystal structure, the volume of the resulting oxide is approximately 0.44 times the thickness of original silicon that is consumed. Thermal oxidation of silicon is not practicable when the surface silicon is very thin, as would be found in advanced IC designs. Such cases require the deposition of SiO_2 onto the silicon.

Typical oxide growth takes place at 101 kPa (1 atm) pressure in a horizontal diffusion tube apparatus, although vertical diffusion equipment is also used. Oxidation can be performed at higher pressures, such as 0.5–1 MPa (5–10 atm). The higher pressures permit the use of lower temperatures, which reduces migration of dopants in underlying layers. The growth rate is much higher under these conditions, and is used with advanced MOS applications. Growth rate is also dependent on the crystallographic orientation of the silicon surface because the orientation determines the concentration of the surface silicon atoms. Consequently, ⟨111⟩ silicon oxidizes faster than ⟨100⟩ silicon (14).

Silicon dioxide properties depend on the techniques used for oxide growth. The index of refraction for dry oxides decreases when higher processing temperatures are used whereas the oxide density increases.

Water as an impurity accelerates the oxidation rate. Figure 4 compares growth curves for SiO_2 under dry and steam conditions. Halogens can also be introduced to the oxidation process, thereby reducing sodium ion contamination. This improves dielectric breakdown strength, and reduces interface trap density (15).

In the oxidation process, a layer of dopant is applied to the surface of silicon and patterned silicon dioxide for subsequent thermal diffusion into the silicon. The masking property of the SiO_2 is based on differences in rates of diffusion. Diffusion of dopant into the oxide is much slower than the diffusion into the silicon. Thus, the dopants reach only the silicon substrate. Oxide masks are usually 0.5–0.7 μm thick.

As the width and thickness of IC layers and patterns continue to shrink into the submicrometer range, SiO_2 layers need to be fabricated of 5–20 nm thickness. These thin oxides have properties that are very sensitive to the substrate cleanliness and uniformity, gas purity, and temperature control.

3.1.3. Dielectric Film Deposition

Dielectric films are found in all VLSI circuits to provide insulation between conducting layers, as diffusion and ion implantation (qv) masks, for diffusion from doped oxides, to cap doped films to prevent outdiffusion, and for passivating devices as a measure of protection against external contamination, moisture, and scratches. Properties that define the nature and function of dielectric films are the dielectric constant, the process temperature, and specific fabrication characteristics such as step coverage, gap-filling capabilities, density stress, contamination, thickness uniformity, deposition rate, and moisture resistance (2). Several processes are used to deposit dielectric films including atmospheric pressure CVD (APCVD), low pressure CVD (LPCVD), or plasma-enhanced CVD (PECVD) (see Plasma technology).

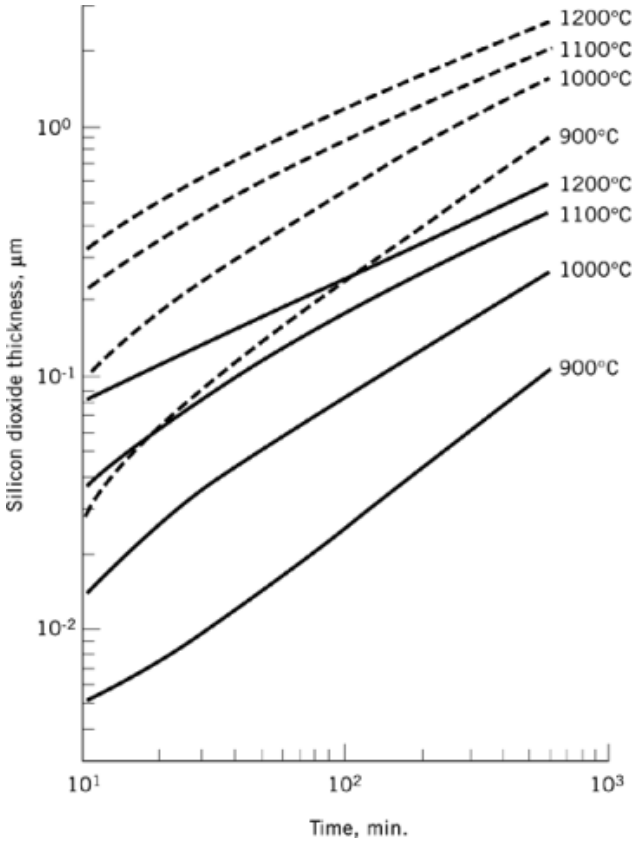

Fig. 4. Silicon dioxide growth rate using a ⟨100⟩ silicon substrate where the solid lines represent a dry oxygen and the dashed lines a steam atmosphere.

The most commonly used dielectric material is SiO_2, which may be deposited by several mechanisms, each distinguished by the specifics of the chemical source, the deposition process, and subsequent applications. Basic SiO_2 can be modified by the addition of boron, phosphorus, or both. SiO_2 and SiO_2-based films can be formed from thermal oxides; silane, SiH_4; tetraethoxysilane (TEOS), $Si(OC_2H_5)_4$; borophosphosilicate glass (BPSG); and spin-on glass (SOG) processes. The majority of CVD SiO_2 films are made from SiH_4 or TEOS.

Gate oxide dielectrics are a crucial element in the down-scaling of *n*- and *p*-channel metal-oxide semiconductor field-effect transistors (MOSFETs) in CMOS technology. Ultrathin dielectric films are required, and the 12.0-nm thick layers are expected to shrink to 6.0 nm by the year 2000 (2). Gate dielectrics have been made by growing thermal oxides, whereas development has turned to the use of oxide/nitride/oxide (ONO) sandwich structures, or to oxynitrides, SiO_xN_y. Oxynitrides are formed by growing thermal oxides in the presence of a nitrogen source such as ammonia or nitrous oxide, N_2O. Oxidation and nitridation are also performed in rapid thermal processors (RTP), which reduce the temperature exposure of a substrate.

Another important use of dielectrics is as intermetal dielectrics (IMDs), where the dielectrics insulate metal lines from each other. The dielectric material must fill small gaps with high aspect ratios (depth to width) while maintaining all other dielectric properties. It is essential that the IMDs are void-free at submicrometer dimensions for both performance and reliability.

Historically, SOG techniques have been used the most for IMD fabrication, but TEOS/ozone ($TEOS/O_3$) processes are more recent developments that have been increasing in popularity based on excellent step coverage and void-free characteristics. $TEOS/O_3$ doped with boron and phosphorus ($BPTEOS/O_3$) has replaced BPSG in small-scale devices, and has been used successfully in 4- and 16-Mb DRAM production (16).

Dopant species can be codeposited with the SiO_2 by introducing small amounts of the dopants in hydride or halide form. P-doped SiO_2, called P-glass, functions as an insulator between polysilicon gates and the top metallization layer of ICs. It is also used as a final passivation layer over devices, and as a gettering source (17).

There are two types of deposited films known as silicon nitride. One is deposited via plasma-enhanced CVD at temperatures $<350^\circ C$ (18). In this process silane and ammonia react in an argon plasma to form silicon imide [14515-04-9], SiNH.

$$SiH_4 + NH_3 \longrightarrow SiNH + 3\ H_2$$

Alternatively, silane can react in a nitrogen discharge.

$$2\ SiH_4 + N_2 \longrightarrow 2\ SiNH + 3\ H_2$$

Plasma-deposited silicon nitride contains large amounts of hydrogen, typically in the range of 20–25 atomic % H, and has polymer-like properties. The electrical resistivity of the film depends on the deposition temperature, the film stoichiometry, and the amounts of hydrogen and oxygen in the film.

A second type of silicon nitride, called stoichiometric silicon nitride, is deposited at much higher temperatures using CVD or LPCVD in the form of Si_3N_4. Stoichiometric silicon nitride can be used as a mask for the selective oxidation of silicon. Here the silicon nitride is patterned over a silicon substrate, and the exposed silicon is oxidized. The silicon nitride oxidizes very slowly compared to the silicon.

3.2. Polysilicon

Polysilicon is used as the gate electrode material in MOS devices, as a conducting material for multilevel metallization, and as contact material for devices having shallow junctions. It is prepared by pyrolyzing silane, SiH_4, at 575–650°C in a low pressure reactor. The temperature of the process affects the properties of the final film. Higher process temperatures increase the deposition rate, but degrade the uniformity of the layer. Lower temperatures may improve the uniformity, but reduce the throughput to an impractical level.

The structure of the polysilicon depends on the dopants, impurities, deposition temperature, and post-deposition heat annealing. Deposition at less than 575°C produces an amorphous structure; deposition higher than 625°C results in a polycrystalline, columnar structure. Heating after deposition induces crystallization and grain growth. Deposition between 600 and 650°C yields a columnar structure having reasonable grain size and ⟨110⟩-preferred orientation.

When required, arsenic, phosphorus, or boron dopants can be added in subsequent steps by diffusion or ion implantation (qv), or by adding dopant gases during deposition. All three methods are used in IC fabrication. The doping elements reduce the resistivity of the polysilicon. The degree of change of resistivities is a function of the deposition temperature, the dopant concentration, and annealing temperature. The dopant concentration affects the etch rate in plasma etching. Oxygen can be used to dope polysilicon to increase its film resistivity, creating a semi-insulating polysilicon (SIPOS) that can be used as a passivating coating for high voltage devices (see Electronics, coatings). SIPOS is formed by depositing silane and nitrous oxide [10102-43-9] together at 600–700°C. A metal or metal silicide, eg, tungsten or tungsten silicide [12039-88-2], may be deposited over the polysilicon to increase its conductivity.

Polysilicon is also oxidized to SiO_2 in dry oxygen at 900–1000°C to form an insulator between the doped polysilicon gate and other conducting layers. Because the surface of the polysilicon is rough relative to the surface of a single-crystal silicon, the SiO_2 grown on polysilicon is itself rough, resulting in lower breakdown fields, higher leakage currents, and higher stresses.

3.2.1. Metallization

Integrated circuits require conductive layers to form electrical connections between contacts on a device, between devices on a chip, between metal layers on a chip, and between chips and higher levels of interconnections needed for packaging the chips. It is critical to the success of IC fabrication that the metallization be stable throughout the process sequence in order to maintain the correct physical and electrical properties of the circuit. It must also be possible to pattern the blanket deposition.

In metal-oxide semiconductor field-effect transistor (MOSFET) devices the control of the on/off function occurs at the gate electrode, which is usually polysilicon. The gate is isolated from the substrate silicon by an insulating layer such as thermally grown SiO_2. The two regions adjacent to the gate are the source and the drain. These features are usually interconnected to first-level aluminum [7429-90-5] metallization through contacts, where the metallization controls the speed of the circuit, and determines the gate-to-source voltage that switches the MOSFET on or off. Vias are small holes in the circuitry that interconnect the first-level metal to the second-level metal, the second to the third, and so on.

Metallization layers are generally deposited either by CVD or by physical vapor deposition methods such as evaporation (qv) or sputtering. In recent years sputter deposition has become the predominant technique for aluminum metallization. Energetic ions are used to bombard a target such as solid aluminum to release atoms that subsequently condense on the desired substrate surface. The quality of the deposited layers depends on the cleanliness and efficiency of the vacuum systems used in the process. The mass deposited per unit area can be calculated using the cosine law of deposition:

$$R_D = M_e/\pi r^2 \cos\phi \cos\theta$$

where R_D is the evaporation rate, M_e is the total mass of the evaporated material, and r, φ, and θ relate the angles between the source and the surface of deposition (19).

A large number of factors can degrade the desired properties of the metallization layer. Impurities in the film, poor adhesion to the substrate, grain size, step coverage, and thickness nonuniformity all reduce the effectiveness of the deposition. Stresses in the layer can occur from differences in TCE between the metal layer and the substrate or from intrinsic stress. Both can result in crack formation. Step coverage can be improved by heating the substrate, optimizing the orientation between the substrate and the source, and using larger-area sources (20).

Aluminum, the most common material used for contacts, is easy to use, has low resistivity, and reduces surface SiO_2 to form interfacial metal-oxide bonds that promote adhesion to the substrate. However, as designs reach submicrometer dimensions, aluminum, Al, has been found to be a poor choice for metallization of contacts and via holes. Al has relatively poor step coverage, which is nonuniform layer thickness when deposited over right-angled geometric features. This leads to keyhole void formation when spaces between features are smaller than 0.7 μm. New collimated sputtering techniques can extend the lower limit of Al use to 0.5-μm applications.

Tungsten CVD (WCVD) is the technique of choice to fill contact and via holes in sub-0.5 μm geometries. Tungsten has extremely good step coverage, good resistance to electromigration, low resistivity, and good adhesion to underlying silicon (in the case of contacts) and aluminum (in the case of vias) (21). The chemistries that are used most frequently in tungsten deposition are mixtures of silane or hydrogen with tungsten hexafluoride, WF_6, where the silane and hydrogen act as reducing agents for the tungsten source. The overall equations are

(22):

$$WF_6 + 3\ H_2 \longrightarrow W + 6\ HF$$

$$WF_6 + 3\ SiH_4 \longrightarrow 2\ W + 3\ SiF_4 + 6\ H_2$$

The blanket deposition is then sputter etched through a resist to pattern the metallization. Selective deposition of W, under development, would deposit metal only in desired areas, and would reduce process steps and costs.

Copper is an attractive metallization element because of its high conductivity. It has been added to Al in low concentrations (AlSi(1%)–Cu(0.5%)) to improve conductive priorities. Selective, low temperature copper CVD processing, using copper(I) β-diketonate compounds, has been carried out (23).

Impurities that can negatively affect the physical and electrical properties of the metallization layer can originate from several sources, particularly the deposition source and the gaseous environment. Impurities stemming from the source bombard the surface of the growing film and get trapped in the metal layer.

Finally, the metallization layer usually requires patterning, which can be done by reactive ion etching (RIE) or back-sputtering. The two processes are similar. In both techniques accelerated ions hit the substrate and forcibly detach atoms or molecules from the surface. RIE uses reactive gases such as chlorine, Cl_2 or trichlorofluoromethane [75-69-4], CCl_3F. Inert gases such as argon or neon are used in back-sputtering.

3.3. Doping of Layers

3.3.1. Diffusion

Another technique for modifying the electrical properties of silicon and silicon-based films involves introducing small amounts of elements having differing electrical compositions, dopants, into substrate layers. Diffusion is commonly used. There are three ways dopants can be diffused into a substrate film: (*1*) the surface can be exposed to a chemical vapor of the dopant at high temperatures, or (*2*) a doped-oxide, or (*3*) an ion-implanted layer can be used. Ion implantation is increasingly becoming the method of choice as the miniaturization of ICs advances. However, diffusion is used in VLSI technology (24, 25).

Theoretical studies of diffusion aim to predict the distribution profile of an exposed substrate given the known process parameters of concentration, temperature, crystal orientation, dopant properties, etc. On an atomic level, diffusion of a dopant in a silicon crystal is caused by the movement of the introduced element that is allowed by the available vacancies or defects in the crystal. Both host atoms and impurity atoms can enter vacancies. Movement of a host atom from one lattice site to a vacancy is called self-diffusion. The same movement by a dopant is called impurity diffusion. If an atom does not form a covalent bond with silicon, the atom can occupy in interstitial site and then subsequently displace a lattice-site atom. This latter movement is believed to be the dominant mechanism for diffusion of the common dopant atoms, P, B, As, and Sb (26).

When the concentration of the solute is low, the measured diffusion profiles are predictable from Fick's second law of diffusion:

$$\frac{\delta C(x,\ t)}{\delta t} = D\ \frac{\delta^2 C(x,\ t)}{\delta x^2}$$

where D is the diffusivity, C is the concentration of solute, x is the coordinate axis in the direction of solute flow, and t is the diffusion time. Diffusivity at low concentration is called intrinsic diffusivity, D_c. When the concentration of the solute is $>N_iT$ where N_i is the intrinsic carrier concentration, and T is temperature in Kelvin, Fick's law as given is insufficient to accurately describe the resulting dopant profile. Additional concentration-dependent diffusivities must be added to the equation. The diffusivities can be determined by experimentally

obtained dopant profiles. At high solute concentrations, the silicon is considered as extrinsic silicon, and the diffusivity is called the extrinsic diffusivity D_e. When arsenic is diffused into silicon at concentrations below $10^{16}/cm^2$ and at temperatures $>1000^\circ C$, nearly all the arsenic ions are ionized and are electrically active. At higher concentrations, and lower diffusion temperatures, the ionized arsenic is only a fraction of the total arsenic. A similar effect is seen for boron (27).

Diffusivities of various elements are determined experimentally. Dopant profiles can be determined. The junction depth can be measured by chemically staining an angle-lapped sample with an HF / HNO_3 mixture. The *p*-type region of the junction stains darker than the *n*-type region. The sheet resistivity can also be measured using a four-point probe measurement. These two techniques are used for process monitoring.

More accurate dopant profile information can be obtained using additional methods. The capacitance–voltage (C–V) method is used to measure the reverse-bias capacitance as a function of the applied voltage. It is common to use Rutherford back-scattering (RBS) to determine the distributions of arsenic, platinum, or gold in silicon. This method cannot, however, be used for measuring boron or phosphorus profiles. Another surface-analytical technique used for these measurements is secondary ion mass spectrometry (sims), which has high sensitivity for many elements including B and As (see Mass spectrometry).

3.3.2. Ion Implantation

Ion implantation (qv) is a technique designed to introduce dopant species into silicon in a controlled, reproducible process. The dopant species are ionized and accelerated toward a substrate surface. The ions have sufficient energy to penetrate the surface and enter the crystal lattice. A series of collisions take place in a cascading effect that affects not just the dopant but also the atoms in the lattice. The depth of penetration, ie, the profile, is controlled by the accelerating energy of the ions. Accelerating energies can range from 1 keV to 1 MeV, allowing penetrations of 100 nm to 10 μm (28). Doping profiles for a bipolar transistor are shown in Figure 5.

Typically, a source gas such as boron trifluoride [7637-07-2], BF_3, is exposed to an ion source that causes the gas to ionize. An analyzer discriminates between all the ionic particles using a magnetic field that can select particles having the correct mass-to-charge ratio to pass through the analyzer to an acceleration tube. The ions are accelerated in the tube and collimated into a beam that is scanned over the substrate wafer. The three primary parameters of any implantation process are the type of dopant species, the accelerating energy used for implantation, and the dose of the source gas. The dose is the total number of ions that enter the wafer. Dose, φ, can be calculated

$$\phi = \frac{1}{QA} \int I dt$$

where Q is the charge on the ionized species; A, the area of the target; and I, the acceleration beams, integrated over time t (29). Doses range from 10^{12} ions/cm^2 for threshold adjustment to 10^{18} ions/cm^2 for buried insulators.

Upon entering a substrate crystal, the implanted ion initiates a series of elastic and inelastic collisions between atomic nuclei and atomic electrons, respectively. The initial collision is between the dopant ion and a silicon lattice atom. The accelerated dopant has sufficient energy to displace the lattice atom, which in turn can displace another lattice atom. An ion entering the wafer and having an initial energy of 100 keV has sufficient energy to initiate a cascade of displacements, because the binding energy of a lattice site is only 10–20 keV.

Channeling is a phenomenon of implantation where impacting ions enter into ordered open spaces, called channels, in the crystal lattice. Channels allow the implanted ion to avoid nuclear collisions and travel further into the lattice than would an ion that collided with atoms in the lattice. Channels can be planar or axial. The primary energy loss in channeled ions is from electron scattering. Channeling can affect the degree of control of the doping profile, especially when dealing with shallow junctions. Consequently, wafers are tilted to avoid channeling, typically at a 7° angle (30).

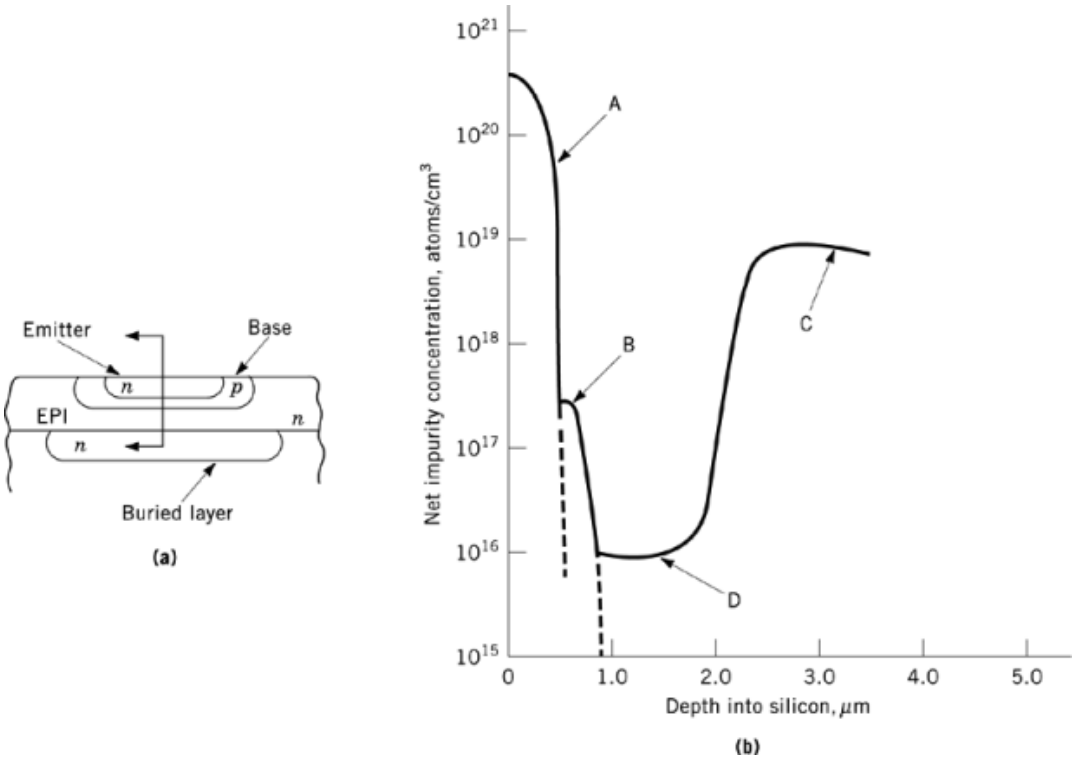

Fig. 5. Bipolar transistor (**a**) schematic and (**b**) doping profiles of A, arsenic ion implanted into the silicon of the emitter (*n*-type); B, boron ion implanted into the silicon of the base (*p*-type); C, antimony ion implanted into the buried layer (*n*-type); and D, the epi layer (*n*-type).

The primary sources of contamination in ion implantation come from metal atoms that may be etched off reactor fixtures, such as reactor walls, wafer holder, clips, and so on. The pump oils used by the vacuum pumps may be a source of hydrocarbon contamination. The dopant sources themselves are not a significant source of contamination because unwanted ions are separated out from the beam during beam analysis.

The implantation of dopant ions into the wafer crystal creates a significant disruption of the lattice structure. An annealing step is performed after implantation to repair the lattice and place the dopant ions into substitutional sites where the dopants can be electrically active with neighboring atoms. Not all of the implanted ions become active, and one primary measure of success of implantation is defined as the fraction of dopant that is electrically active. This is measured experimentally using a Hall effect technique which determines the average effective doping, ie, the integral over local doping densities and local mobilities evaluated per unit surface area

$$N_{\mathrm{Hall}} = \frac{\left(\int_0^{x_j} \mu n dx\right)^2}{\int_0^{x_j} \mu^2 n dx}$$

where u is the mobility; n, the number of carriers; and x_j is the junction depth. If annealing activates all of the implanted atoms, N_{Hall} is then equal to the dose, φ (31).

The annealing must not only repair lattice structure, it must also do so under conditions that minimize diffusion. Furnace annealing is one method that has been used to anneal implanted wafers. In furnace thermal annealing, amorphous silicon regrows by solid-phase epitaxy. An interface exists between the amorphous

and crystalline phases of the silicon, which rises to the surface during the annealing process. The velocity of the movement of the interface depends on the temperature, degree and type of doping, and the crystal orientation. Some impurities such as oxygen, carbon, nitrogen, or argon reduce the recrystallization rate. Boron, phosphorus, and arsenic increase the rate of annealing. Rapid thermal annealing (RTA) is a more recent technology developed for annealing purposes. RTA involves heating the wafer for very short exposures, from 100 down to nanosecond periods. The critical feature of RTA is to minimize diffusion in shallow junctions.

Two newer areas of implantation have been receiving attention and development. Focused ion beams have been investigated to allow very fine control of implantation dimensions. The beams are focused to spot sizes down to 10 nm, and are used to create single lines of ion-implanted patterns without needing to create or use a mask. Although this method has many attractive features, it is hampered by the fact that the patterning is sequential rather than simultaneous, and only one wafer rather than many can be processed at any one time. This limits the production applications of the technique.

Wafer charging is becoming a critical issue as new MOS devices are designed with thinner ($\leq$10 mm) gate dielectrics. Lower energy implanters require optics that compensate for beam divergence which occurs at low energeries (32).

Devices are being designed that have shallow junctions, where the dopants are introduced less than 100 nm beneath the surface. Shallow junctions are formed by using low energies for implantation. A technique being used for submicrometer implantation is large-angle tilt implanted drain (LATID) and source. High (200 keV to 3 MeV) energy implanters are also under development for profile and defect engineering. This implant process is used to produce buried oxide layers, thereby reducing costs by eliminating the need to form seperate epi layers (33).

3.4. Patterning

3.4.1. Lithography

Integrated circuits require that the conducting and nonconducting layers are patterned to form the actual circuits and interconnections needed for a functional device. Typically, the patterns in the functional layers are introduced by first depositing a blanket layer of a polymeric material, called a resist, over the wafer surface. A discrete layer containing the desired pattern information, called the mask, is placed over the resist. The resist is irradiated through the mask with one of several possible media, which include visible or ultraviolet light, electrons, x-rays, or ions. The resist responds to the irradiation by either polymerizing in the exposed regions (negative resists), or preventing polymerization where exposed (positive resist). The unpolymerized portion of the resist is removed in a developer solution. A pattern now exists on the surface of the wafer that can be transferred to the underlying, exposed functional layer by etching the functional layer in the areas not covered by the resist. Alternatively, an additional layer can be selectively deposited into the areas not covered by the resist (see Lithographic resists).

Misregistration between one circuit layer and another is critical for high yield and reliability in IC fabrication. Misregistration can occur at many points during fabrication; for example, when two masks do not align perfectly, or the printer that forms the mask adds to the misalignment. Processing the wafer adds a multitude of possible sources for misregistration, such as etch processing or wafer distortion. The ratio of minimum feature size to registration tolerance is usually in the range of 3–5, permitting high precision of micrometer or submicrometer features. If the contributions are independent, then the total registration error, Δ_t, is given by

$$\Delta_t = \left(\sum_i \Delta_i^2 \right)^{1/2}$$

where Δ_i is an individual error.

3.4.1.1. Optical Lithography. Optical lithography uses visible or ultraviolet light as the exposure media, and is the dominant lithographic process used for patterning IC wafers. The linewidth limit is near 0.4 μm, although some narrower features may be possible (34). The masks typically are made from patterned, opaque chromium films on glass.

Two types of resists are used: negative and positive resists. Negative resists have a chemically inert polymer component, such as rubber, and a photoreactive agent that reacts with light to form cross-links in the rubber. When placed in an organic developer solvent, the unexposed, unpolymerized resist dissolves, leaving a polymeric pattern in the exposed regions. Because the polymer swells in the solvent, the resolution is limited to two to three times the film thickness. Swelling is a critically limiting factor in all negative resists, limiting line resolution. The swelling between two closely spaced lines of resists can create a bridge between the lines, and weaken the adhesion of the resist line to the substrate. An example of a patterning process using a negative resist is shown in Figure 6.

Positive resists have as the photoreactive component a dissolution inhibitor that is destroyed in the regions exposed to the light. The resist is developed in an aqueous solution, where the exposed region dissolves away. The resists do not swell as much in the aqueous developer, allowing higher resolution.

Masks can be placed directly on the resist surface for contact printing allowing submicrometer line resolution. But this procedure may have distortions develop if there are any surface irregularities in the mask or substrate. Contact printing can also cause defects in the mask itself from the physical contact between the mask and the wafer. A second printing process, called projection printing, holds the mask away from the wafer. Projection printers can process 13–50 wafers per hour, depending on the size of the wafer. The different printing methods are shown in Figure 7.

The use of optical lithography for submicrometer resolutions has been extended by developments such as phase-shift lithography (35). A phase-shift reticle (mask) is one which has a patterned layer added to the conventional chrome-on-glass lithographic reticle. Light passing through the glass layer emerges with a different phase than light passing through the phase layer, resulting in an interference pattern that heightens resolution and depth of focus. The next generation of optical lithography is projected to use deep-uv (248 nm) irradiation to achieve 0.20–0.15 μm resolution (32), which may be a necessary but very expensive transition. Manufacturable deep-uv resist technology requires further development (36).

3.4.1.2. Electron Lithography. It is possible to achieve better resolution using electron rather than optical lithography because of the small wavelength of 10–50 keV electrons. Although e-beam direct write systems have always shown great potential for producing fine patterns, production applications have been limited by low throughput to low volume applications such as mask generation (37).

Electron lithography uses both positive and negative resists as does optical lithography. The electron beam induces cross-links between molecules in negative resists, making the exposed areas less soluble in developer. In positive resists, the electron beam causes bond scission, resulting in lower solubility of the resist in the exposed areas when placed in developer. Polymethyl methacrylate (PMMA) offers the highest resolution, and is used both in optical and electron lithography.

Scanning electron beam systems are available commercially, and are commonly used for mask generation. Electron projection systems are also used to obtain resolution over a large field. Current cathode sources have a short lifetime, limiting use in production processes.

3.4.1.3. X-Ray Lithography. The resists used in x-ray lithography are the same as those used in electron lithography, because an x-ray resist is exposed predominantly by the photoelectrons produced during x-ray absorption. The photoelectrons are of much lower (0.3–3 keV) energies than those used in electron lithography (10–50 keV) permitting much higher resolution of the mask image. A 1-μm thick hydrocarbon resist having density of 1 g/cm^3 absorbs only about 10% of the x-ray flux at the $Al_{K\alpha}$ wavelength 0.83 nm. This small absorption yields uniform exposure through the resist thickness. Gold is the most commonly used material for the mask, and is patterned by electroplating (qv) or ion milling (38). However, x-ray lithography has yet to be developed for production applications. It is not considered as cost-effective as deep-uv or phase-shift

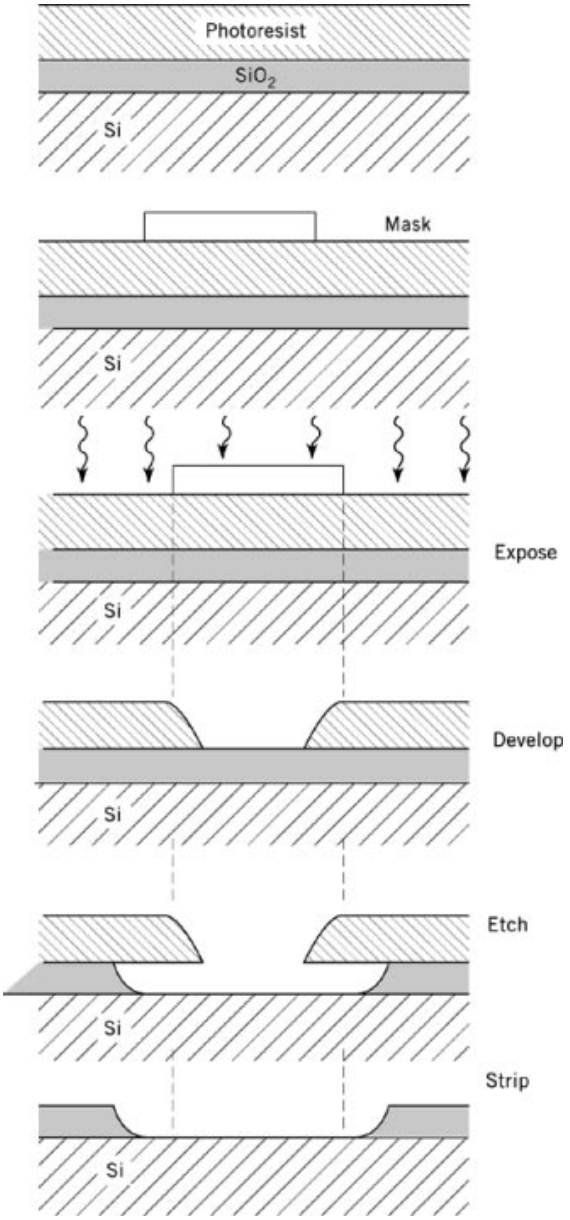

Fig. 6. Schematic illustration of the photolithographic patterning process used for defining features in silicon dioxide using (▨) a positive photoresist that polymerizes light, where (□) represents the mask; (▥) SiO_2; and (▨) Si. Development includes removal of the mask and undeveloped photoresist.

lithography. X-ray lithography is projected to see broader application toward the end of the 1990s as 1 Gb DRAM technology is predicted to require 0.2–18 μm resolutions (37).

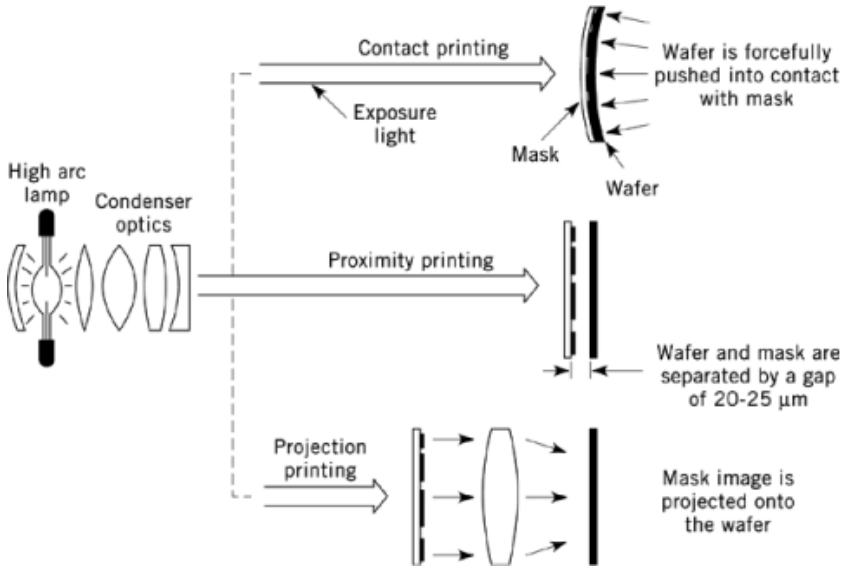

Fig. 7. Schematic illustration of different photolithographic exposure techniques.

3.4.2. Etching

After a resist is patterned on a wafer, the exposed or unwanted substrate is removed by etching processes. Subsequently the resist is removed, leaving a desired pattern in a functional layer of the integrated circuit. Etching is performed to pattern a number of materials in the IC fabrication process, including blanket polysilicon, metal layers, and oxide and nitride layers. The etch process for each material is different, and adapted to the material requirements of the substrate.

The goal of etching processes is to achieve precisely patterned geometries having vertical profiles by removing unwanted material from a substrate or blanket layer inflicting minimal damage to the remaining substrate. The desired etch characteristics are high etch rate, anisotropy (vertical profiles), uniformity, and high selectivity for the material that is to be removed over the material that acts as a resist. An example of isotropic etching is depicted schematically in Figure 8**a,** showing the undercutting that results from etching the exposed surfaces in all directions. The desired anisotropic etch is sketched in Figure 8**b,** showing the vertical profiles of the sidewalls after etching occurs only in the vertical direction.

Etching is performed either by plasma or reactive ions. In plasma etching, a plasma is generated by applying a d-c potential across two conducting electrodes in the presence of a neutral gas. The gas becomes ionized into electrons, singly or multiply charged ions, and neutral atoms, molecules, and molecular fragments. The plasma species physically or chemically interact with the solid surfaces forming volatile products that are removed from the wafer surface.

The physical removal of surface material is called sputtering, where energetic, chemically inert ions such as Ar^+ or Xe^+ are accelerated toward the wafer and physically eject material from the surface. The yield is defined as the ratio of the number of ejected surface atoms to the number of incoming ions per given ion energy.

Chemical removal of surface material is produced through standard bond-breaking reactions. Typically chlorofluorocarbons (CFCs) have been used, eg, $CFCl_3$, CF_2Cl_2, CF_3Cl, CF_4, CHF_3, C_2ClF_5. For example, CF_4 dissociates into F atoms and fluorinated fragments of CF_x in a plasma:

$$CF_4 \rightleftharpoons CF_x + F$$

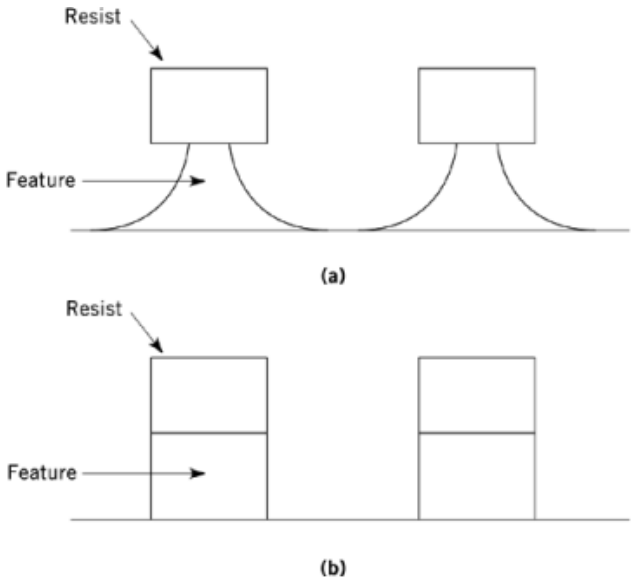

Fig. 8. Schematic of etching directionality showing (**a**) isotropic etch, and (**b**) anisotropic etch.

where $x \leq 3$. Oxygen is added to the plasma mix to react with the CF_x molecules and drive the reaction to the right. In the case of etching Si or SiO_2, F is the active etchant species, reacting with Si and SiO_2 to form volatile SiF_4 (39).

$$Si + 4\ F \longrightarrow SiF_4$$

$$SiO_2 + 4\ F \longrightarrow SiF_4 + O_2$$

A critical feature of plasma etching is managing the directionality of the etching reaction, relative to the resist and the surface layer. Isotropic etching occurs when the etching reactions are equal in all directions on the surface producing an effect called undercutting, as shown in Figure 8**a.** Undercutting is not a problem in patterns having feature height, where the degree of undercutting can be compensated for by making the spacing between the resist narrower to obtain the desired final spacing in the functional layer. Anisotropic etch occurs when the attack of the etching molecules is faster in the vertical direction than in the horizontal direction. This results in deeper channels relative to the width, and straighter wall profiles (Fig. 8**b**). This feature is of primary importance in VLSI fabrication, where narrow spacing is required.

Reactive ion etching (RIE) is an etch process whereby the wafers are held by the radio frequency-driven electrode instead of the grounded electrode, which becomes the chamber itself. The larger grounded area, in combination with lower operating pressures, leads to significantly higher energy bombardment. An advantage of this system is that wafer-to-wafer etching nonuniformities are reduced.

Changes in technology of plasma etching have been driven by two principal forces. The first is the scaling to submicrometer geometries, which has impacted on process design, gas chemistries, and plasma sources. In conventional plasma chemistries, the anisotropy of vertical etching was achieved by adding halocarbons such as CCl_4, CF_4, $CHCl_3$, or CHF_3 to the primary enchant gas to form passivating layers that would protect the

Table 2. Trends in Plasma Etch Chemistry[a]

Material being etched	Conventional chemistry		Replacement chemistry
polysilicon	Cl_2 or BCl_3/CCl_4[b]	$SiCl_4/Cl_2$	no carbon contamination
	$/CF_4$[b]	BCl_3/Cl_2	
	$/CHCl_3$[b]	$HBr/Cl_2/O_2$	increased se-lectivity to SiO_2 and resist
	$/CHF_3$[b]	HBr/O_2	no carbon contamination
		Br_2/SF_6	
		SF_6	higher etch rate
		CF_4	
Al	Cl_2[c]	$SiCl_4/Cl_2$	better profile control
	BCl_3[c]	BCl_3/Cl_2	no carbon contamination
	$SiCl_4$[c]	HBr/Cl_2	
Al–Si(1%)–Cu (0.5%)	same as Al	$BCl_3/Cl + N_2$	N_2 accelerates Cu etch rate
W	$SF_6/Cl_2/CCl_4$	SF_6 only	no carbon contamination
		NF/Cl_2	etch stop over TiW and TiN
single-crystal Si	Cl_2 or BCl_3[c]	CF_3BR	higher selectivity trench etch
		HBr/NF_3	
SiO_2(BPSG)	CCl_2F_2	CCl_2F	all are CFC alternatives
	CF_4	CHF_3/CF_4	
	C_2F_6	CHF_3/O_2	
	C_3F_8	CH_3CHF_2	
Si_3N_4	CCl_2F_2	CF_4/O_2	all are CFC alternatives
	CHF_3	CF_4/H_2	
		CHF_3	
		CH_3CHF_2	

[a]Ref. 45.
[b]Sidewall passivating gas.
[c]Plus sidewall passivating gases.

sidewalls from etching. The typically 0.1-μm thick passivating layers are a significant liability when feature size is less than one micrometer and impossible in under 0.5-μm designs. Engineers having submicrometer etch requirements are turning to single-wafer etchers with precise temperature controls that circumvent the need for passivating layers (40).

Submicrometer etching requires gas chemistries having much greater selectivity and high etch rates. One solution is to use multistep processing, where a number of etch steps are used to etch through one layer. It is possible to have 10–15 procedures to etch through two to three layers. In polysilicon etching, the etch gases must have sufficient selectivity to etch stop over 20-nm thick (or thinner) oxide layers. Polysilicon etch gases have a selectivity to oxide in the range of 40:1 to 100:1 (40, 41). Although bromine-based gases can provide the required selectivity, corrosiveness has prevented widespread use of these gases.

Plasma sources are also being introduced to produce plasmas at lower pressures and process temperatures. Inductively coupled plasma (ICP) and transformer-coupled plasma (TCP) are among the more commonly used sources, operating below 2.6 Pa (20 mTorr) (42). Low temperature RIE processing operates between 26–67 Pa (200–500 mTorr).

A second force that is seriously affecting etch technology is the move to actively eliminate CFC gases to comply with the Montreal Protocol for protection of the ozone layer. These gases, eg, $CFCl_3$, CCl_2F_2, CF_3Cl, and CF_3Br, are being replaced by fluorine-based gases for tungsten etch, such as SF_6, NF_3, and SiF_4;chlorine-based gases for Al etch, eg, Cl_2, BCl_3, and $SiCl_4$; and bromine-based gases, Br_2, and HBr. Table 2 lists conventional and newer chemistries.

Plasma etching can create a variety of damaging effects to a substrate. Typical damage effects include gate oxide breakdown, high reverse leakage current, low minority carrier lifetime, contamination, damage to the

silicon surface charge, and lattice damage (43). The sources of the damage can be typically attributed to either or both of two effects: current flow-induced damage and plasma exposure damage. The first affects primarily the dielectric layers, where voltage across the dielectric produces wear-out from bond-breaking, trapping, or both. The second is a side effect of particle or photon flux impingement on the substrate materials (44). The regions that are most vulnerable to physical damage include the area in the oxide layer, at the layer interfaces, and in the silicon substrate. As of 1994, damage detection, measurement, and analysis are not adequate for submicrometer designs.

3.5. Process Integration

The fabrication of a VLSI i ntegrated circuit involves sequential processing steps. Step interrelationships must be considered in designing a process sequence. The registration of one layer to another layer already present is repeated frequently with resist patterning followed by etching or implantation. Process temperature must be compatible with the condition of the substrate, such that a high temperature process does not damage an underlying layer or destroy a precision dopant profile.

Figure 9 shows a simplified fabrication sequence for an oxide-isolated p-well CMOS process that illustrates many of the essential steps used in IC manufacture. These steps are as follows:

Step 1. Starting with a lightly doped n-type substrate, a thin blanket layer of SiO_2 (the pad oxide) is formed, and a blanket deposition of a thick protecting Si_3N_4 layer follows.

Step 2. A resist is deposited on the Si_3N_4 and patter (mask 1). All Si_3N_4 and SiO_2 not covered by the resist (the resist covers the final transistor area) is etched away.

Step 3. Boron is ion implanted around the perimeter of the resist-protected area to form a p^+-type isolation border (the channel stopper or chanstop). The boron cannot penetrate through the resist.

Step 4. A thick SiO_2 layer (the field oxide) is grown over the p^+-chanstop to isolate the device. This also drives the p^+-region deeper into the substrate.

Step 5. The remaining Si_3N_4, pad oxide, and resist are stripped away and a thin, precisely controlled SiO_2 gate oxide layer is thermally grown.

Step 6. The doping concentration of the p^+-type substrate under the gate oxide is adjusted by another boron implantation. Boron passes through the thin gate oxide. This provides the threshold voltage adjustment for the final device.

Step 7. A blanket layer of polysilicon and pattern (mask 2), such that the resist covers only the polysilicon that is to become the gate, is deposited. All exposed polysilicon is etched away.

Step 8. The n^+-type source and drain regions are created by As ion implantation. The As can penetrate the thin gate oxide, but not the thick field oxide or the polysilicon gate. The formation of the source and gate does not require a separate resist pattern, thus this technique is called self-aligning.

Step 9. SiO_2 is blanket deposited over the substrate. The resist (mask 3) that has openings over the SiO_2 is deposited and patterned. The exposed SiO_2 is etched down to the source, drain, and gate layers, creating contact windows for metallization.

Step 10. The system is metallized, first with a tungsten layer, then with Al. The resist is applied and patterned (mask 4), and unwanted metal is etched away.

Step 11. If no additional metallization layers are required, the substrate is covered with a passivation layer. If additional levels of metallization are to be added to the structure, a blanket layer of a intermetal dielectric (IMD) is deposited. The resist is deposited, patterned (mask 5), and vias down to the Al in the first metal layer are etched. Steps 10 and 11 are repeated to form the second metal layer.

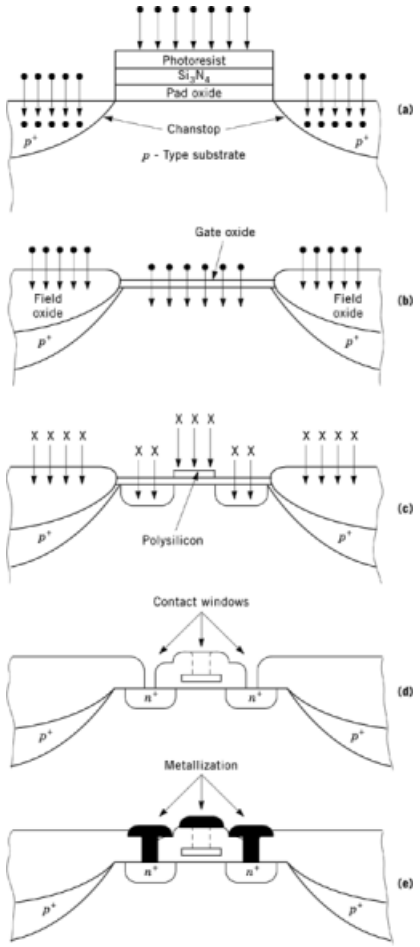

Fig. 9. Fabrication sequence for an oxide-isolated *p*-well CMOS process, where • is boron and X is arsenic. See text. (**a**) Formation of blanket pod oxide and Si_3N_4 layer; resist patterning (mask 1); ion implantation of channel stoppers (chanstop) (steps 1–3). (**b**) Growth of isolation field oxide; removal of resist, Si_3N_4, and pod oxide; growth of thin ($_{<200\text{nm}}$) SiO_2 gate oxide layer (steps 4–6). (**c**) Deposition and patterning of polysilicon gate; formation of $_{n^+}$-source and drain (steps 7,8). (**d**) Deposition of thick SiO_2 blanket layer; etch to form contact windows down to source, drain, and gate (step 9). (**e**) Metallization of contact windows with W; blanket deposition of Al; patterning of metal (steps 10,11). The deposition of intermetal dielectric or final passivation layer is not shown.

These processes are considerably more complex in actual CMOS fabrication. First, the lower layers of a CMOS structure typically have a twin-tub design which includes both PMOS and NMOS devices adjacent to each other (see Fig. 3**b**). After step 1, a mask is opened such that a wide area is implanted to form the *n*-well, followed by a similar procedure to create the *p*-well. Isolation between active areas is commonly provided by local oxidation of silicon (LOCOS), which creates a thick field oxide. A narrow strip of lightly doped drain (LDD) is formed under the edges of the gate to prevent hot-carrier induced instabilities. Passivation sidewalls are used as etch resists. A complete sequence of fabrication from wafer to packaged unit is shown in Figure 10.

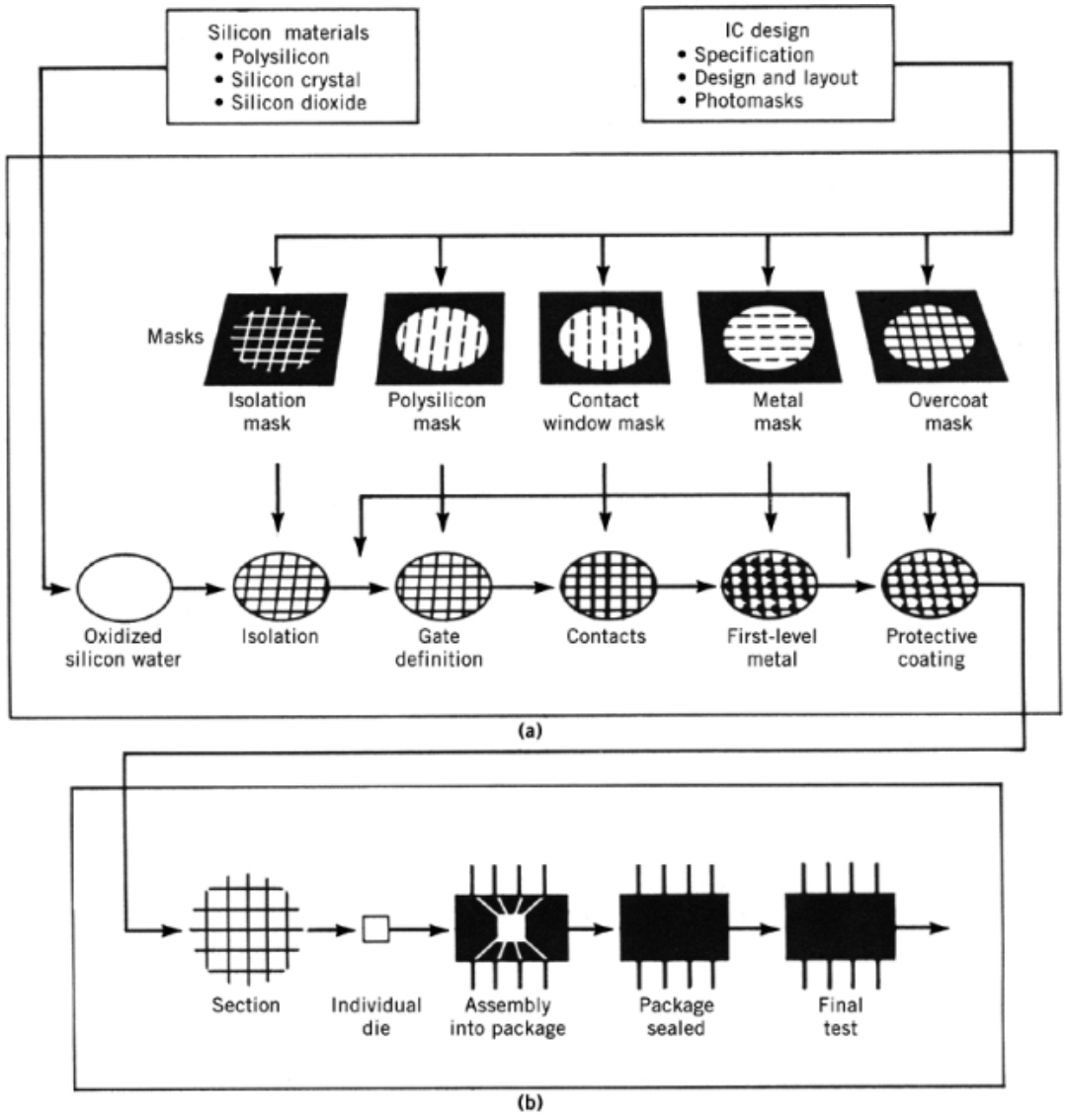

Fig. 10. Complete fabrication sequence for manufacturing a moderately complex silicon device. (**a**) Front end processing, and (**b**) assembly and test.

3.6. Processing Facilities

Two factors essential to successful IC processing are the quality of the process water and the ambient air. Use of ultrapure filtration (see Ultrafiltration) and clean-room facilities addresses these concerns. In addition, three areas of facilities management are undergoing intense development to support the production of submicrometer VLSI and ULSI designs. The first is the area of particle control, where the thrust is to reduce even further the typical allowance of 0.05 particles/cm^2 that are less than 0.35 μm in size (44). There are two approaches to reduce the particle count: prevent particle generation, and prevent particles from depositing on wafers. Both

require *in situ* particle characterization technology that is more advanced than what is available as of this writing, where particles are monitored primarily in the vacuum exhaust yielding only secondary information about the cleanliness of a process (32). Particle contamination is most prevalent during rapid changes in the environment of a wafer, such as during vacuum transition (4). Face-down wafer processing has been introduced based on the observation that fewer particulates settle on a wafer surface positioned this way (46). Room air ionizers have also been found to reduce particle counts (47).

The minienvironment approach to contamination control has been increasing in use. A minienvironment is a localized environment created by an enclosure that isolates the product wafer from contamination and people (48). Another approach is using integrated processing, where consecutive processes are linked in a controlled environment (32). Both require *in situ* sensors (qv) to measure internal chamber temperatures, background contamination, gas flow rates, pressure changes, and particularly wafer temperature (4).

A second area of development that has impacted facility design is the trend to single-wafer processing, allowing enhanced control in processing individual wafers. This should carry greater importance as wafer size goes beyond 200-mm diameter to 300–400 mm.

A third facet of facilities development receiving much attention is the concept of total cost of ownership in evaluating process technology and equipment. Cost of ownership goes beyond simple comparison of technological capability and throughput to consider all aspects of equipment and processing costs, including capital costs, final yields, downtime, maintenance (qv) and repair, cost of consumables, and clean-room requirements.

4. Analytical Techniques

Analytical methodology has had to respond to the rapid scaling of IC designs to submicrometer geometries and to the transition to wafers that are up to 200 mm and larger in diameter. The key requirements that have emerged are not only the need to probe 0.5 μm or smaller features, but also to characterize thin films (qv) that are less than 5 nm thick, detect surface metals present at less than 10^{10} atoms/cm^3, and identify trace organic contaminants (see Trace and residue analysis) (49). These capabilities are essential in process development, process control (qv), and failure mode analysis (FMA), for all stages of fabrication. The purity of gas and liquid starting materials must also be determined (see also Nondestructive evaluation; Surface and interface analysis).

The physical techniques used in IC analysis all employ some type of primary analytical beam to irradiate a substrate and interact with the substrate's physical or chemical properties, producing a secondary effect that is measured and interpreted. The three most commonly used analytical beams are electron, ion, and photon x-ray beams. Each combination of primary irradiation and secondary effect defines a specific analytical technique. The IC substrate properties that are most frequently analyzed include size, elemental and compositional identification, topology, morphology, lateral and depth resolution of surface features or implantation profiles, and film thickness and conformance. A summary of commonly used analytical techniques for VLSI technology can be found in Table 3.

4.1. Electron Beam Techniques

One of the most powerful tools in VLSI technology is the scanning electron microscope (sem) (see Microscopy). A sem is typically used in three modes: secondary electron detection, back-scattered electron detection, and x-ray fluorescence (xrf). All three techniques can be used for nondestructive analysis of a VLSI wafer, where the sample does not have to be destroyed for sample preparation or by analysis, if the sem is equipped to accept large wafer-sized samples and the electron beam is used at low (ca 1 keV) energy to preserve the functional integrity of the circuitry. Samples that do not diffuse the charge produced by the electron beam, such as insulators, require special sample preparation.

Table 3. Analytical Techniques Used in VLSI Technology

Energy range, keV	Secondary signal	Acronym	Technique	Application
			Electron beam	
0.020–0.200	electron	leed	low energy diffraction	surface structure
0.300–30	electron	sem	scanning electron microscope	surface morphology
1–30	x-ray	emp	electron microprobe	surface region composition
500–10	electron	aes	Auger spectroscopy	surface layer composition
100–400	electron	tem	transmission electron microscopy	high resolution structure
100–400	electron, x-ray	stem	scanning TEM	imaging, x-ray analysis
100–400	electron	eels	electron energy loss spetroscopy	local small area composition
			Ion beam	
0.5–2.0	ion	iss	ion scattering spectrometry	surface composition
1–15	ion	sims	secondary ion mass spectrometry	trace composition vs depth
1–15	atoms	snms	secondary neutral mass spectrometry	trace composition vs depth
$\geq$1	x-ray	pixe	particle induced x-ray emission	trace composition
5–20	electron	sim	scanning ion microscope	surface characterization
$>1,000$	ion	rbs	Rutherford back-scattering	composition vs depth
			Photon beam	
$>$1	x-ray	xrf	x-ray fluorescence	composition (μm depth)
$>$1	x-ray	xrd	x-ray diffraction	crystal structure
$>$1	electron	esca.xps	x-ray photoelectron spectroscopy	surface composition
laser	ion		laser microprobe	composition of irradiated area
laser	light	lem	laser emission microprobe	trace elements (semiquantative)
			Neutron beam	
reactor	gamma	naa	neutron activation analysis	bulk (trace) composition

In the secondary electron mode, a 1–20 keV electron beam is rastered across a surface, causing low ($<$50 eV) energy electrons to be emitted from the surface to produce a high magnification, high resolution image of a surface. The wide depth of field yields three-dimensional images that are focused even with a wide variation of surface feature heights. The back-scattered mode is used to obtain images with a better contrast between elements of differing atomic number. Sems are also typically equipped for xrf analysis, where a primary x-ray beam generates fluorescent x-rays that are analyzed to qualitatively or semiquantitatively identify the elemental composition of the surface, or to map the distribution of elements in the surface.

A state-of-the-art sem uses a field-emission (fe) electron gun (fesem) to obtain 0.7-nm lateral resolution, equivalent to the resolution obtained by transmission electron microscopy (tem) but with easier and quicker sample preparation (49). Dramatic improvements in image quality have resulted from the development of immersion lenses; 1.0-nm resolution is possible using small (2 mm) samples in off-line inspection (50). High (50–200 keV) energy beams are used to take three-dimensional measurements of high aspect ratio (up to 10:1) contact holes having good depth of field focus. These high energy sources have a sampling depth of up to 20 μm, and can yield images of layers that are covered by other layers, useful for nondestructive detection of voids in underlying films. Lastly, sems are used in nondestructive, on-line, critical dimension (CD) metrology control, where precision as well as speed are required for monitoring dimensional tolerances during processing.

Auger spectroscopy uses a primary electron beam to generate valence-shell electrons that are analyzed for elemental identification and compositional analysis. The two distinguishing features of Auger spectroscopy are that (*1*) it is very surface sensitive with small spatial resolution. The surface sensitivity stems from the shallow escape depth (0.5–1.0 μm) of the detected electrons, resulting in the characterization of only the uppermost 1–10 monolayers of the surface. (*2*) In general, Auger spectroscopy can provide quantitative analyses of films

as thin as 1.5 nm in areas as small as 15 nm, detecting oxygen and carbon down to 0.1 atomic %. The beam can also be rastered over the surface to provide compositional maps.

Auger spectroscopy is frequently combined with sputter etching to reveal the composition of a surface as a function of depth (depth profile). Used in analysis of ion implantation as well as other applications, the sputter etching removes surface layers to expose underlying layers to the Auger beam. A tenfold improvement in depth resolution has been obtained with the development of the Zalar rotation stage, which reduces crater bottom roughening induced by the ion beam bombardment by rotating the sample during etching (49).

Transmission electron microscopy (tem) is used to analyze the structure of crystals, such as distinguishing between amorphous silicon dioxide and crystalline quartz. The technique is based on the phenomenon that crystalline materials are ordered arrays that scatter waves coherently. A crystalline material diffracts a beam in such a way that discrete spots can be detected on a photographic plate, whereas an amorphous substrate produces diffuse rings. Tem is also used in an imaging mode to produce images of substrate grain structures. Tem requires samples that are very thin (10–50 nm) sections, and is a destructive as well as time-consuming method of analysis.

4.2. X-Ray Radiation

After xrf, the most common method of surface analysis utilizing x-rays as the primary beam is x-ray photoelectron spectroscopy (xps), also known as electron spectroscopy for chemical analysis (esca). The x-rays generate photoelectrons emitted with energies that are equal to the difference between the incident photon energy and the binding energy of the electron, and are characteristic for a given element. Xps is very surface-sensitive, similar to Auger analysis, and is particularly useful in analyzing insulators and surface hydrocarbon contaminations. Xps is also used to obtain information about the chemical structure of a film, as the binding energy of an electron is affected by the electronegativity of the surrounding bonded atoms. For example, the carbon 1*s* signal varies in energy maximum and peak shape depending on the electronegativity of the surrounding bonding atoms. A shortcoming of the technology is the large spot size of the x-ray beam, limiting the lateral resolution of the technique; newer instruments have beam sizes of ca 20 nm.

4.3. Ion Beam Techniques

Secondary-ion mass spectroscopy (sims) uses ion beams having high enough energies to penetrate the surface and break surface bonds, ejecting neutral and ionic species from the surface in a process called sputtering. The primary beam is typically O^+_2. The ejected secondary ions are analyzed and identified according to mass. The sensitivities of ejection, or the ratio of ejected ions to atoms present in the substrate varies greatly according to the particular element, the substrate chemistry, or the substrate structure. The principal advantage of sims analysis is in its very low detection limits, which are applied to the analysis of doping profiles and the detection and identification of surface contaminants. Time-of-flight secondary mass spectrometry (tof-sims) is a new surface-sensitive technique that analyzes both organic and inorganic contaminants in the top monolayer of a surface at ppm detection limits (49). Sims is a destructive analytical process, and requires a large surface area for analysis (5×5 mm) (51) (see Mass spectrometry).

Rutherford back-scattering (RBS) is used to determine the composition and distribution of heavy elements in thin films composed of light elements. $^4He^+$ ions having energies >1 MeV are accelerated toward a surface, penetrate the surface, undergo elastic collisions with substrate nuclei, and back-scatter out of the surface for detection and analysis. The energy distributions of these ions are characteristic of the elements in the substrate and the depth within the surface where the collisions occur. RBS analysis is nondestructive, and requires a large sample size of several mm. Using hydrogen forward scattering (hfs), it is possible to quantitatively profile hydrogen concentrations as low as 0.1 atomic % to $\pm 10\%$ accuracy on both conducting and insulating samples (49).

Newer techniques that are responding to the need for atomic level imaging and chemical analysis include scanning tunneling microscopes (STMs), atomic force microscopes (AFMs) (52), and focused ion beams (FIBs). These are expected to quickly pass from laboratory-scale use to in-line monitoring applications for 200-mm wafers (32).

BIBLIOGRAPHY

"Integrated Circuits" in *ECT* 3rd ed., Vol. 13, pp. 621–648, by E. F. Labuda and J. T. Clemens, Bell Telephone Laboratories, Inc.

Cited Publications

1. C. Hu, *Semiconductor Int.*, 105–114 (June 1994).
2. P. Singer, *Semiconductor Int.*, 56–59 (Apr. 1994).
3. P. Burggraaf, *Semiconductor Int.*, 48–50 (Oct. 1992).
4. M. Moslehi and co-workers, *Solid State Technol.*, 35–45 (Jan. 1994).
5. W. Knausenberger, presentation at the *IPC Technology Marketing Research Council*, New Orleans, La., Dec. 1993.
6. R. Iscoff, *Semiconductor Int.*, 60–66 (Mar. 1992).
7. P. H. Singer, *Semiconductor Int.*, 56–60 (Apr. 1992).
8. P. H. Singer, *Semiconductor Int.*, 44–48 (Jan. 1993).
9. C. W. Pearce, in S. M. Sze, ed., *VLSI Technology*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1988, p. 10.
10. Ref. 9, p. 59.
11. Ref. 9, p. 63.
12. L. E. Katz, in Ref. 9, p. 125.
13. *Ibid*., p. 99.
14. *Ibid*., 108–109.
15. *Ibid*., p. 113.
16. H. W. Fry and co-workers, *Solid State Technol.*, 31–40 (Mar. 1994).
17. A. C. Adams, in Ref. 9, p. 249.
18. R. J. Schultz, in Ref. 9, p. 225.
19. S. P. Muraka, in Ref. 9, p. 375.
20. *Ibid*., p. 400.
21. P. Singer, *Semiconductor Int.*, 57–64 (Aug. 1994).
22. P. Singer, *Semiconductor Int.*, 36–39 (Aug. 1992).
23. A. Jain and co-workers, *Semiconductor Int.*, 128–131 (June 1993).
24. J. C. C. Tsai, in Ref. 9, p. 272.
25. S. L. Hurst, *Custom VLSI Microelectronics*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1992, p. 4.
26. Ref. 24, p. 275.
27. Ref. 24, p. 293.
28. M. D. Giles, in Ref. 9, p. 327.
29. *Ibid*., p. 349.
30. *Ibid*., p. 345.
31. *Ibid*., p. 355.
32. Editorial board, *Semiconductor Int.*, 31–40 (Jan. 1993).
33. L. Peters, *Semiconductor Int.*, 82–87 (June 1993).
34. R. K. Watts, in Ref. 9, p. 142.
35. P. Burggraaf, *Semiconductor Int.*, 42–47 (Feb. 1992).
36. P. Burggraaf, *Semiconductor Int.*, 56–59 (Sept. 1994).
37. P. Burggraaf, *Semiconductor Int.*, 52–56 (Mar. 1992).
38. Ref. 34, p. 174.

39. R. J. Schultz, in Ref. 9, p. 193.
40. L. Peters, *Semiconductor Int.*, 66–70 (May 1992).
41. P. Burggraaf, *Semiconductor Int.*, 56–60 (May 1994).
42. P. Singer, *Semiconductor Int.*, 50–54 (Apr. 1993).
43. P. Singer, *Semiconductor Int.*, 78–81 (May 1992).
44. S. J. Fonash, C. R. Viswanathan, and Y. D. Chan, *Solid State Technol.*, 99–107 (July 1994).
45. L. Peters, *Semiconductor Int.*, 66–70 (May 1992).
46. A. Bowling and C. J. Davis, *Solid State Technol.*, 45–50 (Feb. 1994).
47. A. Steinman, *Semiconductor Int.*, 73–75 (Sept. 1994).
48. X. Pucel, *Semiconductor Int.*, 69–74 (Apr. 1994).
49. P. K. Chu and R. S. Hockett, *Semiconductor Int.*, 142–146 (June 1994).
50. L. Peters, *Semiconductor Int.*, 50–53 (Mar. 1994).
51. J. B. Bindell, in Ref. 9, p. 545.
52. B. S. Meyerson, *Solid State Technol.*, 53–57 (Feb. 1994).

IRENE SAWCHYN PH.D.
Consultant

Related Articles

Silicon and silicon alloys; Computer technology; Microwave technology