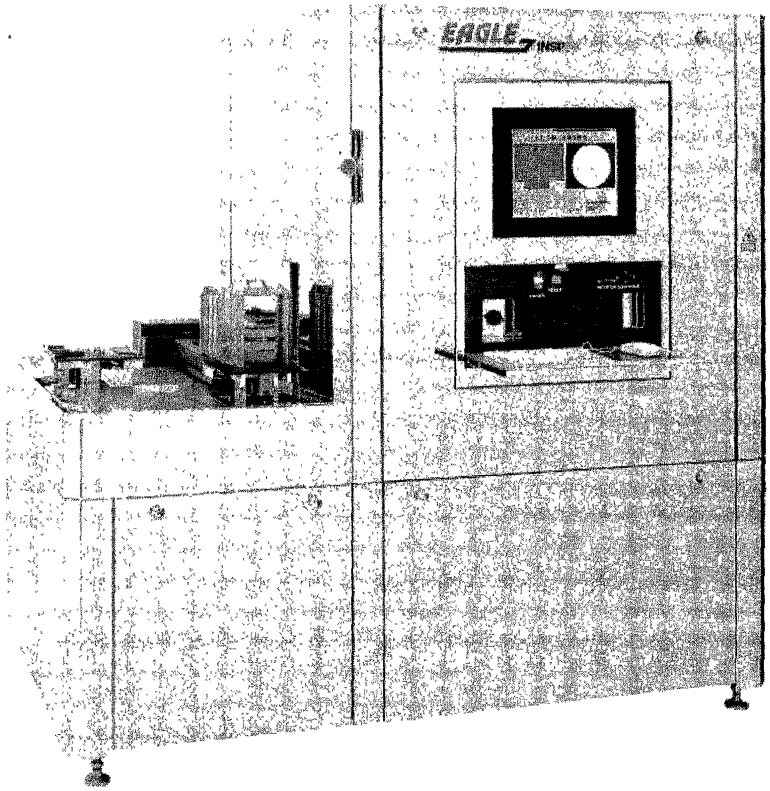


Wafer Inspection System
(Photo courtesy of Inspex)

Optical System. An optical microscope uses light reflection to detect surface defects. The modern optical microscope is integrated into a wafer inspection station that includes robotic wafer handling and software interface with video and defect classification. A typical optical system is shown in Figure 7.16. Microscope manufacturers have achieved optical systems capable of detecting smaller objects by improving the optics and using new light sources with shorter wavelengths (the ability to detect smaller objects improves with shorter light wavelengths). The technology of optics and light are discussed in detail in Chapter 14.

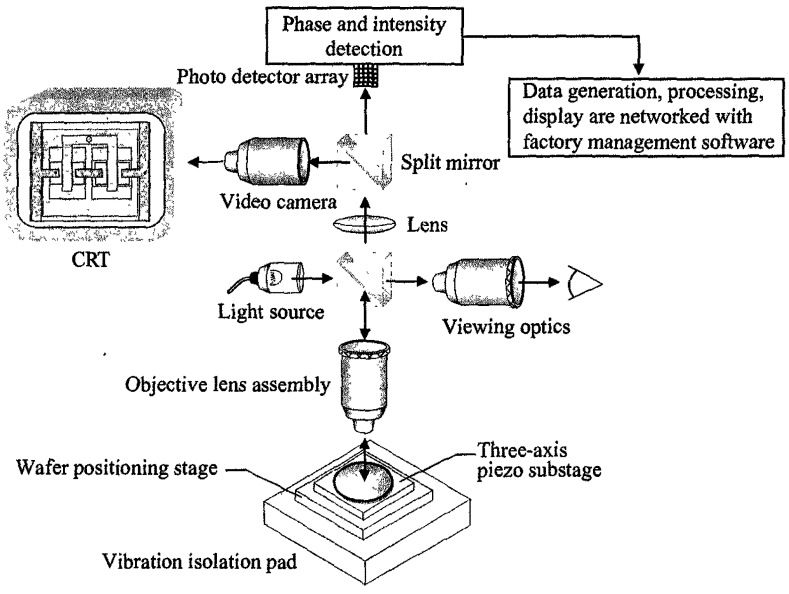


FIGURE 7.16 Schematic of Optical System

TSMC-1014
U.S. Patent 9,337,193

An important parameter for optical microscopy is *contrast,* which is the ability to distinguish between parts of an object. Techniques used in modern optical microscopes to enhance contrast are brightfield and darkfield detection (described on page 161), confocal contrast, and color interference contrast. A *confocal contrast microscope* uses a scanning technique to view a single point of the object at a time, thus providing better image contrast and therefore better visualization of the object (see Figure 7.17). A confocal microscope uses either visible light or laser scanning. *Color interference contrast* splits a beam of light into a direct and a reference beam. The direct beam is altered by the sample and then recombined with the reference beam, creating an image based on interference.

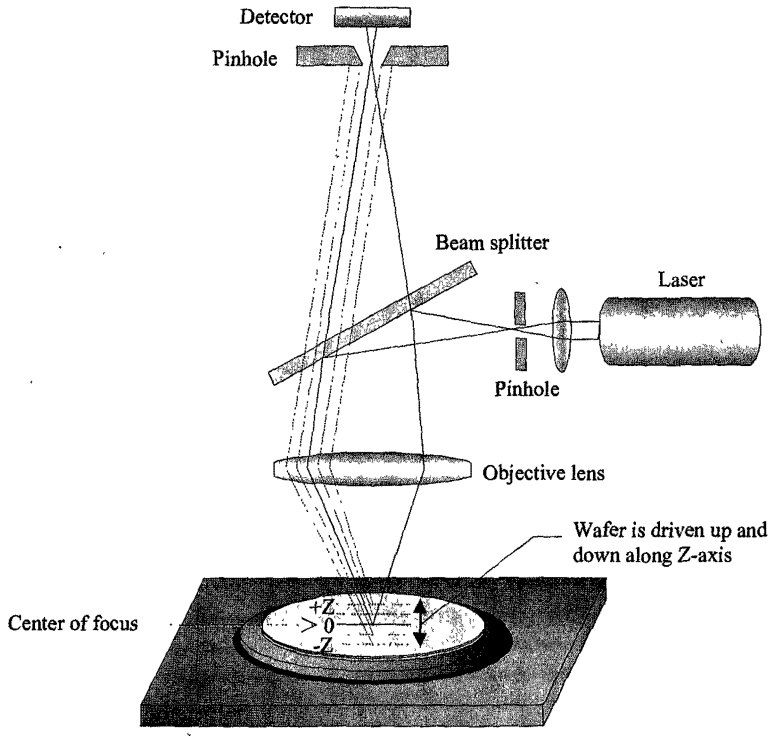


FIGURE 7.17 Principle of Confocal Microscopy

Light Scattering Defect Detection ■ In the beginning of semiconductor manufacturing, particles and surface defects were inspected by an operator using a light source and a microscope. This method was unreliable, leaving much to the subjectivity of the operator. A technique for identifying particles based on *light scattering* (also referred to as *laser scattering* or *scatterometry*) came into widespread use in the mid-1980s. This darkfield-detection method identifies surface particles and other defects by illuminating the surface with laser light and then using optical imaging to detect any light scattered by particles (see Figure 7.18 on page 164). The strength of the scattered light signal received from an individual particle depends on factors such as size, shape, composition, surrounding wafer surface (e.g., roughness and haze), and the type of equipment used. Particle diameters down to about 0.1 μm are currently capable of being detected with light-scattering techniques.[17]

Initially, when light-scattering particle detection became widespread in the mid-1980s, it could only detect particles on polished wafer surfaces. As the technology has advanced, it is now possible to detect particles on patterned product wafers and wafers with a thin film surface. As device geometries decrease, the critical particle size that creates killer defects also decreases. It was thought that the need to detect smaller particles would cause light-scattering technology to disappear, mainly because the particle size detection limit of about 1-μm was on the order of the light wavelength. When the light wavelength is the same as the particle size, it is difficult to detect particles because the scattered light intensity is not clearly defined.[18] Improvements in light scattering have been attained as sensitivity factors such as the choice of wavelength of the incident light beam are reduced. In this manner, light-scattering surface measurement has become a dominant

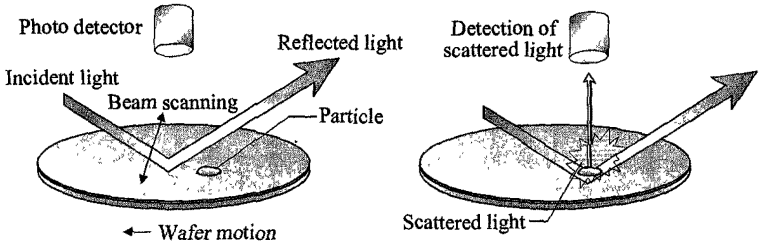


FIGURE 7.18 Particle Detection by Light Scattering

technology for tool monitoring of surface defects with both unpatterned and patterned wafers, primarily because its high-speed measurement provides rapid results during manufacturing process runs.[19]

Particles Per Wafer Per Pass. A practical aspect of measurement tools that detect wafer surface defects is the ability to create particle maps to count the particles on the wafer, showing where they occur as well as the distribution of particle diameters. A major effort in the fab is to perform *particles per wafer per pass* (*PWP*) measurements with unpatterned wafers to check the defect counts from process runs (see Figure 7.19). Examples of processes analyzed for PWP are photoresist spin tracks, dielectric deposition chambers, and chemical mechanical planarization (CMP) equipment. To perform a PWP procedure, a technician uses a wafer inspection tool to count the number of defects on a test wafer, runs the test wafer through the production equipment, and then recounts the number of defects. Corrective action is taken to reduce the number of defects and thereby increase the number of functional die on a wafer. It is often difficult to identify the smaller particles. In this case, the light-scattering detection equipment may work in conjunction with the scanning electron microscope (SEM) to identify particle composition for corrective action.

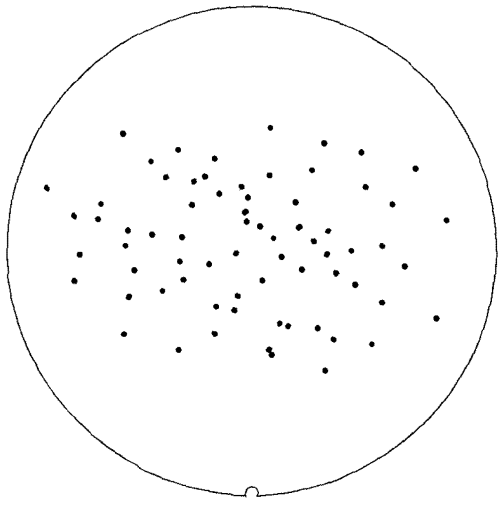


FIGURE 7.19 Particle Map

Patterned Surface Defects

Patterned wafers for fabrication metrology are regular production wafers used for in-line monitoring of surface defects. Production wafers common in IC fabrication may consist of multilayer metals with up to six or more layers that have been grown or deposited in furnaces and cluster tools. These multilayers make it challenging to distinguish between surface defects and the circuit pattern on the wafer. The majority of defects on patterned wafers are detected using light-scattering techniques. Optical microscopy (see previous section) is commonly used to detect surface defects on patterned wafers. There is also a less-common method of digital comparison, which correlates the wafer surface to a defect-free reference wafer.

Light Scattering on Patterned Wafers ■ Patterned wafers are becoming common for tool monitoring during process runs. The major defects on product wafers are particles, scratches, and pattern defects. With reduced geometries and increased process complexity, pattern defects occur more frequently. This condition makes defect detection on patterned wafers all the more critical.

Defect detection on patterned wafers with light scattering is similar to unpatterned wafers. However, the light-scattering process must be modified so that the measurement equipment can distinguish between scattered light from a particle and scattered light from a pattern edge. One technique is to use the regularity of the repeated circuit patterns to identify this particular scattered light and block it using a filter. Other patterned-wafer measurement equipment has a function that electrically cancels the cyclical signal pulses from the edge of the repeated patterns.

Critical Dimension (CD)

An important reason for critical dimension (CD) measurements is to achieve precise control over all line widths of the product being built. In CMOS technology, the transistor's gate structure is very critical. The gate width determines the channel length and the length of the channel affects speed. CD variation often indicates some instability in a critical part of the semiconductor manufacturing process. It is projected that the critical dimension (CD) for the 16-Gbit DRAM will be 0.1 μm in the year 2006.[20] To achieve this CD control, the metrology tool will need a precision and accuracy of better than 2 nm—the size of four silicon atoms side by side.[21] The measurement tool that is capable of attaining this level of metrology measurement is the scanning electron microscope (SEM).

Scanning Electron Microscope (SEM) ■ The *scanning electron microscope* (*SEM*) has been the dominant measurement tool to verify acceptable CD control in all submicron generations since the early 1990s. The first commercial SEMs were produced in the 1960s. The SEM can achieve magnifications up to 100,000 to 300,000X, which is significantly higher than optical microscopes. The resolution (smallest feature detectable) for SEMs is on the order of 40 to 50 Å. Cross sections of wafers viewed with the SEM can provide defect information. Because of the concern for controlling submicron linewidths, SEM technology was developed specifically for CD measurement in the 1980s, and is sometimes referred to as CD-SEM. The SEM measurement tool is often coupled with other analytical techniques like EDX or FIB (see the following section).

SEM Basics. The SEM is a sophisticated microscope that functions by creating a highly focused beam of electrons that scans an object while detectors measure the resulting scattered electrons.[22] It is a nondestructive and noncontact metrology tool. The SEM has an electron gun, focusing elements for shaping the electrons into a beam, and a final electrostatic-magnetic focusing system that makes the electrons strike the sample within a small, 2 to 6 nm spot (see Figure 7.20 on page 166). Since electrons have a very short wavelength (for instance, 1.22 Å for 100 eV electrons, versus about 5,500 Å for visible light), atomic-scale objects can be viewed with SEM electron imaging.

The electron gun produces a beam of electrons in a vacuum chamber of about 10^{-6} torr. It is desirable to have a high current of stable electrons with a narrow energy spread. Near the wafer, the electrons are focused into a narrow beam with a cylindrical magnetic objective lens often combined with electrostatic focusing elements to create a high-energy beam. The beam undergoes x-y deflection with an electrostatic deflector to scan the wafer. When the incident beam strikes the wafer, secondary and backscattered electrons, along with other electrons, X-rays, and photons, are emitted or transmitted due to interaction between the beam and the sample surface. The Everhart-Thornly (E-T) detector is used to collect the secondary electrons and create an electronic image that represents the sample surface. Backscattered electrons are also collected and offer superior compositional contrast between different materials.

The energy of the electron beam is directly related to the image needed. A low-energy beam (<2 keV) has a low accelerating voltage needed for nondestructive, in-line CD measurement. A high-energy electron beam (100 – 200 keV) is used to image underlying or deep structures such as contact holes. High-energy beams make it possible to nondestructively image below the surface of the wafer (say ~ 20 μm for a high-energy beam). Nevertheless, the problems with SEMs are most severe when imaging dense or deep structures, such as very fine resist lines or deep and narrow contact holes on insulator layers. This problem is mainly due to the difficulty of getting a satisfactory signal of

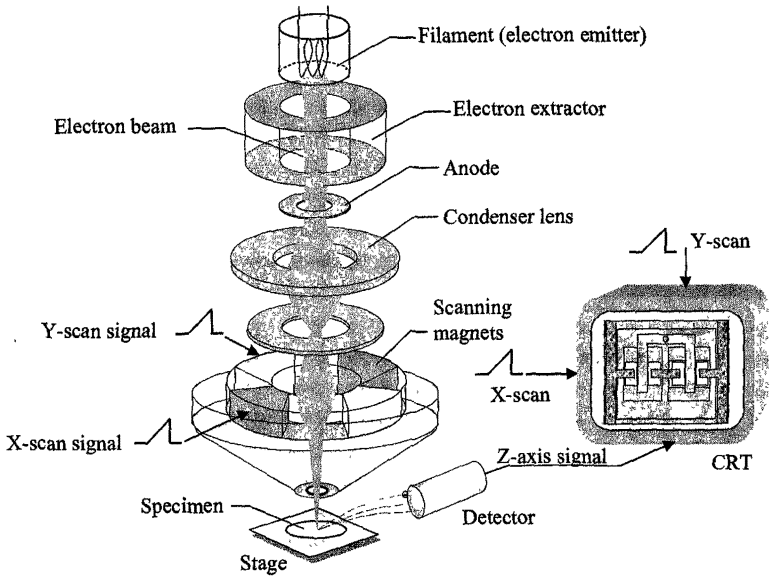
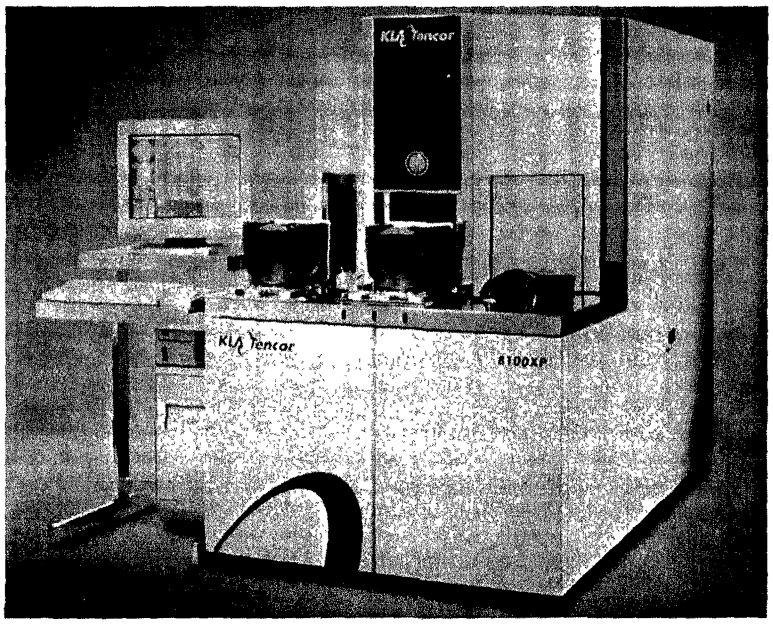


FIGURE 7.20 Simple Schematic of CD-SEM

secondary electrons from such structures. The limitations of the SEM are that it requires a high vacuum and it is necessary to coat insulating samples with conductive films prior to imaging.

CD-SEM. The *CD-SEM* has high resolution imaging with automated equipment control that rapidly evacuates the chamber to the desired vacuum, automatically positions the wafer, and uses a preprogrammed process recipe to select certain measurement sites. It is ideal for the CD-SEM to have a high wafer throughput to support production (up to 70 wafers/hour at five measured sites per wafer). The instrument can examine and measure wafers from all orientations and tilt angles of up to 60° (critical for measuring fine process patterns, sidewalls, and holes). A CD-SEM also can be used to perform defect review and analysis. Additional benefits of SEM imaging include the integration of other metrology tools, such as X-ray compositional analysis, and focused ion beam milling (see the following section).



CD-SEM
(Photo courtesy of KLA-Tencor)

Step Coverage

Because of surface topography during the fabrication of wafers, the ability to achieve conformal step coverage is a desirable material property (see Figure 7.21). *Conformal step coverage* has uniform material thickness in all regions of the step, including the sidewalls and corners.

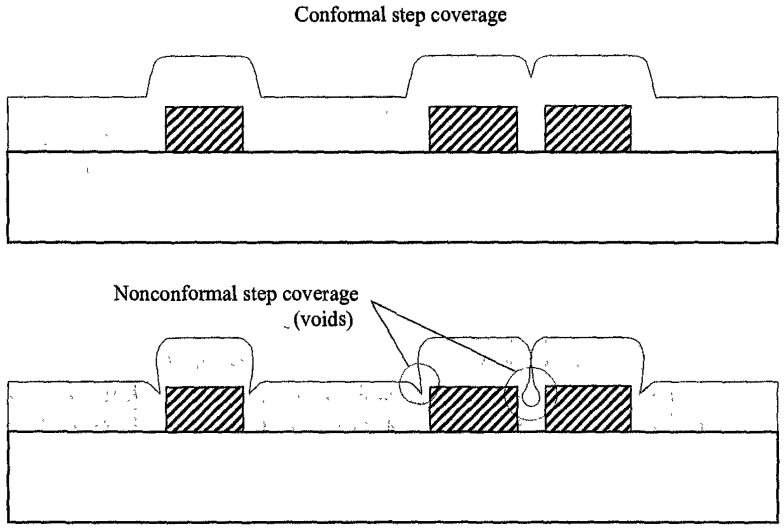


FIGURE 7.21 Step Coverage

A high resolution, nondestructive *surface profiler* with a stylus tip is often used to measure step coverage and other features on the wafer surface. This automated surface metrology tool uses a stylus that comes in contact with the wafer surface with a force as low as 0.05 mg, capable of profiling soft production-wafer films without damage to the wafer surface (see Figure 7.22). The stylus usually has a diamond tip with a radius of 0.1 μm, with older stylus having radii up to 12.5 μm.[23] Current profilers can measure wafer features as small as 0.1 μm with a 7.5 Å step height repeatability.[24]

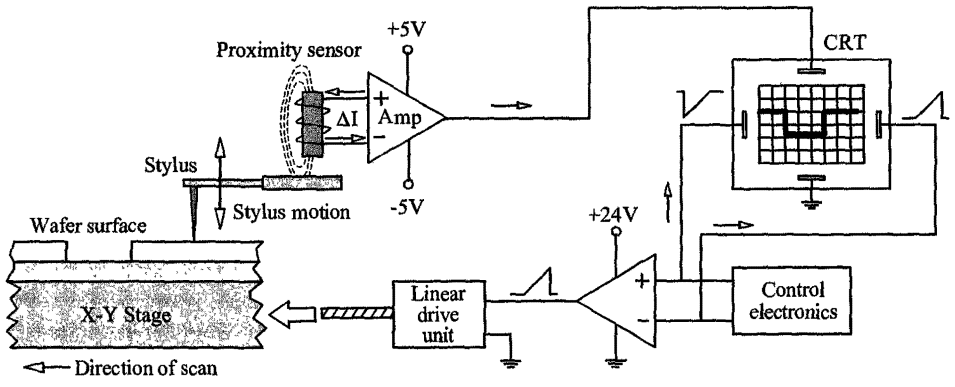


FIGURE 7.22 Surface Profiler

Overlay Registration

Overlay registration is used following the photolithography process to measure the ability of the tools and processes to accurately print photoresist patterns over a previously-etched pattern on a wafer. Due to shrinking feature sizes, the registration of the mask pattern to the wafer is a challenge because of reduced tolerances available for overlay registration. Furthermore, the increased use of chemical mechanical planarization (CMP) creates very low-contrast images on the wafer that are difficult to distinguish. This condition makes alignment of the wafer to the mask more complicated.

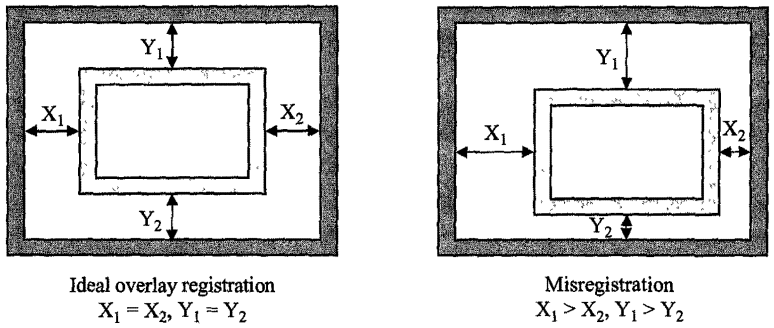


FIGURE 7.23 Overlay Registration Inspection Patterns

An automated-overlay metrology tool is used in photolithography to compare the registration of the special overlay mask pattern transferred to the photoresist with the etched overlay features on the wafer. It is important for this tool to be capable of measuring alignment targets on process layers that have low-contrast and grainy patterns in the wafer. Using brightfield reflected light is not sufficient to focus and measure images of low-contrast alignment targets on the wafer surface.

The primary method used today to measure registration of overlay targets is *coherence probe microscopy* (*CPM*). It is also referred to as a *correlation microscope.* Coherent light has waves with a definite phase relationship to one another, meaning that the individual light waves do not move relative to one another. Using coherent light permits CPM to image not only the amount of light scattered off a surface but also its phase relationship. This method is able to capture wafer surface information in the Z axis along with the wafer surface plane, which improves focusing of the wafer target and enhances the low-contrast image of a polished overlay target. Out-of-focus and diffracted signals are ignored.[25]

Capacitance-Voltage (C-V) Test

MOS device reliability is highly dependent on the thin layer of high-quality oxide in the gate structure. Contamination in the gate oxide region can lead to a shift in the threshold voltage requirement, causing device failure. Mobile ionic contaminants (MICs) and other undesirable charge conditions are detected at the SiO_2 layer after oxidation steps with a test known as *capacitance-voltage test,* or *C-V test.* The C-V test is commonly done to detect ionic contamination after oxidation steps. In addition, the C-V test provides information about gate oxide integrity (GOI), including the dielectric thickness, the dielectric constant, k, the resistivity of the silicon between the electrodes (to characterize majority carrier concentration), and the flatband voltage (a voltage level where there is no potential difference across the oxide structure).

The appropriate model for understanding gate oxide performance is the parallel-plate capacitor, which was explained in Chapter 2. For a MOS device, there are two capacitors in series that are functioning when the threshold voltage is applied. The first is the gate oxide, sandwiched between the doped polysilicon gate and the channel region below the gate structure. The second capacitor is formed in the silicon substrate material. This results from the attraction of charges into the gate region to form the conducting channel (known as inversion) during application of the threshold voltage. The oxide and silicon substrate are modeled as series capacitors during the C-V test (see Figure 7.24).

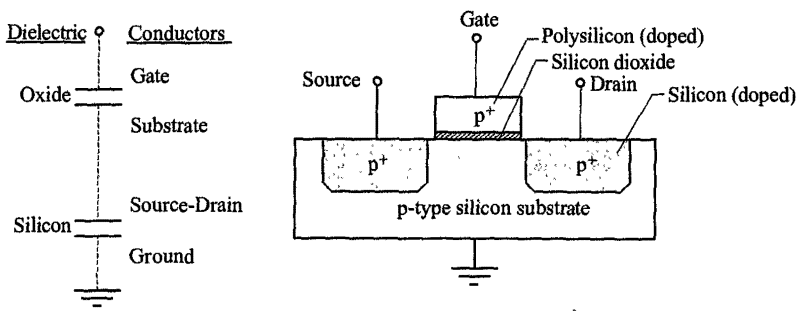


FIGURE 7.24 MOS Model of Two Capacitors at Gate Region

Since two capacitors in series have a combined lower capacitance than any individual capacitor, there is a drop in the capacitance of the gate structure when the threshold voltage is applied. This expected drop in capacitance is used in the C-V test to verify that no undesirable charges (e.g., mobile ionic contaminants) exist.

Steps for C-V Test ■ During a C-V contamination test, the two series capacitors in the gate region are modeled by using a special wafer. A variable voltage is applied between a metallized area on an oxide layer and the lightly-doped silicon beneath the oxide (see Figure 7.25).

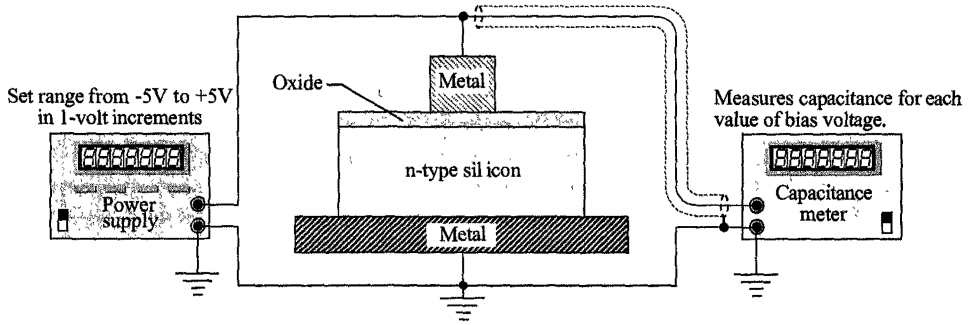


FIGURE 7.25 C-V Test Setup and Plotting

The first step for a C-V test is to apply a variable-voltage bias to both the metal contact of the oxide film under test and the lightly-doped silicon beneath the oxide. The bias is varied from positive to negative for a p-type silicon and negative to positive for n-type silicon. The objective of this first step is to deplete the silicon directly beneath the metallized area of majority carriers. In this manner, the silicon functions as a dielectric, which serves to reduce the total capacitance of the test structure because the oxide and silicon substrate are capacitors in series. The capacitance versus voltage during this test is plotted (see Figure 7.26).

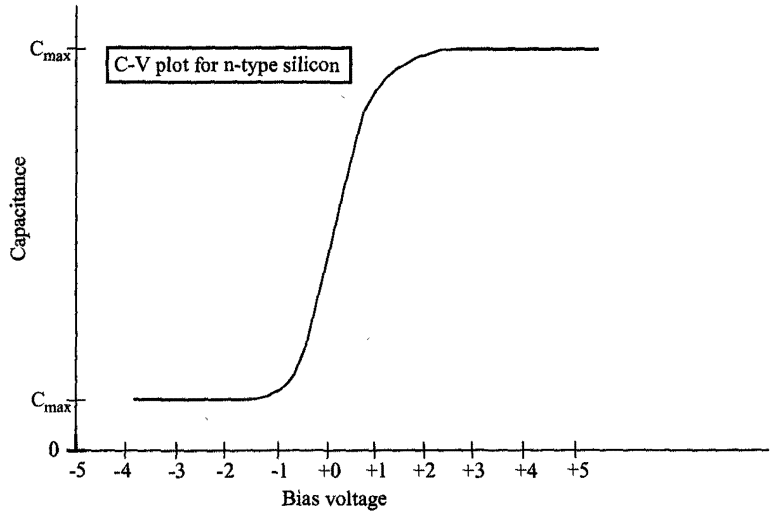


FIGURE 7.26 Capacitance versus Voltage for *n*-Type Silicon (First Step of C-V Test)

The second step is to apply a constant positive DC voltage to the metallized area (its magnitude depends on the oxide thickness) while heating the wafer to 300°C for five minutes followed by a cooldown prior to removing the bias (see Figure 7.27 on page 170). The elevated temperature serves to increase the mobility of the ionic contaminants. The positive voltage bias repels the positive ionic contaminants and drives them toward the oxide-silicon interface.

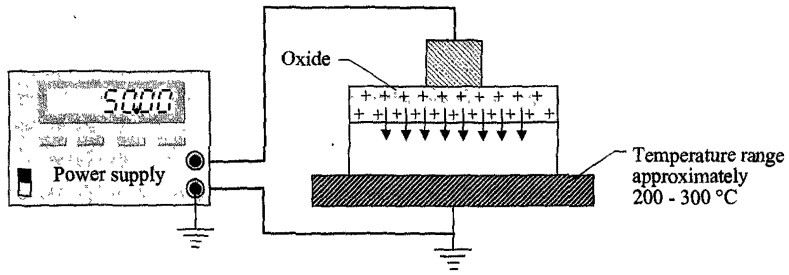


FIGURE 7.27 Ionic Charge Collection in C-V Test

The third step is to repeat the first C-V plot. However, now if there are positive ionic contaminants collected at the Si/SiO_2 interface, then a more negative voltage is required to get an equivalent charge in capacitance. This is a voltage shift, Vs, that is a measure of the amount of contamination in the oxide (see Figure 7.28). The magnitude of the voltage shift is proportional to the ionic contamination in the oxide, the oxide thickness, and the wafer doping. The actual amount of ionic contamination can be calculated from this plot.

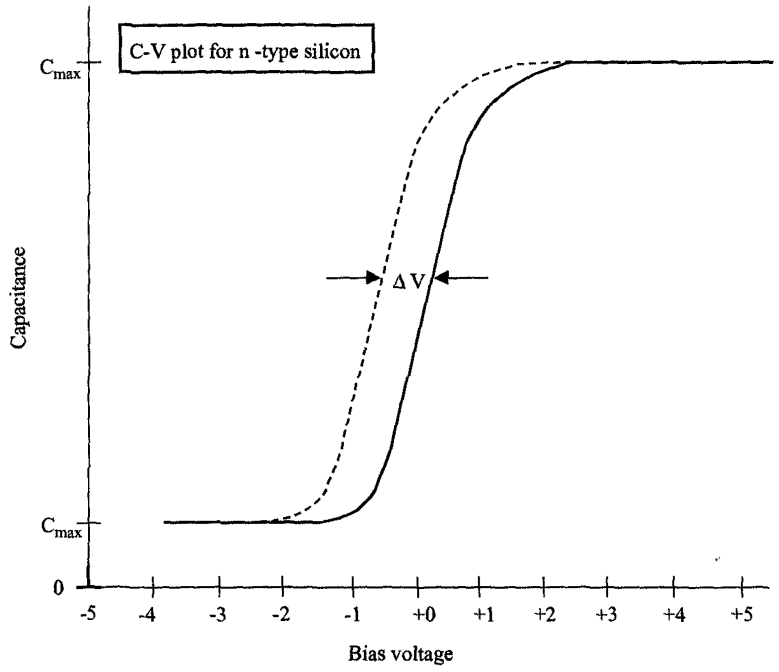


FIGURE 7.28 Voltage Shift in n-Type Silicon

The final step is to verify that this shift is due to contamination and not oxide charging. A negative voltage is applied while heating the substrate to move the ionic contaminants away from the Si/SiO_2 interface. Performing the C-V test again should produce the original plot, verifying that it was actually the ionic contaminants that produced the voltage shift.

Benefits of C-V Test ■ The C-V analysis is used to monitor the wafer to ensure that the cleaning process removes ionic contaminants. This analysis is done either on fabricated test wafers or production wafers with a special test structure (test structures are described in Chapter 19). The analysis cannot tell you where the contamination comes from, such as the wafer surface, cleaning step, equipment maintenance, or furnace process. The C-V test is usually a part of any evaluation for process changes. In this case, wafers are tested in two groups. One group undergoes normal processing prior to C-V testing while the other group follows the new process and then is C-V tested. The results are compared to verify that the proposed process is acceptable. There is a small amount of voltage shift that is acceptable, depending on the oxide thickness and the sensitivity of the measurement equipment.

Contact Angle

Contact-angle meters are used to measure adhesion of liquids to the wafer surface and to calculate surface energies or adhesion tension. This measurement characterizes wafer surface parameters such as wettability, cleanliness, finish, and adhesion (see Figure 7.29). The contact (tangent) angle formed between a liquid drop and its supporting surface is relative to the forces at the liquid/solid or liquid/liquid interface and can be used as a wafer-test specification or as a quality characteristic. Both direct-angle measurements and indirect dimensional measurements methods are employed to obtain highly accurate and repeatable contact angle measurements.

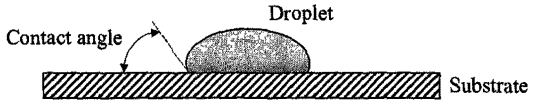


FIGURE 7.29 Contact Angle

ANALYTICAL EQUIPMENT

This section provides an overview of the major analytical equipment used to support wafer fabrication. These analytical tools provide highly accurate wafer measurements and are typically located in an off-line lab to support production problems. Figure 7.30 shows when some of these instruments were first used and how important each one is or is projected to become in either process development or manufacturing. The analytical equipment reviewed includes:

- Secondary-ion mass spectrometry (SIMS)
- Time of flight secondary-ion mass spectrometry (TOF-SIMS)
- Atomic force microscope (AFM)
- Auger electron spectroscopy (AES)
- X-ray photoelectron spectroscopy (XPS)
- Transmission electron microscope (TEM)
- Energy- and wavelength-dispersive spectrometer (EDX and WDX)
- Focused ion beam (FIB)

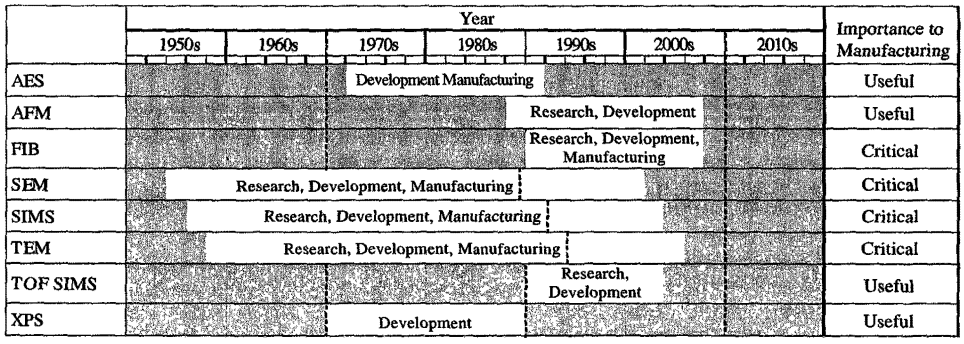


FIGURE 7.30 Relative Importance of Analytical Equipment

Secondary-Ion Mass Spectrometry (SIMS)

Secondary-ion mass spectrometry (*SIMS*) is a method of eroding a wafer surface with ions accelerated in a magnetic field to analyze the surface material composition. These ions strike the surface of the wafer and dislodge, or sputter, other ions, some of which are known as secondary ions (see Figure 7.31 on page 172). Secondary ions contain the wafer material and any impurities such as dopants. They are collected and analyzed in a vacuum chamber with a mass spectrometer that can

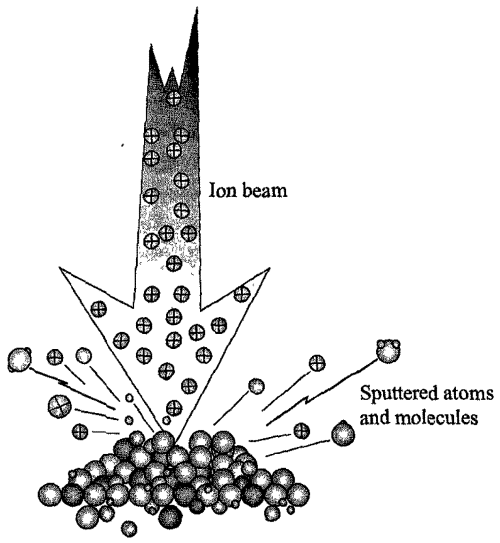


FIGURE 7.31 Ion Beam Sputtering of Surface Material

identify the type of dopant and its concentration in the silicon.[26] SIMS is essentially a destructive measurement technique and requires an ultrahigh vacuum environment (around 10^{-10} to 10^{-12} torr).

Recall that ions are charged particles, able to be accelerated or deflected in a magnetic field. When accelerated ions strike a wafer surface, this ion-beam sputtering process involves violent collisions at the atomic level. This action could be compared to a sandblaster, which propels sand with such energy that it removes surface pieces but leaves behind craters in the base material. However, with an ion beam, there are also chemical reactions that take place during the collision, creating many complex molecules in the sputtered material. The particular incident ion could be nearly any element, but is typically chosen as Cs^+, O_2^-, O_2^+ or Ar^+ and is selected based on the collision chemistry that will best produce secondary ions.

SIMS Tool Description ■ A common method to produce ions used in SIMS is the duoplasmatron (see Figure 7.32). A gas, commonly O_2 or Ar, is converted into a plasma by a low-pressure arc formed between a hot filament and an anode. An extraction electrode is used to draw charged atoms from the plasma.

Cs^+ ions are produced by a different method. A pellet of cesium chromate is vaporized in a heated reservoir. The vapor feeds into a hot (~1100°C) and porous tungsten plug in an ionization chamber. The vapor diffuses through the pores of the plug and becomes ionized, followed by an extractor that collects the ions and forms them into an ion current.

After the ions are produced, they go through an analyzer magnet to select the ions of interest (e.g., Cs^+ or O^-). These ions are then either focused into a small spot with a magnet, known as an ion microprobe, or used to flood the surface of the sample. The latter is referred to as an ion microscope and is the most common technique today. With an ion microscope, the sputtered secondary ions are collected simultaneously from many points on the surface and then identified by analyzing their mass-to-charge ratio in the mass spectrometer. A common mass spectrometer, the quadrupole, is based on an oscillating electric field that separates ions into defined oscillations based on their mass, permitting certain ions to pass through an aperture and be identified.

The rate of material removal using SIMS can range from a high sputter rate, known as *dynamic SIMS* and often used for ion implanter characterization, to a slow sputter rate of a single monolayer of atoms every few hours. The latter is referred to as *static SIMS*, and is frequently used to analyze contamination in thin oxide and nitride films. The SIMS measurement tool is sensitive enough to measure ppba (parts per billion–atomic) impurities in a junction or contact that is only 1 to 10 μm across. It has become the primary measurement tool used to verify the performance of ion implanters because it can characterize the dose and depth of the junction plus reveal any undesirable metal impurities at the junction.[27]

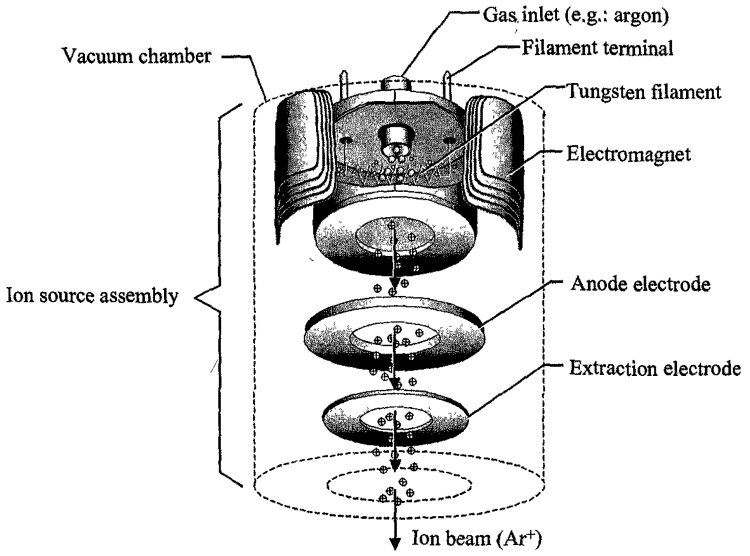


FIGURE 7.32 Ion Production in a Duoplasmatron

Time of Flight SIMS (TOF-SIMS) ■ A limiting factor for SIMS is that the mass spectrometer only detects around 0.001% of all the ions emitted from the sample. Another approach, referred to as *time of flight SIMS (TOF-SIMS)*, can detect as many as 10 to 50% of the ions emitted by the sample. The TOF-SIMS identifies ions of different mass by measuring the time it takes for an ion to travel the length of a fixed path, since the charged particle velocity is a function of mass (see Figure 7.33). This fact makes it possible to reduce the incident beam current by a factor of up to 10^5 relative to a SIMS with a quadrupole spectrometer. The rate of material removal is so low that only a fraction of a single-surface monolayer is removed in an hour. For this reason TOF-SIMS is essentially nondestructive and is ideal for very thin films on the wafer surface.[28]

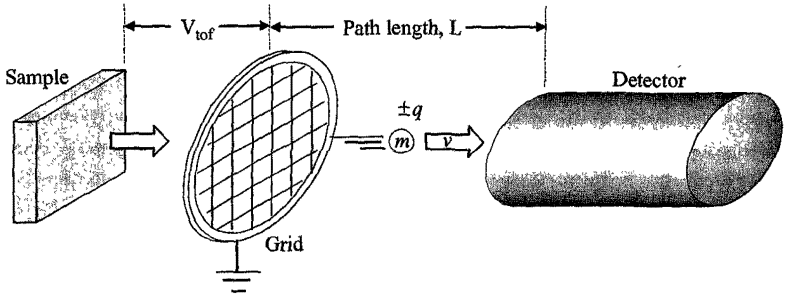


FIGURE 7.33 TOF-SIMS Mass Spectrometer Principle

Atomic Force Microscope (AFM)

The *atomic force microscope (AFM)* is a surface profiler that scans a small, counterbalanced tip probe over the wafer to create a 3-D surface map.[29] It was first demonstrated in 1986 and uses optical technology, direct tip contact, and a laser to sense the position of the tip relative to the surface (see Figure 7.34 on page 174). Probe and surface separation is so small (on the order of 2 Å) that atomic forces affect the probe between the surface and the tip. The tip geometry is critical and must be characterized to provide accurate measurements. AFM measurements are exceedingly slow, which is a problem for in-line measurements in a production environment.

There are several variations of AFM technology.[30] In the simplest system, a laser beam reflects off the top surface of the probe tip and is directed to a photodiode. When the surface topography moves the probe, it changes the position of the laser in the photodiode, creating an electronic map of the surface. A more advanced method has the probe oscillate very near the surface. Atomic van der Waals forces interact with the vibrating tip, inducing phase shifts in the vibration that are detected by an imaging system. The benefit of this approach is that it can measure vertical

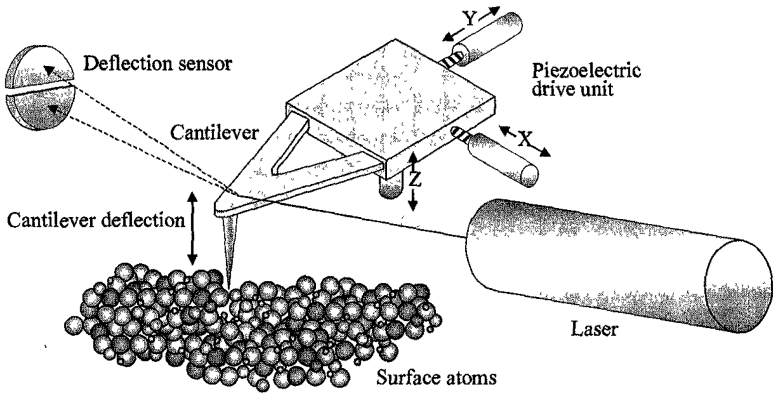


FIGURE 7.34 Schematic of Atomic Force Microscope

photoresist profiles. AFM is a recent measurement technology and is considered a promising alternative to SEMs.

Auger Electron Spectroscopy (AES)

Auger electron spectroscopy (*AES*) measures the energy of auger electrons emitted by the surface of a sample when struck by an incident electron beam. It is very sensitive to the surface, with a depth of only 10 to 50 Å. Auger (pronounced ōzhā´) electrons were discovered in 1923 by Pierre Auger in France, and make up a very small percent (< 0.1%) of the total electrons produced in a sample (secondary electrons are the dominant species). The energy associated with Auger electrons provides a distinct link back to the parent atom that is used for identification of the sample elements.

Because Auger electrons are easily absorbed by the sample, only those Auger electrons on the outer monolayers of the surface escape and are detected in AES. This makes Auger technology especially suited for analyzing the surface of materials, usually to a depth of about 2 nm. Oxides of metals, silicides, and wafer surfaces are readily sensed by Auger spectroscopy. Furthermore, the Auger spectrometer uses a highly focused electron beam, with the smallest beam spots on the order of 12 nm, which is beneficial in microcircuit analysis. AES requires an ultrahigh vacuum environment in order to reduce contaminant formation on the sample.

X-Ray Photoelectron Spectroscopy (XPS)

X-ray photoelectron spectroscopy (*XPS*) is primarily used to identify chemical species on a sample surface, analyzing to a sample depth of about 2 nm (equivalent to AES). All elements are detectable except hydrogen and helium (their detection requires a good spectrometer). In XPS, X-ray photons are directed toward the wafer surface and interact with certain core-level electrons referred to as XPS electrons (see Figure 7.35). If the X-ray energies exceed the XPS electrons' binding energy, then the electrons are emitted from the sample. Because the binding energy of an electron is influenced by its chemical surroundings, XPS identifies both the element and its chemistry.[31]

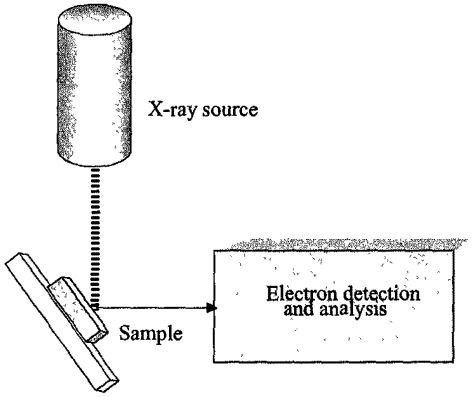


FIGURE 7.35 Schematic of XPS Measurement

Transmission Electron Microscope (TEM)

The principle of operation for the *transmission electron microscope (TEM)* is similar to an SEM, with the major difference that the beam of electrons is transmitted through an ultrathin slice of the sample (on the order of 10 to 100 nm thick). Based on factors such as the electron wavelength, accelerating voltage and specimen thickness, an image is formed and magnified on a screen with a resolution of about 2 Å (see Figure 7.36).

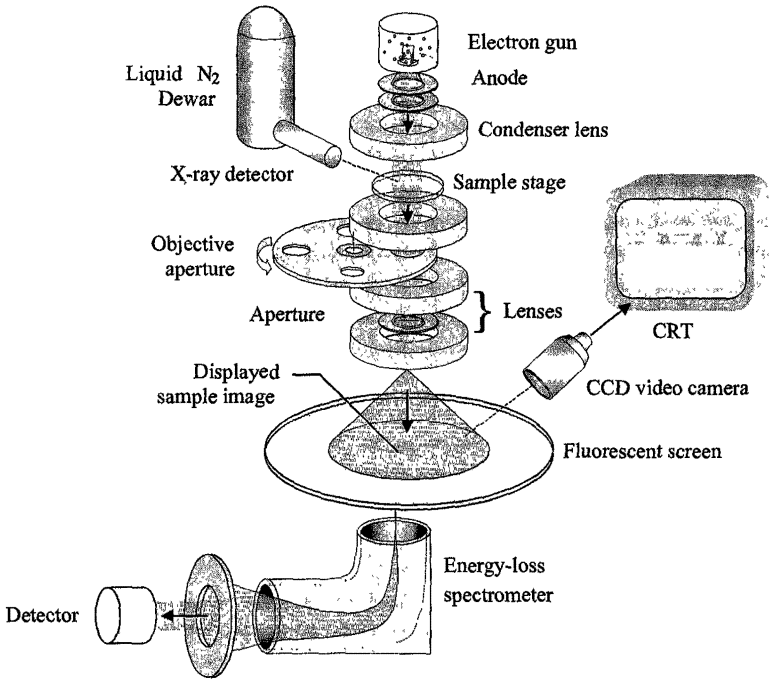


FIGURE 7.36 Schematic of TEM

TEM is the only metrology tool capable of quantifying some very small features on a wafer. For instance, silicon crystal point defects, such as single dislocations introduced into an active junction by ion implant (which leads to junction leakage), can be imaged on an atomic scale by TEM. Table 7.3 outlines some process areas of wafer fabrication that benefit from TEM analysis.[32] The most difficult aspect of TEM is sample preparation, with a wide array of techniques used, such as mechanical polishing, chemical etching, and ion-beam milling.

TABLE 7.3 Sample TEM Applications in Semiconductor Manufacturing

	TEM Measurement
Silicon Material	Dislocation and stacking fault densities in Si shallow junctions
Patterning	Precise sidewall profile in polysilicon and metal structures
Metallization	Characterization of metal silicides and alloys
Implantation	Surface and buried implant damage
Contamination	Thin organic and oxide films at submicron contact interfaces

Energy- and Wavelength-Dispersive Spectrometer (EDX and WDX)

A useful signal generated by the incident electron beam of an SEM is the characteristic X-ray produced by the energetic primary electrons emitted by the sample material. These X-rays carry information related to the atom species present in the sample, permitting the identification of a variety of thin films, particles, and defects created in semiconductor manufacturing. X-ray detectors are found on nearly all electron microscopes today.

The *energy-dispersive spectrometer* (*EDX*) is the most widely used X-ray detection method for identifying elements and is complementary to SEM. This is primarily because the EDX can quickly detect X-rays of all energies simultaneously, which means it is a relatively fast measurement. However, because it penetrates well beyond the sample surface, it should not be considered a surface analysis. EDX operation is based on a large diode made from a high-quality doped Si crystal and isolated from the SEM vacuum chamber by a thin (~25 μm) beryllium window (see Figure 7.37). An X-ray passing through the window produces a series of electron-hole pairs, which are electronically detected and identified back to the energy level of the X-ray. EDX can acquire all energy peaks and make a spectrum plot within a few minutes.[33]

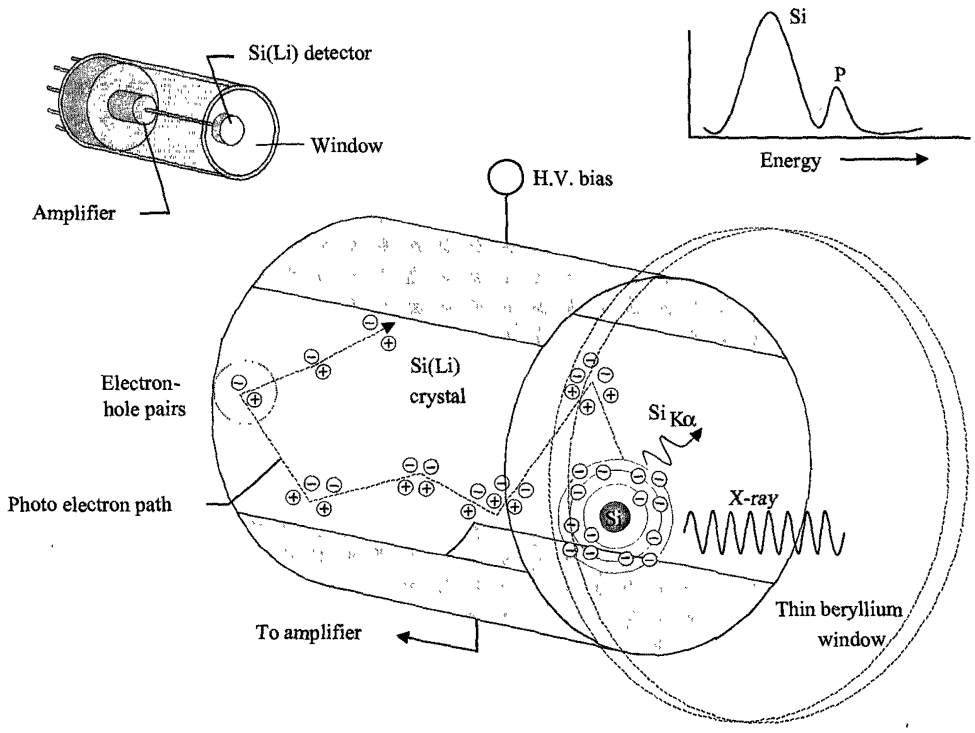


FIGURE 7.37 Energy-Dispersive Spectrometer (EDX)

The *wavelength-dispersive spectrometer* (*WDX*) operates based on a diffracting crystal and photo counter. The crystal separates and disperses the incoming X-rays by wavelength, which are then collected by a photo counter detector. The appropriate crystal is selected and placed in the path of the X-rays. Critical parameters are X-ray wavelength, lattice spacing of the crystal, incident angle, and order of reflection. WDX is a slow measurement but yields superior measurements with high accuracy.

Focused Ion Beam (FIB)

Traditionally, and often still today, a wafer is analyzed by removing it from the cleanroom to a lab where it is cross-sectioned and examined on an SEM. This process started changing when *focused ion beam* (*FIB*) systems became prevalent in the early 1990s because of their convenient cross-sectioning capabilities. The FIB is a destructive technique similar to an SEM in design and operation,

except that the primary beam is made of Ga^+ ions instead of electrons. These ions are focused through a set of lenses into a small spot. At the location where they impact the wafer, atoms are ejected into the vacuum, creating a small void of precisely controlled shape and depth in the sample material. This precision gives the FIB the possibility of making cross sections in specific locations of a wafer. The FIB technique is, however, a time-consuming process.

Focused ion beam (FIB) milling is capable of carving a 10 to 100-nm thick cross section from any area on a wafer inside the cleanroom (see Figure 7.38). It can cut through metal, polysilicon, oxide, and nitride layers with little or no damage to adjacent structures. Typically a high-current beam is made for the initial cut and is followed by a low-current beam with a tighter focus for final sample polishing. FIB can make the thin samples necessary for TEM. For this reason, equipment suppliers make custom FIB designs specifically for this application.

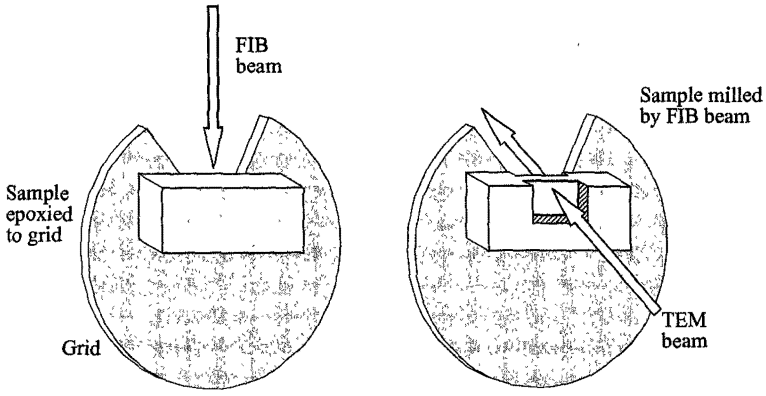


FIGURE 7.38 Focused Ion Beam (FIB) Milling

SUMMARY

IC metrology measures the performance of fabrication processes, typically with monitor or production wafers. Metrology tools are either stand-alone or integrated into the process tool, and they often use advanced defect-analysis software. Specified quality measures are used at different process locations to verify acceptable performance. Film thickness is most often measured with the four-point probe sheet resistance (opaque films), ellipsometry (transparent films), and reflection spectroscopy. Lesser-used methods are X-ray and photoacoustic. Film stress on a wafer is determined by measuring the wafer curvature. The refractive index of a transparent substance is measured using ellipsometry. The dopant concentration is measured using four-point probe, infrared interference, and spreading resistance probe. Surface defects are measured with unpatterned or patterned wafers, with the latter becoming more cost-effective. The optical microscope is used, with advanced features to improve wafer-surface contrast. Light scattering is often used to detect particles on the wafer surface, with contour maps generated to highlight the defects. The recent trend is for more surface defect analysis to be done with patterned wafers. Critical dimensions are measured using the scanning electron microscope (SEM). Enhanced techniques are used to measure overlay registration. The capacitance-voltage test provides information on the quality of the gate oxide structure, primarily to ensure the oxide is contamination-free. The contact angle verifies acceptable wafer surface quality. Sophisticated analytical tools are used to characterize the wafer and verify acceptable process performance.

KEY TERMS

metrology
defects
defect density
monitor wafers (test wafers)
yield
stand-alone measurement tool
integrated measurement tool
automatic defect classification (ADC)
quality measures
film-thickness metrologies
sheet resistance, R_s
sheet resistivity, ρ_s
four-point probe
Van der Pauw method
contour maps
ellipsometry
variable-angle spectroscopic ellipsometry (VASE)
spectroscopic ellipsometry
reflection spectroscopy (reflectometry)
X-ray fluorescence (XRF)
Total-reflection XRF (TRXRF)
photoacoustics
film stress
refraction
dopant concentration (dose)
thermal-wave system
spreading resistance probe
unpatterned wafers
brightfield detection
darkfield detection
optical microscope (light microscope)
contrast
confocal contrast microscope
color interference contrast
light scattering (laser scattering or scatterometry)
particles per wafer per pass (PWP)
patterned wafers
scanning electron microscope (SEM)
CD-SEM
conformal step coverage
surface profiler
overlay registration
coherence probe microscopy (CPM)
correlation microscope
capacitance-voltage test (C-V test)
contact angle meters
secondary ion mass spectrometry (SIMS)
dynamic SIMS
static SIMS
time of flight SIMS (TOF-SIMS)
atomic force microscope (AFM)
Auger electron spectroscopy (AES)
X-ray photoelectron spectroscopy (XPS)
transmission electron microscope (TEM)
energy-dispersive spectrometer (EDX)
wavelength-dispersive spectrometer (WDX)
focused ion beam (FIB)

REVIEW QUESTIONS

1. What is metrology? What is the purpose of metrology in IC fabrication?
2. Define a defect. What is meant by wafer defect density?
3. Discuss the difference between monitor wafers and production wafers for metrology measurements.
4. What is the difference between a stand-alone measurement tool and an integrated measurement tool?
5. Describe automatic defect classification (ADC), and explain how this analysis can improve metrology.
6. Define semiconductor quality measures. List twelve different quality measures used in IC fabrication. State the process areas where the different quality measures are used.
7. List the two general areas of film-thickness metrology.
8. State the formula for sheet resistance for a thin film. Explain how sheet resistance is interpreted for a square sample. What are the units for sheet resistance?
9. What is the formula for the sheet resistivity of a thin film?
10. Describe the four-point probe and give its benefits for measuring sheet resistance.
11. What is the Van der Pauw method?
12. Explain what a contour map is.
13. Explain the basic principles of ellipsometry. What are the benefits for using ellipsometry to measure thin film thickness?
14. Explain what VASE is. What applications does VASE offer improved ellipsometry?

15. Describe spectroscopic ellipsometry.
16. Describe reflection spectroscopy. What is a dual-beam spectrometer?
17. How is film thickness measured using X-rays? What does XRF stand for? What is total reflection XRF?
18. Describe how photoacoustic technology is used for film-thickness measurement.
19. How is thin film stress measured on a wafer?
20. How is refraction measured on a wafer? What does a variation in the refractive index indicate?
21. Can the four-point probe be used to characterize dopant concentration in a film?
22. Explain the principle of thermal wave for measuring dopant concentration.
23. Describe the spreading resistance probe method for dopant concentration.
24. What is brightfield detection? What is darkfield detection?
25. Describe the general state of optical microscopy for wafer-surface defect detection.
26. What is contrast? Discuss two methods to improve inspection contrast on the wafer surface.
27. Explain how light scattering can be used to detect surface defects.
28. Explain particles per wafer per pass (PWP).
29. What is the dominant measurement tool for wafer critical dimension?
30. Explain the basic operation of an SEM.
31. What is a CD-SEM?
32. How is step coverage often measured?
33. What is overlay registration? State and discuss the primary technique for measuring overlay registration.
34. Why is the capacitance-voltage test done? Describe four steps necessary to perform this test.
35. State the purpose of the contact-angle metrology measurement.
36. Describe secondary-ion mass spectrometry (SIMS).
37. What is TOF-SIMS and in what condition is it desirable to use?
38. Explain atomic force microscopy (AFM).
39. Describe the technique for using the Auger electron spectroscope.
40. What is X-ray photoelectron spectroscopy?
41. Explain transmission electron microscopy.
42. Explain the difference between EDX and WDX.
43. Describe focused ion beam (FIB) milling and explain its benefits.

METROLOGY EQUIPMENT SUPPLIERS' WEB SITES

Applied Materials	http://www.appliedmaterials.com/products/
Carl Zeiss Microelectronics	http://www.zeiss.de/
Cerprobe Corp.	http://www.cerprobe.com
FEI Company	http://www.feic.com
Gaertner Scientific Corp.	http://www.gaertnerscientific.com/
Hitachi	http://www.hitachi.com/semiequipment/products.html
Inspex	http://www.inspex.com/
International SEMATECH	http://www.sematech.org/public/index.htm
JA Woollam Co. Inc.	http://www.jawoollam.com/
JEOL	http://www.jeol.com/
Kaman Instrumentation	http://www.kamaninstrumentation.com/
Keithley Instruments	http://www.keithley.com/
Kemco Instruments Co.	http://www.kerncoinstr.com/cam.htm
Kevex Spectrace	http://www.kevexspectrace.com/
KLA-Tencor	http://www.kla-tencor.com/splash.html
Leica	http://www.leica.com/
Leybold Inficon Inc.	http://www.leyboldinficon.com/
The Micromanipulator Co. Inc.	http://www.micromanipulator.com/
Nanometrics	http://www.nanometrics.com/
National Institute of Standards	http://www.nist.gov/
Nicolet Instruments	http://www.nicolet.com/
Nikon	http://www.nikonusa.com/
Olympus America Inc.	http://www.olympus.com/
Perkin-Elmer	http://www.perkinelmer.com/
Rudolph	http://www.rudolphtech.com/home/
Schlumberger	http://www.1.slb.com/ate/diagsys
SEMI	http://www.semi.org/

Sonoscan Inc.	http://www.sonoscan.com/
Therma-Wave	http://www.thermawave.com/index.htm
Veeco Instruments	http://www.veeco.com/

REFERENCES

1. S. Butler, "Process Control Through Integrated Metrology," *Solid State Technology* (January 1999): p. 34.
2. C. W. Pearce, "Crystal Growth and Wafer Preparation," *VLSI Technology,* 2nd ed., ed. S. Sze, (Boston: McGraw Hill, 1988), p. 300.
3. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation,* 2nd ed., (New York: McGraw-Hill, 1998), p. 117.
4. J. C. C. Tsai, "Diffusion," *VLSI Technology,* 2nd ed., ed. S. Sze, (Boston: McGraw-Hill, 1988), p. 300.
5. Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors: Technology Needs,* (San Jose, CA: SIA 1997), p. 80.
6. P. Burggraaf, "Thin Film Metrology: Headed for a New Plateau," *Semiconductor International* (March 1994): p. 58.
7. R. DeJule, "Advances in Thin Film Measurements," *Semiconductor International* (May 1998): p. 64.
8. C. Morath et al., "Ultrasonic Multilayer Metal Film Metrology," *Solid State Technology* (June 1997), p. 85.
9. M. Dax, "X-Ray Film Thickness Measurements," *Semiconductor International* (August 1996), p. 98.
10. R. DeJule, "Advances in Thin Film Measurements," *Semiconductor International* (May 1998): p. 56.
11. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* (Boston: Kluwer Academic, 1995), p. 371.
12. Semiconductor Industry Association, *National Technology Roadmap,* p. 80.
13. E. Rimini, *Ion Implantation: Basics to Device Fabrication* (Boston: Kluwer Academic, 1995), p. 70.
14. D. Schroder, *Semiconductor Material and Device Characterization,* 2nd ed., (New York: Wiley, 1998), pp. 31–35.
15. J. Baliga, "Defect Detection on Patterned Wafers," *Semiconductor International* (May 1997): p. 64.
16. A. Braun, "Defect Detection Overcomes Limitations," *Semiconductor International* (February 1999): p. 52.
17. A. Braun, "Defect Detection and Review Enter New Era," *Semiconductor International* (May 1998): p. 61.
18. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation,* 2nd ed., p. 227.
19. A. Braun, "Defect Detection and Review Enter New Era," *Semiconductor International* (May 1998): p. 66.
20. Semiconductor Industry Association, *National Technology Roadmap,* p. 14.
21. M. Davidson and A. Vladar, "The Physics of Metrology Instruments," *Solid State Technology* (June 1998): p. 135.
22. Material for description of CD-SEM taken from M. Davidson and A. Vladar, "Physics of Metrology Instruments," pp. 136-142.
23. A. Braun, "Analytical Techniques for Process Problem Solving," *Semiconductor International* (October 1997): p. 112.
24. KLA-Tencor, Product literature for *High Resolution Profiler, KLA-Tencor HRP-220,* (August 1998).
25. KLA-Tencor, Product literature for *Automated Overlay Metrology Tool, KLA-Tencor 5200 XP,* (May 1998).
26. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation,* 2nd ed., p. 88.
27. Ibid., p. 402.
28. Ibid., p. 418.
29. K. Wilder, B. Singh, and W. Arnold, "Novel In-Line Applications of Atomic Force Microscopy," *Solid State Technology* (May 1996), p. 109.
30. M. Davidson and A. Vladar, "Physics of Metrology Instruments," p. 144.
31. D. Schroder, *Semiconductor Material and Device Characterization,* 2nd ed., p. 701.
32. W. Runyan and T. Shaffner, *Semiconductor Measurements and Instrumentation,* 2nd ed., p. 332.
33. Ibid.

CHAPTER 8

GAS CONTROL IN PROCESS CHAMBERS

Semiconductor manufacturing is a cyclic repetition of several major process steps. Many fabrication processes involve chemical reactions that take place in process chambers. The driving force for these chemical reaction processes is to optimize the required chemical reactions by introducing the correct chemicals in the proper environment (e.g., vacuum) while providing energy to drive the reaction. At the same time, detrimental aspects of the reaction are minimized, such as exposure to moisture, the ambient environment, and contaminants. This optimum condition is reached by carefully introducing the necessary mix of precursor chemicals into the process chamber, often as a gas, and then monitoring the chemical reaction to achieve the desired conditions on the wafer surface.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain why process chambers are used in semiconductor manufacturing.
2. Describe the benefits of a vacuum, the vacuum ranges, and appropriate pumps.
3. Explain the need for gas flow in process chambers, and describe how it is controlled.
4. Explain what an RGA is and why it is beneficial in process chambers.
5. Describe what plasma is and how it is obtained.
6. Discuss the effects of contamination in process chambers and explain how to minimize it.

INTRODUCTION

During the beginning of the semiconductor industry, only two wafer fabrication steps required a vacuum chamber for processing. A vacuum was used for the evaporation of aluminum for the one and only metal layer and for the evaporation of gold on the back of the silicon wafer so that the circuit die could be mounted in its transistor package.[1] Vacuum processes in those early days occurred in a chamber known as a bell jar (see Figure 8.1 on page 182).

Present-day wafer processing often requires chemical reactions that take place in process chambers. The *process chamber* is a controlled vacuum environment where intended chemical reactions occur under controlled conditions. For this reason, a process chamber for chemical reactions is sometimes referred to as a *reactor*. Process chambers have many functions:

- Controlling how gas chemicals flow into and react in the chamber in close proximity to the wafer.
- Maintaining a prescribed pressure inside the vacuum environment.
- Removing undesirable moisture, air, and reaction by-products.
- Creating an environment for chemical reactions such as plasma to occur.
- Controlling the heating and cooling of the wafer.

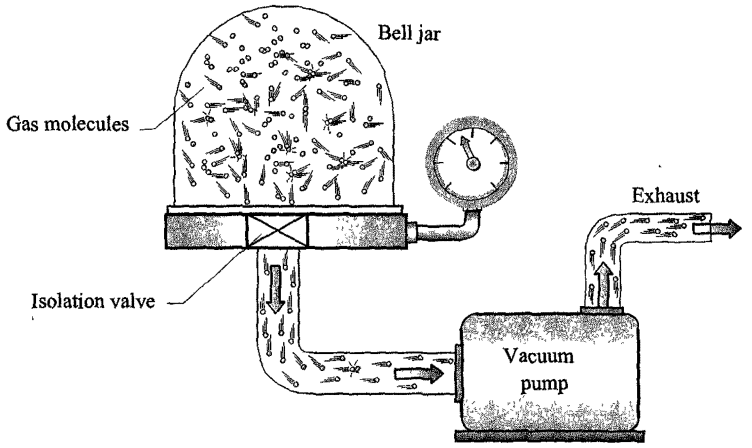


FIGURE 8.1 Early 1960s Vacuum Bell Jar

Gas is typically the material state used to transport the various chemicals needed in the process chambers. Gas flows into the process chamber as a result of the pressure differential between the supply system and the vacuum conditions in the chamber. Sometimes solid materials are used in the process chamber. An example is the solid metal target used as a source of material during sputtering (see Chapter 12).

Since the late 1980s, process chambers have been configured in the form of a *cluster tool.* The multiple process chambers are clustered around a central transfer chamber with a wafer transport system that is typically a robot arm (see Figure 8.2). This equipment design permits integration of multiple process steps. Wafers are transported from process chamber to process chamber under vacuum, eliminating native-oxide formation and reducing contamination on the wafer. Cluster tools also improve wafer manufacturing throughput (defined as the number of wafers processed per unit time) because there is no need to vent the chamber during wafer transfer steps.

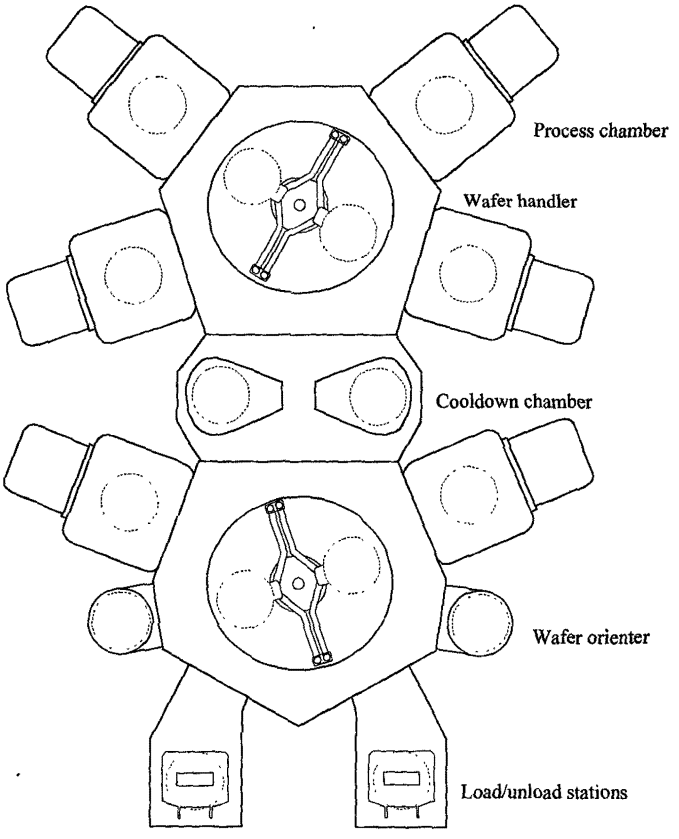


FIGURE 8.2 Integrated Cluster Tool
(Used with permission from Applied Materials, Inc.)



Cluster Tool with Integrated Process Chambers
(Photo courtesy of Applied Materials, Inc.)

VACUUM

Many chemical reactions in wafer fabrication are performed in a vacuum environment. A *vacuum* exists when there is less pressure in an enclosed volume than in the surrounding atmospheric pressure. The benefits of a vacuum in semiconductor manufacturing are shown in Table 8.1.[2]

TABLE 8.1 Benefits of Vacuum in Semiconductor Manufacturing

	Vacuum Condition	Benefit
1.	Create clean environment	Removes particles, unwanted gases, moisture, and contaminants.
2.	Low molecular density	Reduces the number of molecules in the system to reduce contamination and to move a gas out of the way (lower the molecular interference).
3.	Extend distance between collisions of molecules (Mean Free Path)	Provides the necessary condition for creating the plasma needed in semiconductor processes such as sputtering and etch.
4.	Accelerate reactions	Helps accelerate processes by lowering the vapor pressure of materials so they can react faster with other chemicals.
5.	Create a force	Creates a force, such as a vacuum pickup on a robot arm for wafer handling.

Vacuum Ranges

The following terms describe the different vacuum ranges: low vacuum, medium vacuum, high vacuum (also called high vac), and ultrahigh vacuum (UHV). These ranges are shown in Table 8.2 on page 184.[3] As a reference, the vacuum of deep space is about 10^{-16} torr.

Low vacuum (also referred to as *rough vacuum*) has two important characteristics:[4] gas flow is primarily by collisions between molecules (also known as *viscous flow*) and the pressure is high enough that true mechanical pressure gauges can be used. Low vacuum is commonly used in fabrication processes that depend on gas-phase chemical reactions, momentum transfer between

molecules and/or a high rate of interactions between the gas and surfaces. *Medium vacuum* is from about 1 torr to 10^{-3} torr and is a transition between low and high vacuum. *High vacuum* is characterized by having few collisions between gas molecules (molecular flow). This condition results in very clean wafer surfaces. *Ultrahigh vacuum* is a continuation of high vacuum with stringent control of the vacuum chamber design and materials to minimize undesirable gas contaminants.

TABLE 8.2 Vacuum Ranges

	Vacuum Ranges in Torr				
Wafer Fab Processes	**Rough 759 - 10^0**	**Medium 10^0 - 10^{-3}**	**High 10^{-3} - 10^{-6}**	**Ultrahigh 10^{-6} - 10^{-9}**	**Chapter in book**
Oxidation	Wafer handlers Atmospheric tools				10
Photo	Vacuum chucks Wafer handlers				13 - 15
Polish	Wafer handlers Slurry removal				18
Etch	Plasma resist strippers Plasma etchers				16
Deposition	Batch deposition tools Single wafer deposition tools				11
Metallization	Metal evaporators Metal sputtering tools				12
Ion Implant	Batch ion implanters Single wafer ion implanters				17
Metrology	Wafer inspection tools for quality checks and diagnostics Analytical tools for research and failure analysis				7

Mean Free Path

The average distance a gas molecule moves before it strikes another molecule is known as the *mean free path (MFP)*. When the pressure is lowered in a vacuum, the space between the gas molecules increases, which is an important factor for how gases flow through the system and for creating a plasma in the process chamber. The mean free path for air at different pressure regimes and standard temperature is given in Table 8.3.

TABLE 8.3 Mean Free Path and Molecular Density versus Pressure

	760 Torr (atmosphere)	**1×10^{-3} Torr**	**1×10^{-9} Torr**
# of molecules/cm^3	3×10^{19} (30 million trillion)	4×10^{13} (40 trillion)	4×10^{7} (40 million)
Mean Free Path	2×10^{-6} inches	2 inches	30 miles

VACUUM PUMPS

There are many different vacuum pumps used in semiconductor manufacturing. For our purposes, they can be categorized into two general types: roughing pumps and high vacuum pumps. *Roughing pumps* have several purposes: to achieve a rough to medium vacuum (pressure down to 10^{-3} torr) in a chamber, to evacuate the entry area for wafers into a cluster tool (known as the loadlock), and to exhaust a high vacuum pump (see Figure 8.3). *High vacuum pumps* are used for achieving high and ultrahigh vacuum from 10^{-3} torr to 10^{-9} torr. Modern vacuum pumps for new wafer fabs are dry, meaning they contain no oils or lubricants that can backstream into the process chamber and contaminate wafers as well as the process chamber. This chapter addresses only dry pumps as the intent is to focus on current technologies needed to support sub-quarter micron processes.

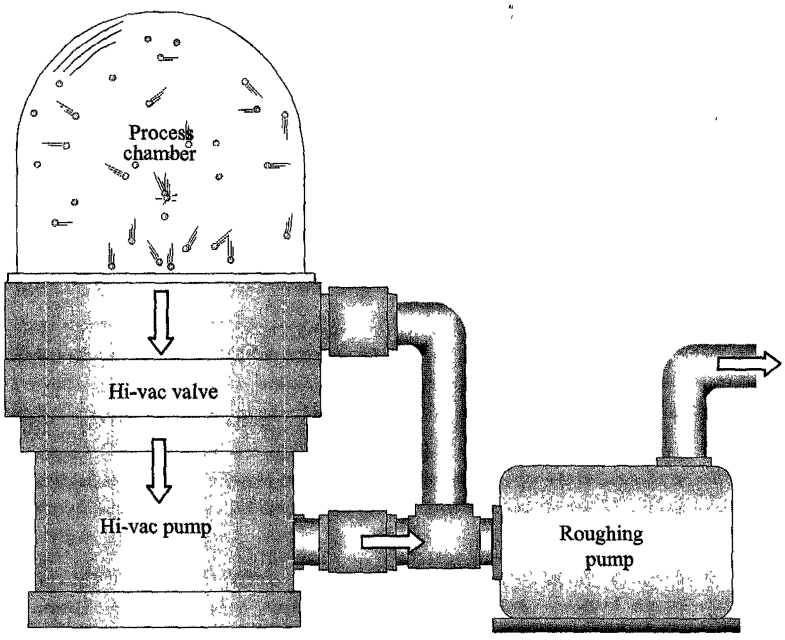


FIGURE 8.3 Roughing Pump Exhausting a High Vacuum Pump

Roughing Pump

When chamber pumpdown starts, roughing pumps remove more than 99.99% of the initial air or contaminants in the chamber. Removing this material with the roughing pump permits the high-vacuum pump to be optimized for removal of the remaining water vapor and gas molecules attached to the walls of the chamber. There are various types of roughing pumps, each with characteristics that work for a specific application. Two types of roughing pumps are:

- Dry mechanical pump
- Blower/booster pump

Dry Mechanical Pump ■ A *dry mechanical pump* employs mechanical devices to remove gases, such as the rotary claw dry pump (see Figure 8.4). The pump principle is typically based on increasing the chamber volume, thus lowering the pressure (i.e., Boyle's law). A mechanical pump often uses nonmetallic materials on the moving surfaces to avoid the use of sealing or lubricating fluids.

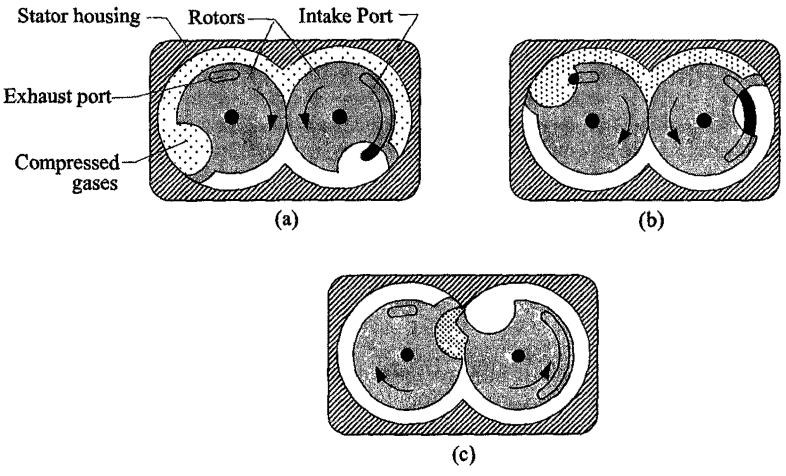


FIGURE 8.4 Rotary Claw Dry Mechanical Pump
(Used with permission from International SEMATECH)

Blower/Booster Pump ■ The *blower* or *booster pump* (also referred to as a *lobe pump*) is a widely used mechanical pump because it provides for high throughput of gas and requires no lubricants. It is desirable for systems where a high volume of gas must be pumped at a rough vacuum range. A blower is also referred to as a *Roots blower* or a *Roots-type blower.* The principle for how a blower works with the mating lobes is shown in Figure 8.5. A blower is often exhausted into a roughing pump because it will not pump at atmosphere under viscous flow. Newer blower pumps are being designed to exhaust directly to the atmosphere without the use of a roughing pump.

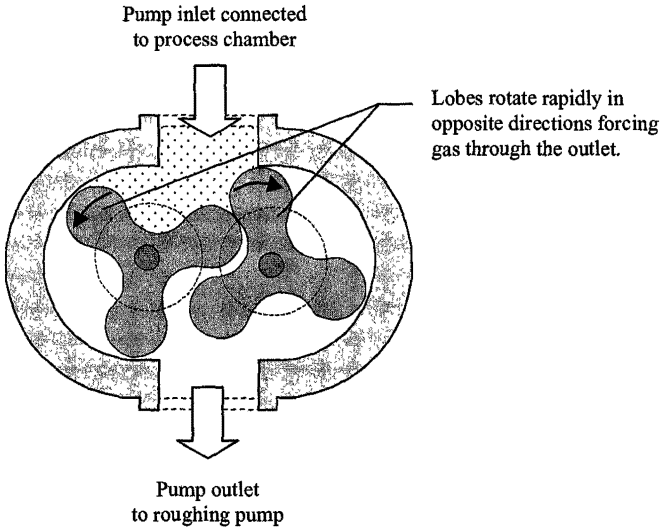


FIGURE 8.5 Roots Blower Pump

High Vacuum Pump

Two common high vacuum pumps are:

- Turbomolecular pump
- Cryopump

Turbomolecular Pump ■ The *turbomolecular pump* (usually called a *turbo pump*), is a versatile, reliable, and clean pump that is widely used in wafer fab equipment. The turbo pump is capable of reaching a pressure of 10^{-10} torr if a bakeout is done to drive out moisture in the process chamber. This pump also gives a fast start-up to full pumping action once turned on.

A turbo pump works on the principle of mechanical compression. There are a number of high-speed rotating blades (resembling blades in a jet engine) positioned between fixed blades that impart momentum and direction to the gas molecules (see Figure 8.6). Each set of rotating and fixed blades is a compression stage—a pump may have from ten to forty of these stages. Turbo pump blades will rotate at speeds up to 90,000 RPM. A turbo pump exhausts into a roughing pump because it cannot pump with viscous flow at atmospheric pressure.

Turbo pumps are specially designed for semiconductor applications. Magnetically levitated bearings (referred to as mag-lev) are common to eliminate the need for lubrication and special antivibration designs. The turbine blades can be designed to produce high gas throughput and high compression ratio for fast pumpdown of light gases.

The most common causes for failure of a turbo pump are exposing the pump suddenly to atmospheric pressure (called dumping the pump), particulate entering the pump, or physical shock. When a turbo pump is suddenly dumped to atmosphere, the turbine blades flex and come in contact with each other, causing catastrophic failure. Because the rotors on a turbo pump are precisely balanced, the pump should never be moved or bumped while it is in use. Extensive maintenance is not required on turbo pumps; most fabs run pumps until they crash and then change them.[5]

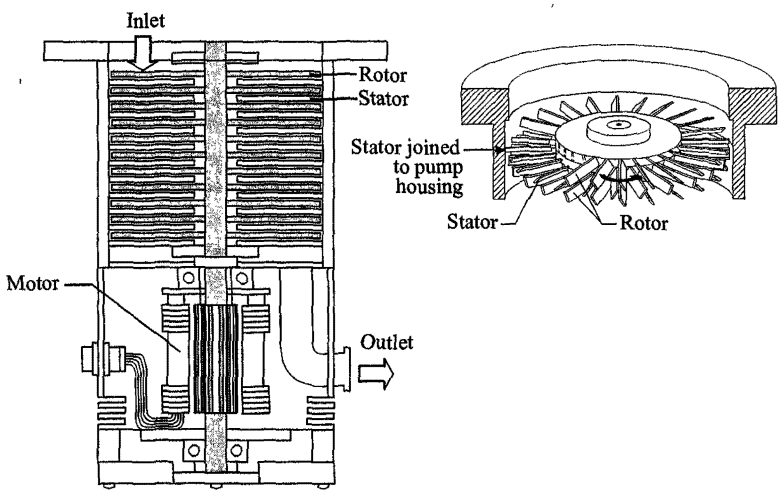


FIGURE 8.6 Turbo Pump Blades
(Used with permission from Varian Vacuum Systems)

Cryopump ■ A *cryopump* is a capture pump that removes gases from the process chamber by making them so cold that they are frozen and captured in the pump, thus providing the pumping action. The cryopump has become the industry standard in semiconductor equipment for pumping to the high and ultrahigh vacuum range. It has high gas throughput while pumping, plus has a very high water vapor pumping speed. This feature is useful when pumping down from atmosphere to remove the moisture in the chamber. The cryopump is extremely clean, with no oils or moving parts exposed to the vacuum. This attribute is desirable for wafer fabrication, which is a reason why this pump is common on new production equipment. There are two main components to a cryopump: a gaseous helium compressor and a pump module with a cold head, baffle, and body (see Figure 8.7).

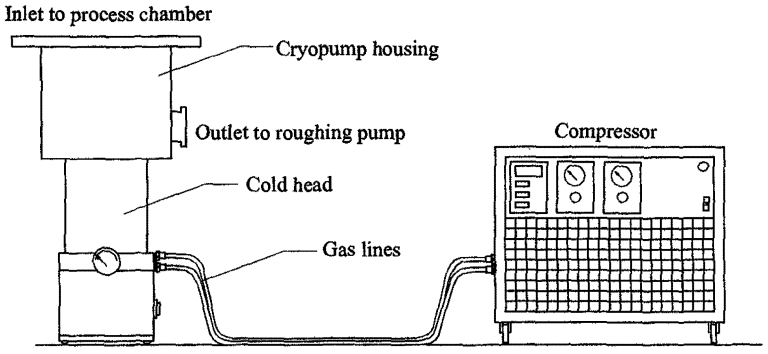


FIGURE 8.7 Cryopump Compressor and Pump Module
(Used with permission from Varian Vacuum Systems)

The helium compressor works similar to a regular refrigeration compressor (e.g., in the air conditioning system at home), except that it supplies high-pressure, high-purity, room-temperature helium to the expander module in the pump. When the gas expands in the pump module from a high-pressure to a low-pressure condition, the helium takes in heat and refrigeration occurs. This process produces cryogenic temperatures from 80 K down to 50 K. The pump has a second stage that expands the helium again and reaches approximately from 20 K to 10 K. Once the helium expands, it comes in contact with and cools multiple surfaces called *cryoarrays* (see Figure 8.8 on page 188). It is on these surfaces that the gases from the vacuum chamber are cooled and condensed or adsorbed. The condensed gases are immediately frozen on the cold cryoarray surfaces and trapped, which is essentially the pumping action.

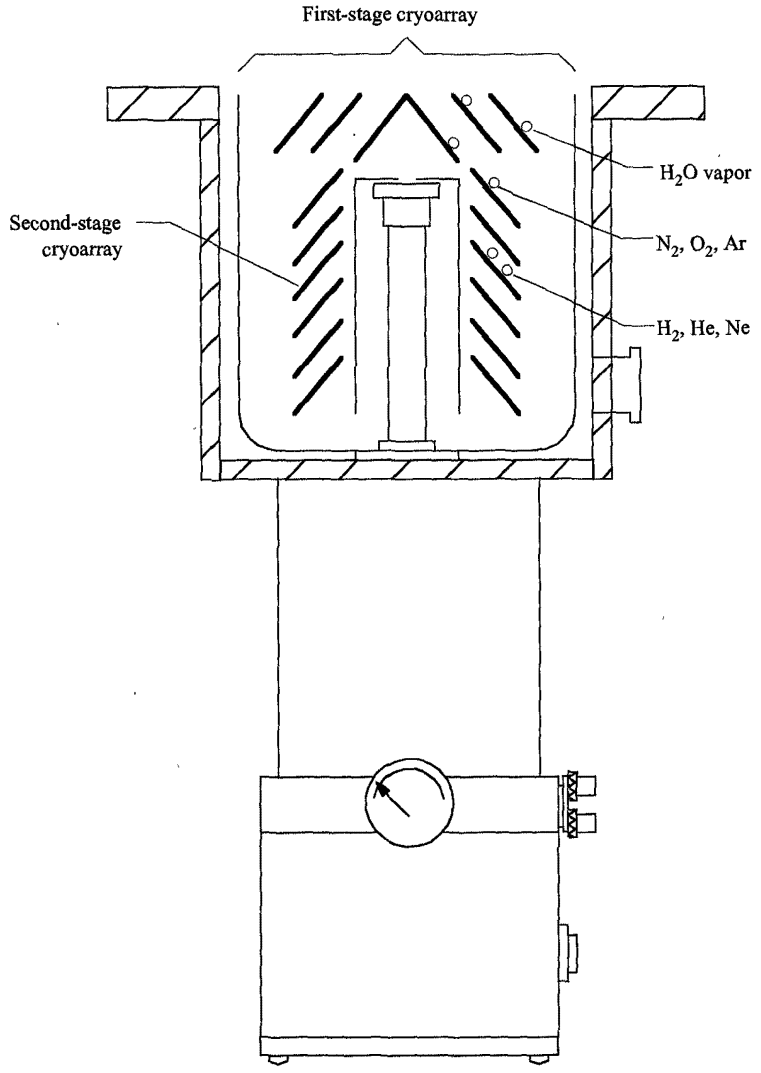


FIGURE 8.8 Cryoarray Surfaces in Pump Module
(Used with permission from Varian Vacuum Systems)

Since cryopumps capture gases by freezing the molecules rather than compressing and expelling them, the gases actually accumulate as frozen solids on the cryoarray surfaces. This process is similar to ice (frozen water) that forms on the inside of a freezer. These captured gases are removed periodically through a process called *regeneration,* where the pump is warmed to room temperature or above and the gases are vented through an appropriate vent line. Cryopumps require a roughing pump to remove the air from the pump and vacuum system.

Vacuum in Integrated Tools

The vacuum environments of an integrated cluster tool depend on the requirements of each individual process chamber (see Figure 8.9).[6] Chambers are isolated from one another, with a progressively better level of vacuum from the loadlock to the process chamber. The *loadlock* is where wafers enter the cluster tool, isolating the inner regions of the tool from the workplace environment. The system is designed to provide a well-controlled, low-contamination environment for wafer preparation and processing.

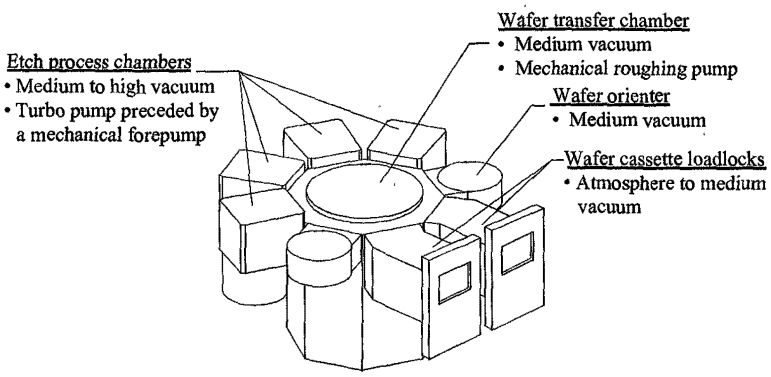


FIGURE 8.9 Cluster Tool Layout with Vacuum Environment

PROCESS CHAMBER GAS FLOW

The flow of gases into the process chamber is critical to attain the desired chemical reaction. The basic process chamber requirements for gas flow are:[7]

- Ability to handle a wide variety of bulk and specialty gases, many of which are corrosive and toxic.
- The control of gas flow into the process chamber is accurate and repeatable.
- The gas mix proportions are able to be controlled during the process run.
- Materials used in the chamber are not affected by the process gases and do not introduce contaminants into the gas stream.

When discussing gas delivery systems or vacuum, the mass quantity of gas flow at standard conditions is called *throughput* (Q). Throughput is the net number of gas molecules that pass through a point of the vacuum system in a specified period of time. Throughput defines the volumetric gas flow in a system at standard conditions. The most common units for throughput are torr-liters per second, standard cubic centimeters per minute (sccm), or standard liters per minute (slm).

Pumps are specified for a certain *pump speed,* which indicates how effectively the pump can remove gases. Pump speed is typically expressed in units of volume per unit time (such as liters/second or cubic feet per minute).

Throughput and pump speed are important in many semiconductor process steps. The chemical reactions that take place at the surface of the wafer may require high gas flow rates (which means high throughput). Factors such as the type of pump, its pumping speed, its location in the system, and the gas throughput are important variables that can determine whether an acceptable chemical reaction occurs at the wafer surface.

Mass Flow Controllers

Chemical reactions involve physical processes where molecular quantities are important for proper control of the reactions. From the ideal gas law we know that the number of gas molecules in a given volume changes in proportion to the absolute pressure and temperature. Thus, controlling gas flow into a chamber only by volume will not always yield the same number of gas molecules, which is undesirable for controlling chemical reactions.

To overcome this problem, gas flow into process chambers is controlled by use of a *mass flow controller (MFC)*, shown in Figure 8.10 on page 190. MFCs use the heat-transfer property of the gas to directly measure the mass flow rate into the chamber. It employs a thermal sensor to detect changes in the mass flow of the gas. Integrated tools will typically use many MFCs to control the flow of the various gases into the process chambers. A pressure regulator is required ahead of the MFC to ensure that a constant specified pressure is delivered to the MFC.

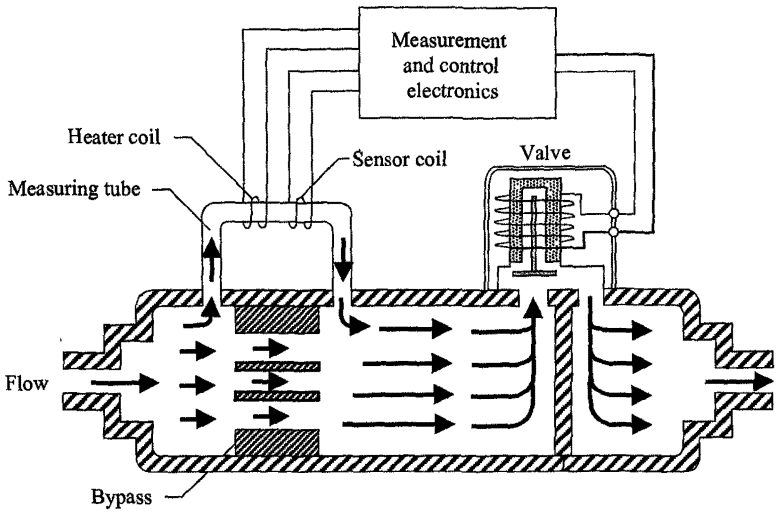


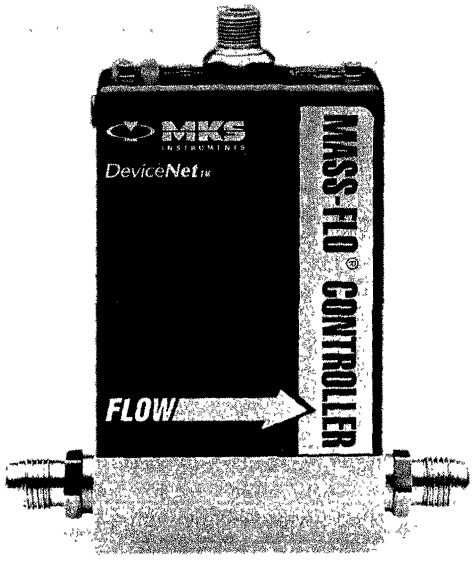
FIGURE 8.10 Thermal Mass Flow Controller
(Used with permission from International SEMATECH)

RESIDUAL GAS ANALYZER (RGA)

The *residual gas analyzer* (*RGA*) is an important process chamber instrument to identify the types of gas molecules remaining in an evacuated system. In this manner, it can be used for leak detection, and analysis of contamination in process chambers, and as a troubleshooting tool to solve vacuum-based problems in chambers.[8] Its most common applications are leak detection and process troubleshooting.

RGA Basics

The principle of the RGA is to separate, identify, and measure the quantity of all gas molecules in the chamber. An RGA measures the partial pressure contribution of each gas present in the vacuum system, as well as the total pressure from all gas molecules. It typically operates only in the high and ultrahigh vacuum ranges, but it is acceptable to use an RGA for vacuum conditions up to 10 millitorr.



Mass Flow Controller
(Photo courtesy of MKS Instruments, Inc.)

There are four basic parts to an RGA: an ionizer, an aperture, an analyzer, and a detector (see Figure 8.11). These are essentially the parts of a mass spectrometer, but an RGA is much smaller and is designed for attaching to process tools. These four parts are located in the sensing head of the RGA behind a process-specific valve inlet attached directly to the process chamber.

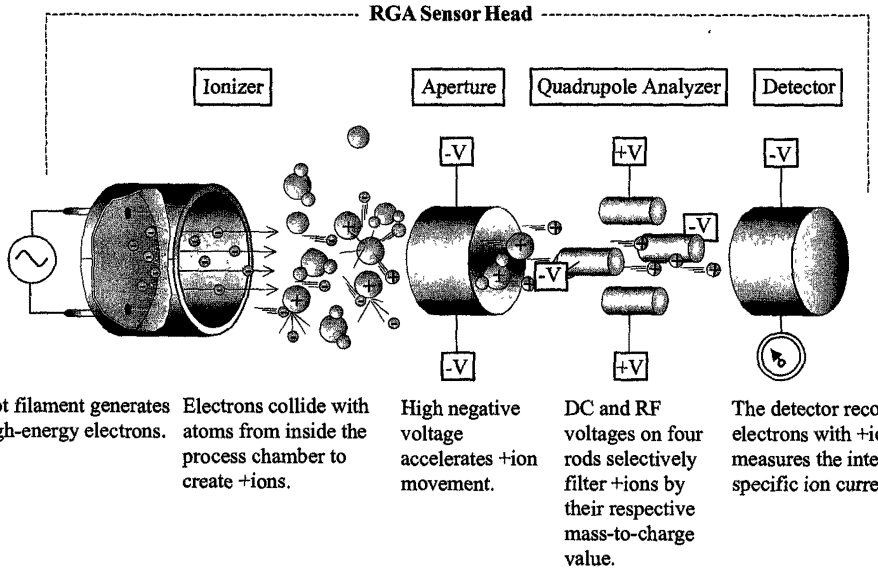


FIGURE 8.11 Basic Parts of Residual Gas Analyzer (RGA)

The ionizer creates ions from the gas molecules present in the system by bombarding the molecules in the chamber with electrons generated in the ionizer. The bombardment removes an electron from the gas molecules and creates positive ions. These ions are directed toward the analyzer by applying a voltage potential or magnetic field in the aperture. The analyzer is where the ions are separated by mass. There are several different types of analyzers; a common type is the *quadrupole mass analyzer* (*QMA*). It consists of four cylindrical rods that have both a constant DC potential and a high-frequency RF component (see Figure 8.12). For a given voltage level applied to the cylinders, only ions of a given atomic mass-to-charge pass through the filter. All other ions are grounded on the cylinder rods. The detector on modern RGAs will have better than 1 atomic mass unit (amu) resolution to differentiate between the various ions. As voltages are progressively changed on the filter, different ions are passed through the cylinders. In this manner, various gases of different masses are separated and identified.

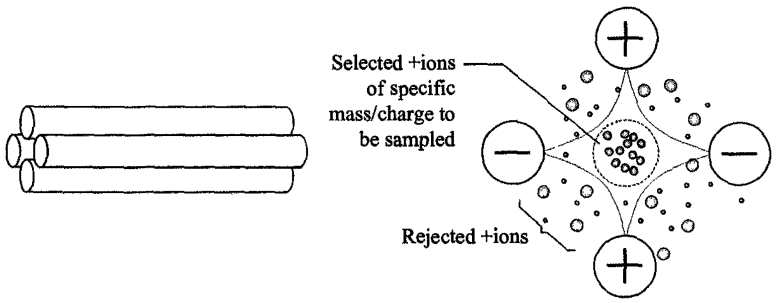


FIGURE 8.12 Quadrupole Mass Filter

RGA as Real-Time Monitor

RGA data can be an important part of verifying that a process chamber is ready for processing, especially for large-diameter wafers (see Figure 8.13 on page 192). This is because RGA data can provide real-time information about the cleanliness and stability of the process chamber during pumpdown.

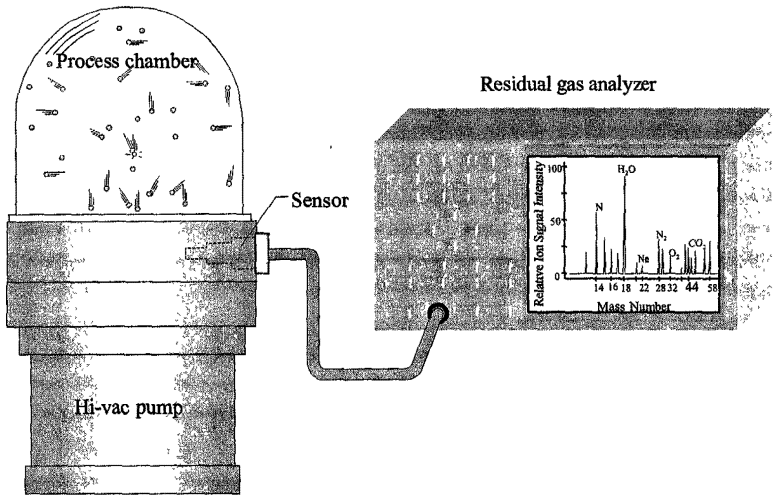


FIGURE 8.13 RGA Data Output

Using the RGA as an in situ (real-time) process monitor is a recent development. While the wafer is undergoing processing, the RGA will directly monitor the chemical constituents in the chamber to verify that the process is proceeding correctly.[9] The RGA can be used to replay conditions in the event of a wafer quality problem. This replay ability allows for more rapid diagnosis of process problems.

There is a growing need for RGAs to monitor chemical processes, such as plasma etching and plasma-enhanced chemical vapor deposition (PECVD.[10] For plasma applications, an RGA can follow the different process chemicals as the reaction occurs in the chamber and give insight into plasma behavior and how the chemical species change over time.

PLASMA

Plasma is a neutral, highly energized, ionized gas consisting of neutral atoms or molecules, positive ions, and free electrons. Positive ions and free electrons are formed when a valence electron is removed from a neutral atom. For instance, fluorine is neutral when there are an equal number of protons and electrons in its atomic structure. Fluorine is ionized when an electron is separated from its host atom (see Figure 8.14). Ionization of gas atoms in a confined process chamber can occur by subjecting the gas to strong DC or AC electromagnetic fields or by bombarding the gas atoms with some sort of electron source. These methods will be covered later in greater detail.

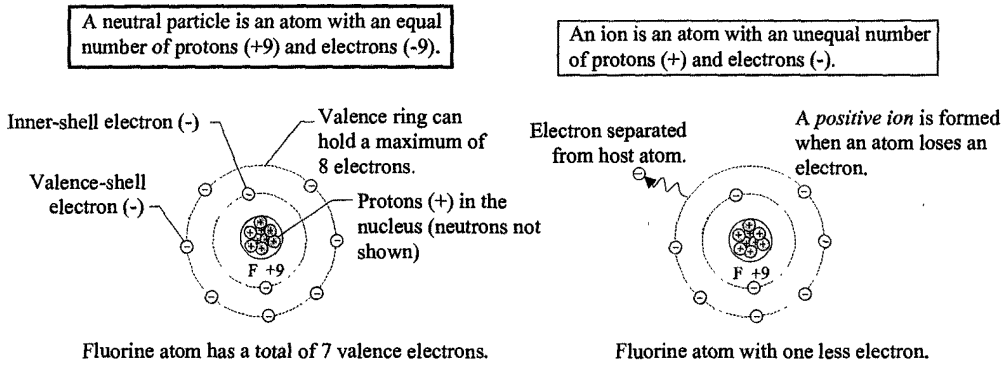


FIGURE 8.14 Creation of an Ion

Plasma is used at various process steps during wafer fabrication because it supplies much of the energy needed to support a gas reaction near the wafer surface in a process chamber. For example, plasma is used in lieu of thermal energy to ionize and excite a source gas to deposit thin films in high-density plasma chemical vapor deposition (HDP-CVD), discussed in Chapter 11. Another plasma application is to selectively remove metal through plasma etching (see Chapter 16). The most common indication that a plasma exists in a process chamber is the characteristic observable light referred to as a glow discharge (see Figure 8.15).

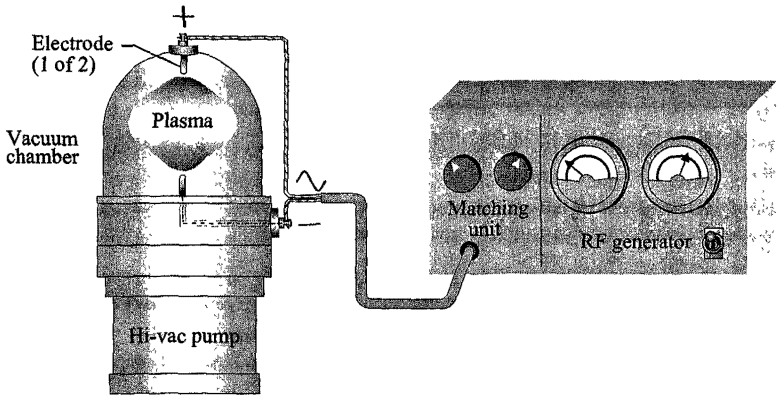


FIGURE 8.15 Plasma Glow Discharge

Glow Discharge

Although a glow discharge across a gas mixture can be created by applying DC power, the most common methods use AC power in the radio frequency (RF) range. When the electrical power is first applied, free electrons in the gas mixture are greatly influenced by the presence of electric fields. Electrons are accelerated through the gas mixture and collide with atoms and molecules, releasing additional electrons during the collisions.

In a weakly ionized plasma, or glow discharge, such as that commonly used in wafer fabrication, the high-energy electrons collide with neutral atoms or molecules and excite them. These excited atoms or molecules are short-lived, with lifetimes measured in nanoseconds. When an excited atom or molecule returns to its lowest energy state, energy is released during this relaxation in the form of a radiated photon (or light). This release causes the characteristic glow in the glow discharge (see Figure 8.16), with different glow colors for different species (e.g., oxygen, nitrogen, fluorine, and so on). The high-energy electrons in a plasma transfer energy to the neutral atoms and molecules by impact and initiate a reaction which could not have occurred in a high pressure environment. Typical parameters sustaining a glow discharge include the RF power and frequency, pressure, gas mixture and flow rate, vacuum pumping speed, and surface temperature.

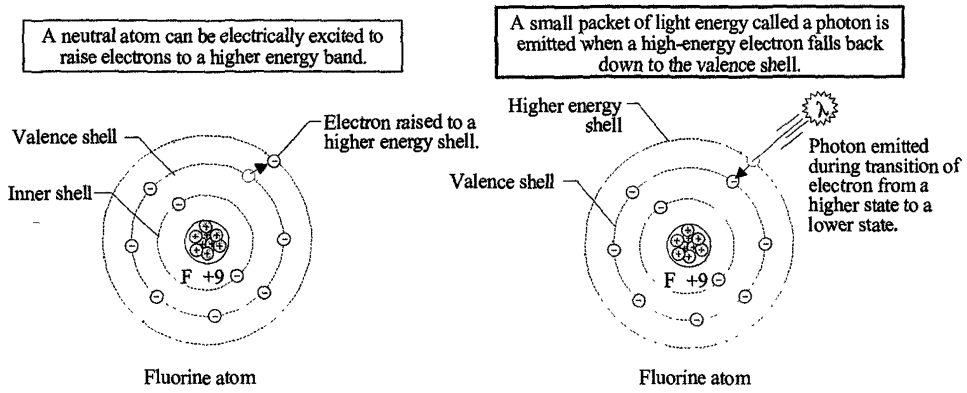


FIGURE 8.16 Electrically Exciting and Relaxing an Atom

Radicals ■ Radicals are highly reactive chemical species. As electrons collide with atoms or molecules, different kinds of products are created. The dissociation (or fragmentation) of a molecule occurs when a large molecule is broken into smaller particles (see Figure 8.17). Recall that the number of electrons in the valence shell are related to the chemical properties of an atom. A *radical* is created when a neutral molecule is bombarded by an energetic electron. This action cleaves a bond without adding or removing an electron, creating a highly reactive species. This radical is an uncharged atom or molecular fragment that has incomplete bonding (or unpaired electrons) in the valence shell.

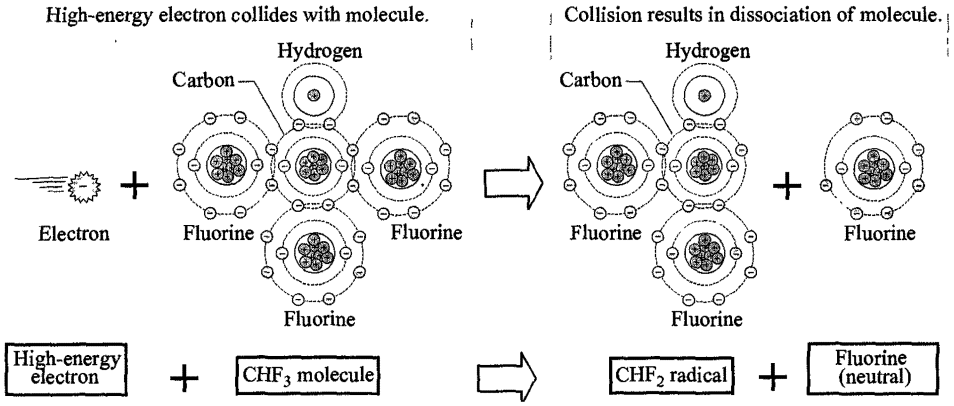


FIGURE 8.17 Dissociation of a Molecule

RF Energy ■ The energy of plasma is sustained through the absorption of RF radiation in an applied alternating current (AC) electric field of several hundred volts rms. The use of RF power, typically at 13.56 MHz (chosen for industrial purposes by the FCC, or Federal Communication Commission), creates a high-efficiency plasma. In recent years, different frequencies have started being used for plasma generation, such as 400 KHz, 2 MHz, 4 MHz, and 2.45 GHz. The frequency has a direct influence on the ion mobility, thus affecting process uniformity and the rate of the process (e.g., the metal removal rate for etching). The RF field is applied between a negative cathode and a positive anode, referred to as electrodes. The wafer typically will be held against the grounded electrode, with the RF power applied to a parallel electrode (see Figure 8.18).

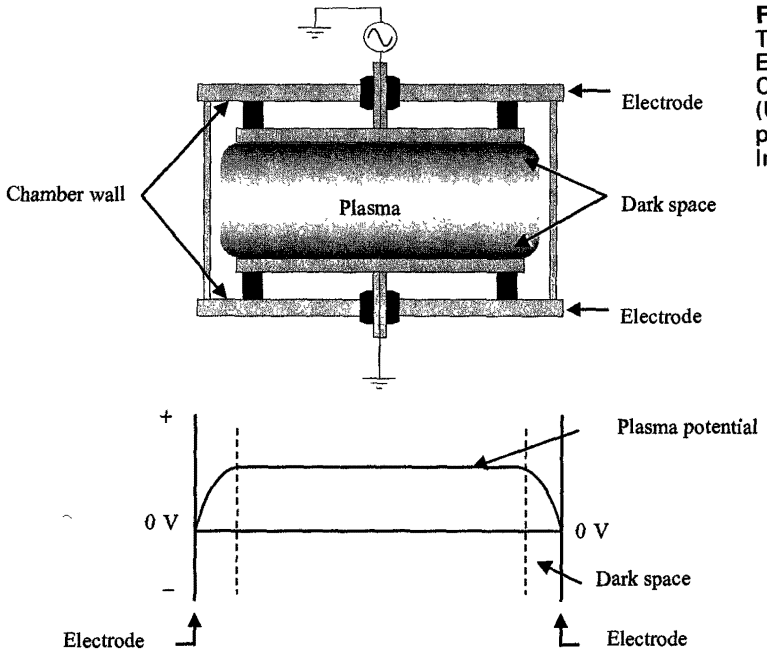


FIGURE 8.18 Typical Plasma Electrode Configuration (Used with permission from International SEMATECH)

The electrons, ions, and other species generated in a glow discharge will move toward the electrodes. Electrons in the glow discharge can move faster toward the positive electrode when compared to the slow moving massive positive ions. This movement produces a positive potential in the glow discharge region. The positive ions move toward the cathode and cross through a *dark space,* or ion sheath, adjacent to the cathode. The dark space results from the lack of electrons in this region. With respect to the glow discharge region, the dark space has a large voltage drop and strong electric field. This condition accelerates the positive ions toward the cathode, which then ejects secondary electrons. These secondary electrons are repelled from the cathode and cross back through the dark space where they maintain the glow discharge.

PROCESS CHAMBER CONTAMINATION

Wafer fabrication equipment with vacuum can be viewed as two general types: tools that process large batches of wafers (e.g., 150 or 200 wafers) or single-wafer, multichamber tools with loadlocks and several integrated process chambers. In general, chamber contamination on wafers is lower in single-wafer reactors than in large batch tools. The improvement in single-wafer cluster tools is attributed mostly to maintaining the processing chambers under relatively constant temperature and pressure conditions, which minimizes factors that contribute to particle generation.

Trace amounts of water are probably the most significant source of contamination in process chambers with vacuum.[11] Water is a problem in vacuum systems since it is reduced to ions as well as H_2 or O_2 gas molecules because of the reactor chemistry and vacuum. It sticks tenaciously to surfaces, outgasses slowly, and is a chemical poison in reactors. For example, the tenacity of water sticking to chamber walls is so high that it is estimated that there are 10,000 times more water molecules on the chamber walls than in the vacuum space of one cubic meter of processing chamber. It is also believed that water creates particle contamination during pumpdown by the formation of particles.[12]

In order to minimize water adsorption contamination in process chambers, it is desirable to reduce the need to open or disassemble process reactors for cleaning purposes. This goal is often accomplished by using in situ cleaning techniques (cleaning while the wafers are processed). Nevertheless, process equipment still requires maintenance actions, which necessitate equipment shutdown and repair. The specific steps followed by the maintenance technician during servicing have a significant effect on the control of contamination in the equipment. Recommendations for minimizing contamination during the servicing of wafer fabrication equipment are listed in Table 8.4.[13]

TABLE 8.4 Recommendations to Minimize Contamination During Equipment Servicing

Recommendations
1. Maintain good temperature and humidity control in the cleanroom environment where the equipment is located.
2. Control the equipment's pump and vent cycles to minimize turbulence and prevent particle generation when processing wafers.
3. Avoid abrasive cleaning materials.
4. Use exact replacement parts and materials to avoid subtle sources of equipment contamination.
5. Use low particle-generating gas-handling components, such as regulators and automatic valves that have a tendency to generate particles.

SUMMARY

Process chambers have many functions in wafer fabrication, such as controlling gas flow. A vacuum in a process chamber creates a clean environment with low molecular density and high mean free path. Vacuum ranges are low, medium, high, and ultrahigh. The most common vacuum pumps in the wafer fab generally can be categorized as roughing (dry mechanical and blower/booster) and high vacuum (turbo pump and cryopump). The mechanical pump is a low vacuum pump. The blower is a low to medium vacuum that has high throughput. The turbo pump is a common high vacuum pump that uses turbine blades, while the cryopump is a capture pump that removes gases by freezing them. Gas flow in the process chamber is controlled using a mass flow controller (MFC). The residual gas analyzer (RGA) is used to detect leaks or to analyze residual gas in a chamber after vacuum pumpdown. It also can do real-time monitoring of the reaction in a chamber during a process run. Plasma is a highly energized gas that is commonly used in process chambers to excite source gases during a reaction. It involves a glow discharge that results from collisions between energetic electrons that produce radicals. Plasma energy is sustained by the absorption of RF radiation. Cleanliness in the process chamber is critical, with water moisture being the most serious contaminant.

KEY TERMS

process chamber
reactor
cluster tool
vacuum
torr
low vacuum (rough vacuum)
medium vacuum
high vacuum
ultrahigh vacuum
mean free path (MFP)
roughing pump
high vacuum pump
dry mechanical pump
blower (booster pump or lobe pump)
turbomolecular pump (turbo pump)
cryopump
cryoarrays
regeneration
loadlock
throughput
pump speed
mass flow controller (MFC)
residual gas analyzer (RGA)
quadrupole mass analyzer (QMA)
plasma
glow discharge
radical
dark space

REVIEW QUESTIONS

1. What is a process chamber? What are its five functions?
2. Describe a cluster tool, and explain why it is beneficial in IC fabrication.
3. What is a vacuum?
4. What are the benefits of vacuum in semiconductor manufacturing?
5. State the most common vacuum unit, and describe how to interpret it.
6. List and describe four vacuum ranges.
7. What is mean free path? Why is this condition important?
8. Give reasons for using a roughing pump and a high vacuum pump.
9. Describe two types of roughing pumps. Which one has the higher gas throughput?
10. Describe two types of high vacuum pumps.
11. What are the most common reasons for failure of a turbo pump?
12. Describe the purpose of the cryoarray, and explain how it creates pumping action.
13. What is regeneration?

14. Draw a picture of a cluster tool and show chambers with at least three different vacuum levels.
15. What is the purpose of the loadlock in a cluster tool?
16. State the four basic process chamber requirements for gas flow.
17. Describe gas throughput.
18. What is pump speed?
19. What is the purpose of a mass flow controller?
20. What does a residual gas analyzer (RGA) do?
21. List and describe the three basic parts to an RGA.
22. How does the quadrupole mass analyzer work on an RGA?
23. What is plasma? How is plasma beneficial in a process chamber?
24. Describe the plasma glow discharge region.
25. What is a radical in a plasma?
26. Why is RF energy used in plasma?
27. Why is moisture a problem in process chambers?
28. List the recommended steps to minimize contamination during equipment servicing.

VACUUM EQUIPMENT SUPPLIERS' WEB SITES

Alberta University Vacuum Page	http://nyquist.ee.ualberta.ca/~schmaus/vacf/
Alcatel Vacuum Products	http://www.alcatel.com/
Apiezon Products	http://www.apiezon.com/
AVS, American Vacuum Society	http://www.vacuum.org/
BOC Edwards	http://www.boc.com/edwards/
CTI Cryogenics	http://www.ctivacuum.com/
Ebara Technologies	http://www.ebaratech.com/
Granville-Phillips	http://www.helixtechnology.com/
Inficon Inc.	http://www.leyboldinficon.com/
Leybold Vacuum	http://www.leyboldvac.de/
Millipore Corp.	http://www.millipore.com/
MKS Instruments	http://www.mksinst.com/
Omega Engineering Inc.	http://www.omega.com/
Osaka Vacuum Ltd.	http://www.osakavacuum.com/
Parker Hanniflin Corp.	http://www.veriflo.com
Pfeiffer Vacuum Tech., Inc.	http://www.pfeiffer-vacuum.com/
SEMI	http://www.semi.org/
Unit Instruments	http://www.unit.com/
Varian Inc.	http://www.varianinc.com/
Varian Vacuum Technologies	http://www.varianinc.com/vacuum/
VAT Valve	http://www.vatvalve.com/
Veeco Instruments Inc.	http://www.veeco.com/

REFERENCES

1. R. Waits, "Semiconductor Manufacturing and Vacuum Technology: A Memoir," *Solid State Technology* (May 1997): p. 105.
2. H. Tompkins, *The Fundamentals of Vacuum Technology,* (New York: American Vacuum Society, 1997), p. 6.
3. C. Tilford and J. P. Looney, "Vacuum Measurement: The Basics," *Semiconductor International* (May 1994): p. 73.
4. Ibid.
5. J. Baliga, "Vacuum Pump Designs Adjust to Harsher Conditions," *Semiconductor International* (October 1997): p. 88.
6. S. Hansen, *Introduction to the Creation and Control of the Vacuum Process Environment,* (Andover: MKS Instruments, 1995), p. 118.

7. Ibid., p. 110.
8. L. Peters, "Residual Gas Analysis: A Technology at a Crossroads," *Semiconductor International* (October 1997): p. 95.
9. T. Banks, G. Diamond, and S. Ruck, "Integrating Mass Spectrometry Data into the Fab Environment," *Semiconductor International* (June 1997): p. 138.
10. Ibid., p. 98.
11. A. Rapa and A. Bross, "Contamination Control in Multilevel Interconnection Manufacturing," *Handbook of Semiconductor Interconnection Technology,* ed. G. Schwartz, K. Srikrishnan, and A. Bross, (New York: Marcel Dekker, 1998), p. 552.
12. Ibid.
13. Ibid., p. 553.

CHAPTER 9

IC FABRICATION PROCESS OVERVIEW

The typical IC wafer fab process may take as many as six to eight weeks and involve 350 or more process steps to complete the full manufacturing flow. The complexity of this process can be overwhelming.

Recall that most semiconductor processing occurs in the top few microns of a silicon wafer. This activity corresponds to the first part (front end) of the manufacturing process flow. All materials on top of the silicon are part of the layering strategy needed to interconnect the many devices on the chip. To add multiple metal and insulating layers, the process flow requires the wafer to cycle through the different process steps. Once you understand the process, you will realize that only a few process areas are used many times over to fabricate a high-performance microchip.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Draw a diagram showing how a typical wafer flows in a sub-micron CMOS IC fab.
2. Give an overview of the six major process areas and the sort/test area in the wafer fab.
3. Describe the primary purpose of each of the 14 CMOS manufacturing steps.
4. Discuss the key process and equipment used in each CMOS manufacturing step.

INTRODUCTION

This chapter describes a simple manufacturing process for producing 0.18-μm CMOS integrated circuits on a silicon wafer. This is presented to help the reader gain a better understanding and appreciation for semiconductor manufacturing. Specific details about each process step is provided later in this text in the process chapters.

An overview of the entire wafer fab process is first presented, with a model that depicts how wafers repeatedly cycle through a few major process areas. This explanation simplifies the concept of keeping wafer flow to a manageable level in a fab. Recognize that when changes are made to the process, such as parameter or tool changes, the results of the change may not be known until many days or weeks later during wafer test at the end of the process. This fabrication complexity makes it important for each process area to flawlessly perform its tasks by meeting the defined quality measures at each step.

CMOS PROCESS FLOW

IC manufacturing is a complicated sequence of chemical and physical operations that are performed on a silicon wafer. Simply stated, the operations fall into four basic categories: layering, patterning, etching, and doping. Figure 9.1 on page 200 illustrates how complex the process can be even for manufacturing a single MOS transistor.

Since CMOS technology is the most popular of the process families, we have chosen it as an example to describe wafer process flow. The specific example is that of a 0.18 μm CMOS integrated circuit process flow. And, since this is an overview of IC manufacturing, you will be introduced to a variety of terms and concepts, which will be further explained in subsequent chapters. As you study this chapter, keep in mind the manufacturing area where specific process operations are being performed. Note the purpose of each operation, the type of equipment and materials that are used, and the quality measures that are followed to determine the integrity of the process at each step.

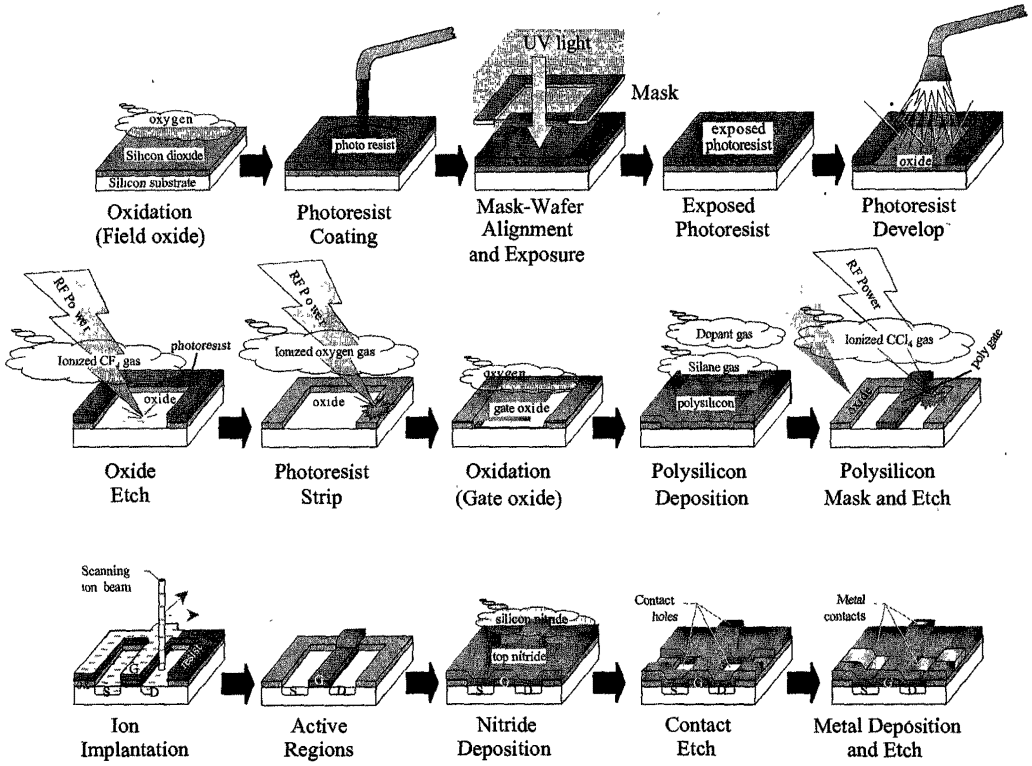


FIGURE 9.1 Major Fabrication Steps in MOS Process Flow
(Used with permission from Advanced Micro Devices)

Overview of Areas in a Wafer Fab

ICs are manufactured in a wafer fab. As shown in Figure 9.2, wafer fabs are generally divided into six distinct production areas: diffusion (which includes oxidation, film deposition, and doping processes), photolithography, etch, thin films, ion implant, and polish. These six major production areas and their related process and metrology tools are all housed in the ultraclean cleanrooms of the wafer fab. The polish area is a new addition to semiconductor manufacturing for high-performance ICs and is gaining popularity in the industry. Although located near the wafer fab, the test/sort area for testing individual die on the wafer is not housed in the same cleanroom environment as the other areas of the fab. Assembly and packaging plants are generally located in other facilities, perhaps even in other countries.

Diffusion ■ The *diffusion* bay is recognized as the area where high-temperature processing and film depositions are performed. The primary tools in the diffusion area are a *high-temperature diffusion furnace* and a *wet cleaning station.* High-temperature diffusion furnaces (see Figure 9.3) can operate at temperatures near 1200°C and are configured to run a variety of processes, including oxidation, diffusion, deposition, anneals, and alloys. These processes will be covered in detail in later chapters. Wet cleaning stations are the secondary tools used in the diffusion area. Wafers must be cleaned thoroughly to remove contamination and native-grown oxide on the surface before inserting the wafers into the furnaces.

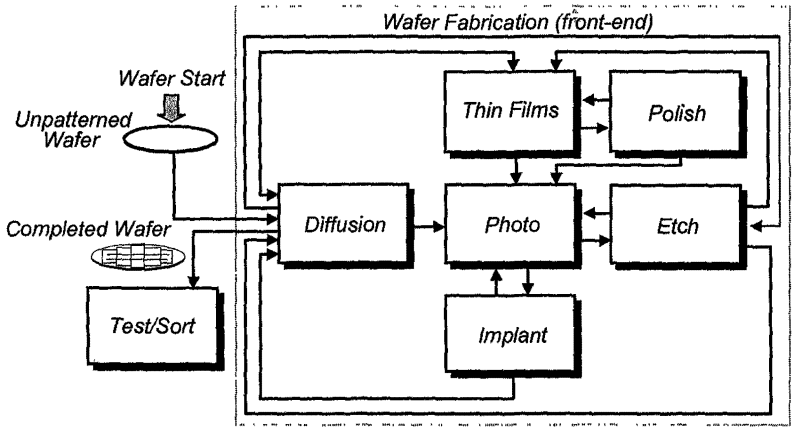


FIGURE 9.2 Model of Typical Wafer Flow in a Sub-Micron CMOS IC Fab (Used with permission from Advanced Micro Devices)

Photolithography ■ The *photolithography* bay is recognizably different than the other areas in the fab due to the yellow fluorescent tubes that light up this production bay. The purpose of photolithography is to photograph the image of a circuit pattern onto the photoresist that coats the wafer surface. Photoresist is a light-sensitive chemical that captures the image of a mask pattern resulting from exposure to ultraviolet (UV) light. Photoresist is sensitive to certain wavelengths of light, such as UV and white light, but the wavelengths of yellow light do not affect it.

The *coater/developer track* is a cluster tool used to perform many of the operations in photolithography. This tool primes the wafer, coats it with photoresist, spins the wafer to smooth out the photoresist, bakes the wafer, and uses robotics to transfer the resist-coated wafer to the alignment and exposure tool. The purpose of the *stepper* is to align the wafer to an array of die patterns etched on a chrome-coated quartz reticle. When properly aligned and focused, the stepper exposes a small area of the wafer, then steps to the next field and repeats the process until the entire wafer surface has been exposed to the die patterns on the reticle (see Figure 9.4 on page 202). When completed, the wafer returns to the coater/developer tracks where the resist is developed, then the wafer is rinsed and baked again.

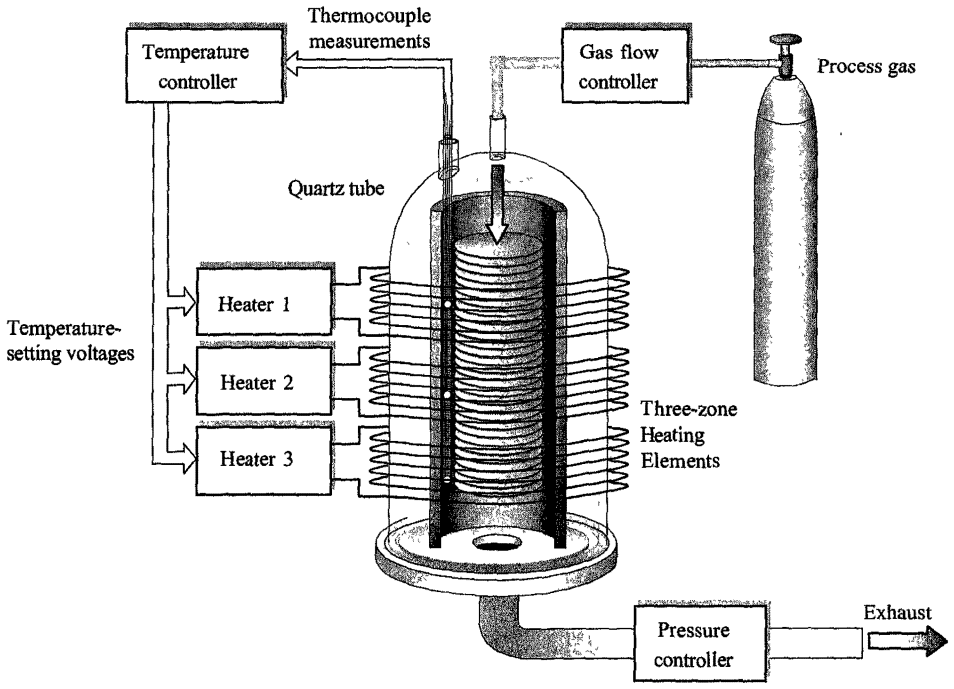
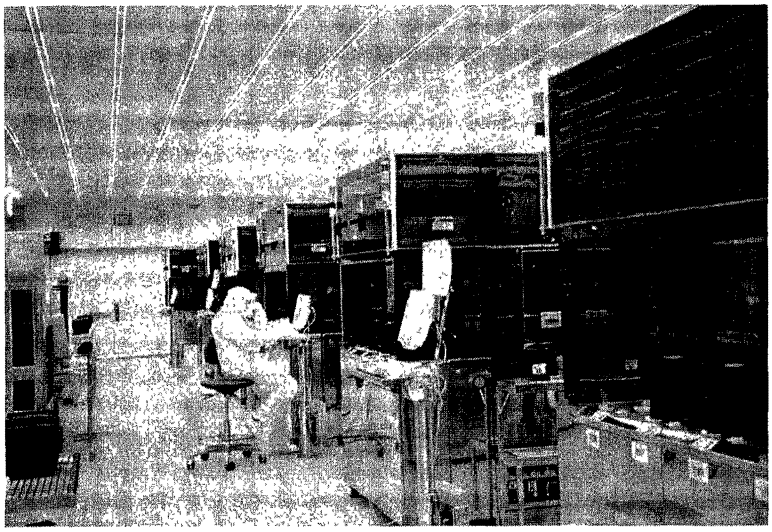


FIGURE 9.3 Simplified Schematic of a High-Temperature Furnace



Photolithography Bay in a Sub-Micron Fab
(Photo courtesy of Advanced Micro Devices)

Referring once again to Figure 9.2 on page 201, photolithography is shown to be at the center of the wafer fab. This position is due to the fact that wafers flow into photolithography from all other areas of the fab. Contamination control is especially important here because particles or defects can become imbedded in or on the resist films during the photographic process. A defect in the photographic mask or reticle or a particle on the stepper can be photographed onto all wafers that are processed with these tools.

To reduce contamination, open containers of chemicals are prohibited from being used in this area. Thus, cleaning stations and photoresist strippers are usually located in other areas of the fab other than photolithography. Note that in Figure 9.2 on page 201, wafers flow from photolithography into only two other areas: etch and ion implant. Consequently, there are only three production areas where photoresist-coated wafers can be found—photo, etch, and ion implant.

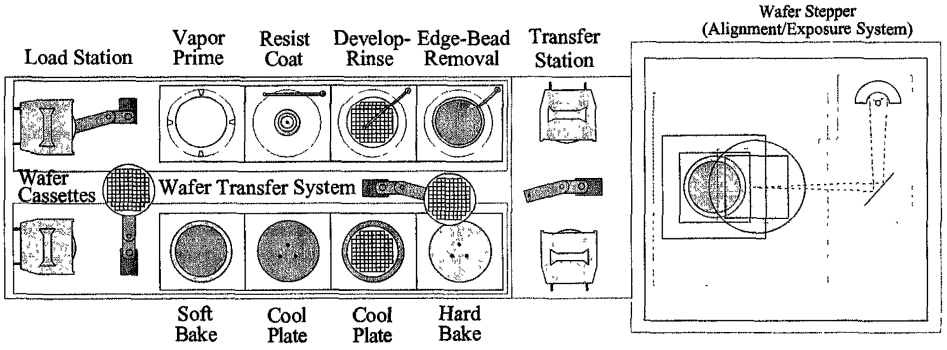


FIGURE 9.4 Simplified Schematic of a Photolithography Processing Module

Etch ■ The *etch* process creates a permanent pattern on the wafer in areas not protected by the photoresist pattern. The most common tools in the etch bay are the *plasma etcher,* the *plasma resist stripper,* and the wet cleaning station. Today most etching steps are done with dry plasma etchers (see Figure 9.5), although some wet etch processes are still in use. The plasma etchers are tools that use radio frequency (RF) energy to ionize gas molecules inside a vacuum chamber. The plasma is the glow given off by the electrically-excited gases. The gases react with the top layer of material on the wafer. After the etch process, another plasma system, called the plasma stripper, uses ionized oxygen to remove the photoresist from the wafer. This step is followed by thoroughly cleaning the wafers with a combination of chemicals.

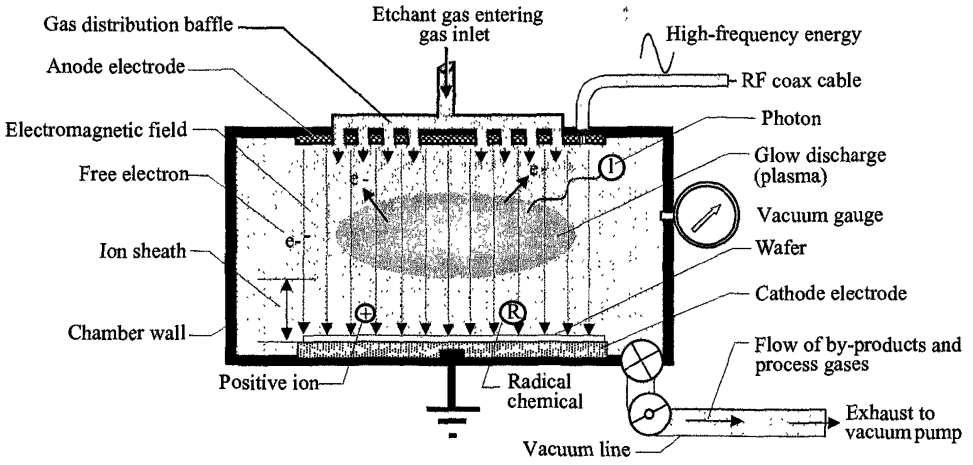


FIGURE 9.5 Simplified Schematic of a Dry Plasma Etcher

Ion Implant ■ The *ion implanter* is the most popular tool for doping wafers in a sub-micron process. Gases carrying the desired dopant, e.g., arsenic (As), phosphorus (P), and boron (B), are ionized inside the implanter (see Figure 9.6). High voltages and magnetic fields are used to control and accelerate the ions. The high energy of the dopants penetrates through the surface of the resist-coated wafer. After implantation is completed, the photoresist is stripped off and the wafer is thoroughly cleaned.

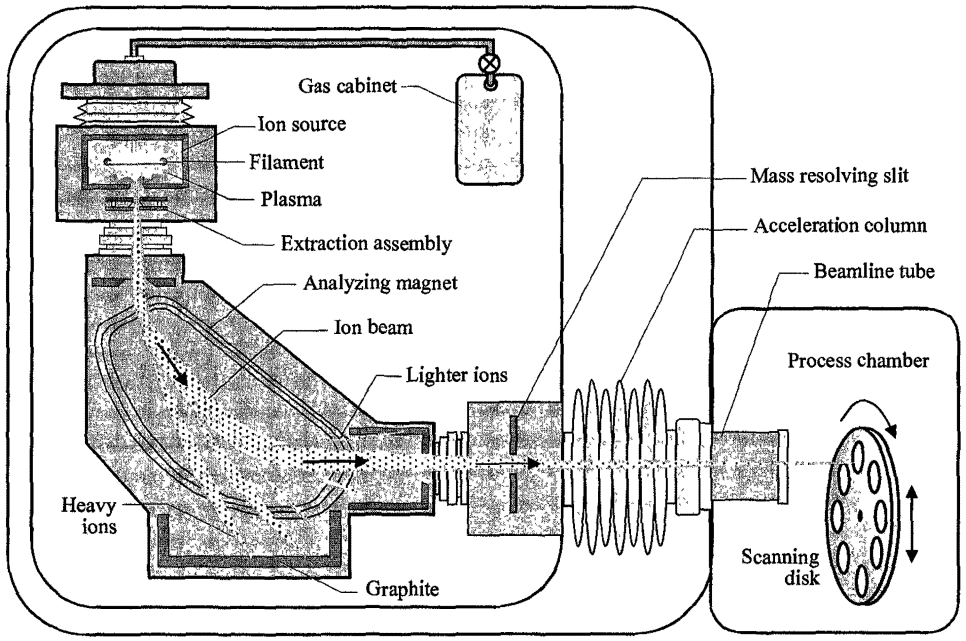


FIGURE 9.6 Simplified Schematic of an Ion Implanter

Thin Films ■ The *thin films* bay is primarily responsible for depositing dielectric and metal layers during different steps of the manufacturing process. The processes in thin films operate at lower temperatures than the furnaces in the diffusion area. There are many diverse tools in this area. All thin film deposition equipment operates under low to medium vacuum conditions (see Figure 9.7 on page 204), including chemical vapor deposition (CVD) and metal sputtering tools (PVD, or physical vapor deposition). Other tools used in this area may include the spin-on-glass (SOG) system,

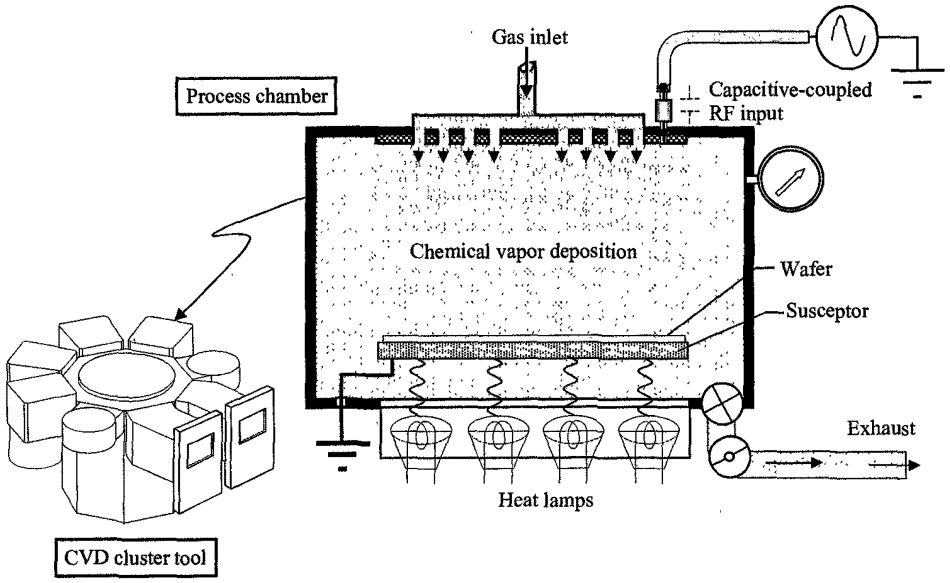
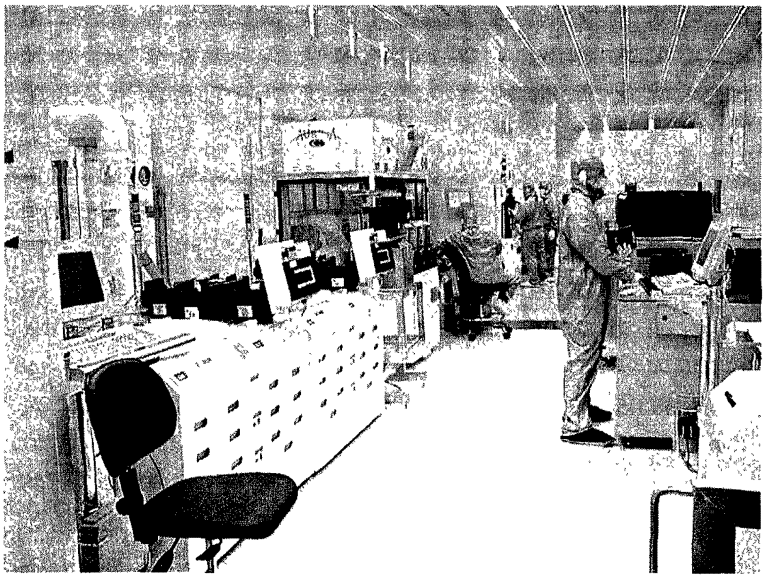


FIGURE 9.7 Simplified Schematics of a CVD Cluster Tool and Process Chamber

rapid thermal processor (RTP) system, and wet cleaning station. The SOG is used to fill in low areas of the wafer for the purpose of planarizing (smoothing) the surface of the wafer. RTAs are used to anneal ion implant damage of the silicon substrate and for metal alloy processing steps.

Polish ■ The purpose of the CMP (*chemical mechanical planarization*) process is to planarize the top surface of the wafer by lowering the high topography to be level with the lower surface areas of the wafer. CMP minimizes uneven wafer surfaces that make further processing difficult. The polisher is the primary tool in the CMP area, and this process is also referred to as *polish*. CMP uses a combination of chemical etching and mechanical abrading to remove a desired amount of the upper layer of the wafer. Other tools that support CMP include *wafer scrubbers,* cleaning stations, and metrology tools.



Polish Area in a Sub-Micron Wafer Fab
(Photo courtesy of Advanced Micro Devices)



Thin Film Metallization Work Bay
(Photo courtesy of Advanced Micro Devices)

CMOS MANUFACTURING STEPS

The remainder of this chapter focuses on the manufacturing steps that a wafer follows in a typical CMOS process. Although there are some operations in the process flow that can be performed in batches of wafers at one time, this process flow describes the manufacturing of a single wafer. For further simplification, this process describes the production of only a microscopic area of a single CMOS inverter consisting of two transistors—nMOS and pMOS. Cross-sectional views of the product will be provided for each major operation.

The CMOS manufacturing steps to be described are:

1. Twin Well Process
2. Shallow Trench Isolation Process
3. Poly Gate Structural Process
4. Lightly Doped Drain (LDD) Implant Processes
5. Sidewall Spacer Formation
6. Source/Drain (S/D) Implant Processes
7. Contact Formation
8. Local Interconnect Processes
9. Via-1 and Plug-1 Formation
10. Metal-1 Interconnect Formation
11. Via-2 and Plug-1 Formation
12. Metal-2 Interconnect Formation
13. Metal-3 to Pad Etch and Alloy
14. Parametric Testing

1. Twin Well Process

The first step in some CMOS wafer fabrication processes is to define the active regions of the MOSFETs. Many of today's sub-quarter micron processes use a *twin-well* (also referred to as *twin-tub*) approach to define the active regions of the nMOS and pMOS transistors. A twin-well consists of a p-well and an n-well, with each well requiring at least three to five major processing steps to fabricate. A *retrograde implant* technique is used to obtain optimum electrical characteristics for the FETs. This technique starts with a high-energy, high-dose implant that penetrates approximately 1 μm into the epilayer. Subsequent well implants are done at the same active site with a progressive

decrease in energy, junction depth, and dose of dopant material such as phosphorous or boron (measured in atoms/cm^2). The results help set the threshold operating voltage for the FETs as well as prevent common CMOS problems, such as *latchup* and other reliability issues. Figure 9.8 and Figure 9.9 illustrate, respectively, the major steps required in the formation of the n-well and p-well.

n-well Formation ■ The five major steps in n-well formation are described below in the following process and illustrated in Figure 9.8.

Description of Five Major Steps in n-well Formation

	Process Step	Description
1.	Epitaxial Growth	The silicon wafer, having already a thin layer of *epitaxial* silicon (epilayer), arrives in the diffusion bay. The epilayer has the exact crystal structure as the substrate, except with improved purity and fewer crystal defects. The epilayer has already been lightly doped with p-type dopant (boron).
2.	Initial Oxide Growth	In diffusion the wafer is cleaned in a series of chemical baths to remove particles, organic and inorganic contamination, and *native oxide* growth (natural silicon dioxide growth) from the wafers. After the wafer is rinsed and dried, the wafer is placed in a high temperature (~1000 °C) furnace process chamber. Oxygen is flowed into the process chamber to react with silicon to grow approximately 150 Å of oxide. This initial oxide serves several functions: (1) it protects the top surface of the epi silicon from contamination, (2) it prevents excessive damage to the silicon during implantation, and (3) as a screen oxide, it helps control the depth of the dopants during implantation.
3.	1st Mask, n-well Implant	In photolithography the wafer undergoes a series of process steps in the coater/developer track tool. The "tracks," as commonly referred to, prime the top surface of the wafer, coat the wafer with photoresist (liquid photographic film), spin the wafer, and bake the wafer. An internal automatic wafer-handling system transfers the wafer between process stations within the tracks. Another wafer handler removes resist-coated wafers, one at a time, and transfers them to the *alignment-and-exposure system* (an extremely complex and precise camera). The alignment-and-exposure system photographs the image of a specific mask layer directly onto the resist-coated wafer. In this case, the mask layer defines the areas of the product that are to be implanted in order to form the n-wells for the pMOS transistors. The exposed wafer is then transferred back to the tracks, where the image first appears when the wafer is sprayed with the chemical developer. The developed wafer undergoes another bake step and is inspected before being transferred to the ion implant bay. The printed pattern is inspected for proper *linewidths,* called *critical dimensions* (*CD*). In the event of a major defect, the photoresist can be stripped off and the wafer can then be reworked. Photolithography is the only area in the fab where a wafer can be easily reworked.
4.	n-well Implant (High energy)	The patterned wafer arrives in the ion implant bay. The photoresist pattern covers specific areas of the wafer that are to be protected from ion implantation. Windows, or openings, in the photoresist allow high-energy positive dopant ions to penetrate into the upper surface of the epilayer (~1 μm *junction depth*). In this case, phosphorus is the desired dopant. The *ion implanter* is the main process tool in this bay. Its purpose is to ionize dopant atoms, accelerate them with high voltage (~200 KeV), select the most appropriate dopant species to implant, focus the ions into a narrow beam, and, finally, scan the wafer to provide uniform doping across all unprotected areas of the wafer. At this point the dopant ions penetrate the crystal lattice of the silicon causing damage to the covalent atomic structure. This damage will be repaired later in a diffusion and annealing step. *Note:* After each ion implant operation, an oxygen-plasma reactor tool strips the photoresist off each implanted wafer. The wafer is then cleaned by a wet chemical process to remove residual photoresist and polymers created by the plasma process. This note applies to all subsequent implant operations but will be listed only this one time.
5.	Anneal	The implanted wafer is transferred to diffusion where the wafer is cleaned before being inserted into an *anneal* furnace. Four things occur as a result of this anneal process. (1) A new barrier oxide layer is grown over the bare silicon. (2) The higher temperatures cause the movement of dopants further into the silicon (called diffusion). (3) The implant damage is repaired. (4) The bonds between the dopant atoms and the silicon atoms are activated, making the dopant atoms a part of the crystal lattice structure (electrical activation).

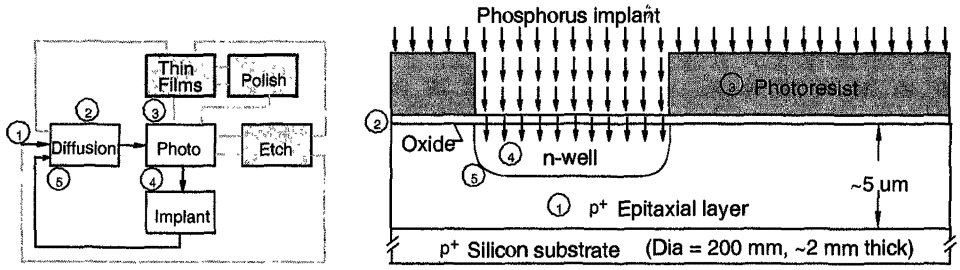


FIGURE 9.8 n-well Formation

p-well Formation ■ The three major steps in p-well formation are described in the following process and illustrated in Figure 9.9.

Description of Three Major Steps in p-well Formation

Process Step	Description
1. 2nd Mask, p-well Implant	The photolithography steps for the p-well implant mask are exactly the same as for the first mask. The only difference is the mask is the direct opposite of the n-well implant mask. Compare Figure 9.8 and Figure 9.9 to see the differences.
2. p-well Implant (High energy)	The retrograde p-well implant energy levels are considerably lower than the n-well. This condition is due to the difference in the masses of the elements being implanted. Comparing the atomic mass unit (amu) of boron (~11) to phosphorus (~31), boron is about one-third the mass of phosphorus. Thus, the energy needed to implant boron should only require about one-third of that used to implant phosphorus to the same junction depth.
3. Anneal	This anneal step is basically the same as the first annealing step.

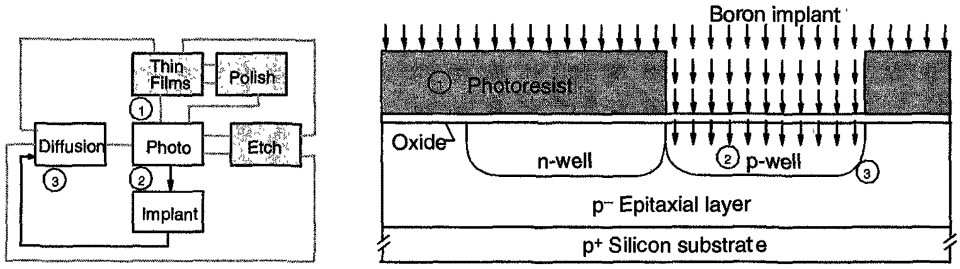


FIGURE 9.9 P-well Formation

2. Shallow Trench Isolation Process

Shallow trench isolation (STI) is an alternative method is for creating isolation regions between active transistor areas on a substrate.[1] This method is especially useful in the manufacturing of quarter-micron devices. The preferred isolation technique up until the late 1990s had been the LOCOS (local oxidation of silicon) technique that was instituted in the early 1970s.[2] Despite its complexity, STI is gaining in popularity with ULSI chip manufacturing. STI formation is explained as follows in three major steps: trench etch (see Figure 9.10 on page 208), oxide fill (see Figure 9.11 on page 209), and oxide polish (see Figure 9.12 on page 209).

STI Trench Etch ■ The four major steps in STI trench etch are described in the following process and illustrated in Figure 9.10.

Description of Four Major Steps in STI Trench Etch

	Process Step	Description
1.	Barrier Oxide	The wafer arrives in the diffusion bay where the wafer is again cleaned to remove contamination and oxide. After rinsing and drying, the wafer is inserted into a high-temperature oxidation furnace. Approximately 150 Å of new oxide is grown on the wafer. This layer will form a barrier to protect active regions from chemical contamination during nitride strip.
2.	Nitride Deposition	The wafer is inserted into a high-temperature (~750 °C) *LPCVD* (*low-pressure chemical vapor deposition*) furnace. Inside the process chamber ammonia and dichlorosilane gases react to form a thin layer of silicon nitride (nitride, Si_3N_4) on the wafer. The nitride will serve two functions throughout the STI formation process steps. (1) Nitride is a durable masking material, which protects active regions during the STI oxide deposition process and (2) nitride serves as a polish-stop material during the chemical mechanical planarization (CMP) step.
3.	3rd Mask, STI	The wafer arrives in photolithography from diffusion. This photolithography process step is similar to the previous masking step, except a different mask is used. This masking step is much more critical than the first mask due to the smaller dimensions. Inspection of the wafer includes CD measurements as well as checking for *defect inspection* (*DI*), and optical *visual inspection* (*VI*). These measurements require checking the accuracy of the alignment-exposure system relative to the written specifications.
4.	STI Trench Etch	The photoresist pattern is designed to protect areas of the silicon that are not to be etched. The windows in the photoresist allow ions and highly reactive *radical chemicals* to etch nitride, oxide, and silicon in unprotected areas. The preferred tool for etching these deep trenches is the *dry plasma etcher.* The etcher uses high-powered radio frequency (RF) energy to ionize a fluorine- or chlorine-based gas inside a vacuum process chamber. The RF energy dissociates molecules and ionizes atoms to create a chamber filled with a variety of plasma components. These plasma components provide the chemical and physical etching that results in the removal of silicon in locations designated as *isolation regions.* The slanted profile and the rounded bottom of the trenches improve the filling process and the electrical characteristics of the isolation structure. *Note:* Following completion of each etch operation, the wafer is stripped of photoresist and wet cleaned in a series of chemical baths. Key inspection procedures include measurements that verify proper step height (Å), etch rate (Å /min), CD, and DI. This note applies to all subsequent etch operations but will be listed only this one time.

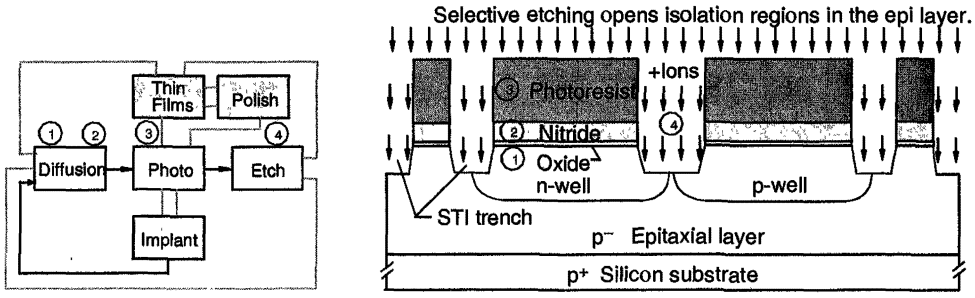


FIGURE 9.10 STI Trench Etch

STI Oxide Fill ■ The basic steps in STI oxide fill are described in the following process and illustrated in Figure 9.11.

Description of Basic Steps in STI Oxide Fill

	Process Step	Description
1.	Trench Liner Oxide	The wafer is again cleaned in the diffusion bay to remove contamination and oxide. After rinsing and drying, the wafer is inserted into a high-temperature oxidation furnace. Approximately 150 Å of oxide grows in the exposed walls of the isolation trenches. The nitride mask prevents oxygen diffusion into the active regions. The purpose of the liner oxide is to improve the interface between the silicon and the trench CVD oxide.
2.	Trench Fill with CVD Oxide	This deposition process is performed either in diffusion using an LPCVD furnace or in thin films using a variety of oxide CVD systems. Higher throughput and deposition rates are obtained with the furnace.

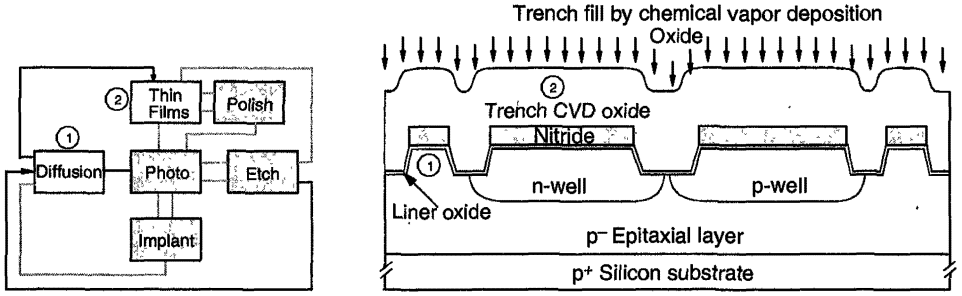


FIGURE 9.11 STI Oxide Fill

STI Oxide Polish-Nitride Strip ■ Planarization of the wafer surface is accomplished by several methods. In the past, filling the gap to planarize the wafer could be done using *spin-on-glass* (*SOG*), consisting of 80% solvent and 20% silicon dioxide. Baking the SOG after the deposition evaporated the solvent, leaving oxide inside the gaps. A complete surface *etchback* also could be performed to reduce the overall thickness. However, no method has been as effective a planarization technique as CMP, chemical mechanical planarization (also referred to as *polish*). The following table and Figure 9.12 on page 210 describe and illustrate the process.

Description of Basic Steps in STI Oxide Polish and Nitride Strip

	Process Step	Description
1.	Trench Oxide Polish (CMP)	In the CMP area the wafer is turned upside down and held by vacuum to a wafer carrier. The wafer rotates in the same direction as a polishing pad attached to a rotating polishing table. A chemical slurry keeps the pad and wafer wet during the polishing process and provides reactants to help abrade the oxide surface. The nitride, being a harder material than the oxide, serves as the polish-stop material to prevent over-polishing the isolation structures. *Note:* The following statement applies to all polishing operations and will not be repeated again. The polishing process creates excessive particles and chemical contaminants that must be thoroughly removed before transferring the wafer to other areas in the wafer fab. Special scrub stations clean the top and bottom of the wafer and ensure it is rinsed and dried. Inspections for film thickness, particles, and defects are also conducted before wafer is transferred.
2.	Nitride Strip	The diffusion bay is traditionally equipped with hot phosphoric acid stations for stripping the nitride from the wafer. In the diffusion bay the wafer is stripped, cleaned, rinsed, and dried, then inspected to check the thickness of the isolation oxide.

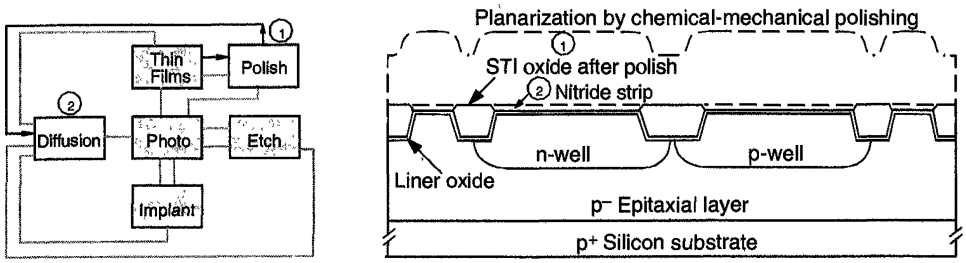


FIGURE 9.12 STI Formation

3. Poly Gate Structural Process

Forming the structure of the transistor gate is a critical step in the process because it includes thermally growing the thinnest *gate oxide* and patterning and etching the *polysilicon gate,* one of the smallest physical structures in the entire IC process (see Figure 9.13). The width of the polysilicon gate is often the most critical CD linewidth of the entire wafer.

Description of Basic Steps in Forming the Poly Gate Structure

	Process Step	Description
1.	Gate Oxide Growth	The wafer is cleaned to remove contamination and oxide. There is only a matter of a few hours before the wafer must be inserted in the oxidation furnace. Oxide will form as long as the silicon is exposed to oxygen in the atmosphere. The wafer is inserted in the oxidation furnace and a thin oxide of approximately 20 to 50 Å is grown.
2.	Polysilicon Deposition	The wafer is immediately loaded into an LPCVD furnace where silane (SiH_4) is introduced into the process chamber. The silane dissociates and polysilicon is deposited on the wafer. Approximately 5000 Å of polysilicon (also referred to as poly) is deposited. Some processes may call for doping the polysilicon immediately following the deposition step. This action can be done in the same process chamber or in a separate furnace.
3.	4th Mask, Poly Gate	In photolithography, deep UV lithography (assuming sub-0.25 μm technology) techniques are used to pattern the fine structures of the polysilicon gates. An *antireflective coating* (*ARC*) is commonly applied between the poly and the photoresist to reduce undesirable reflections. The resist linewidths for the gates are the narrowest structures on the IC; therefore, various quality measurements are required, such as CD and overlay registration (OL), and defect inspection (DI).
4.	Poly Gate Etch	By far one of the most critical etch steps in the IC process requires the use of the best *anisotropic plasma etchers* available in the fab. These single-direction etchers etch the polysilicon to provide the vertical profile as shown in Figure 9.13.

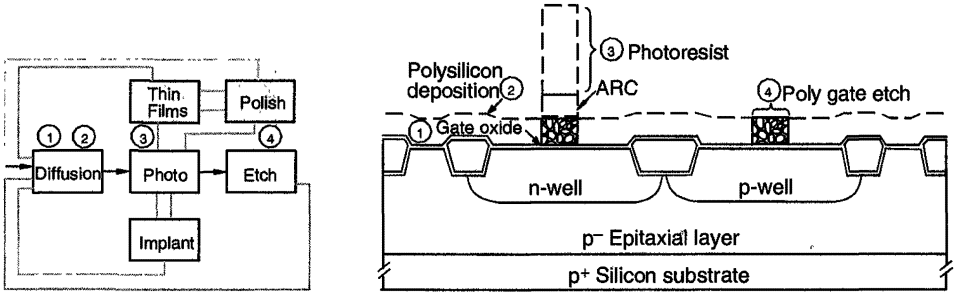


FIGURE 9.13 Poly Gate Structural Process

4. Lightly Doped Drain (LDD) Implants Process

As gate widths get smaller, the *channel length* below the gate structure (the silicon region between the source and drain) also decreases. This reduction increases the potential for charges to *punchthrough* the transistor source and drain and cause undesirable channel leakage current. There are techniques that are used to reduce the occurrence of this leakage in the channel.

The next series of ion implant steps begins to define the source and drain regions of transistors. Each pMOS and nMOS transistor will be implanted twice—once with a shallow implant called the *lightly doped drain (LDD) implant,* then followed by a medium or high dose *source/drain implant.* Lightly doped drain (LDD) implants (see Figure 9.14 below and Figure 9.15 on page 212) use the larger mass of arsenic and BF_2 dopant materials to create an *amorphous* (disorganized with no long-range order or crystal structure) upper layer of silicon. The combination of large mass and amorphous surface conditions helps maintain a shallow junction, which also helps reduce channel current leakage effects between the source and drain.

n^- LDD Implant ■ The steps for creating n^- lightly doped drain implants are described in the following process and illustrated in Figure 9.14.

Description of n- Lightly Doped Drain Implant

	Process Step	Description
1.	5th Mask, n^- LDD Implant	The purpose of this masking step is to pattern the wafer in a manner that opens windows in the photoresist where n-channel transistors can be implanted. All other areas are protected by the remaining photoresist.
2.	n^- LDD Implant (Low energy, shallowjunction depth)	Arsenic ions are selectively implanted through the windows in the patterned resist. The energy, dose and depth are significantly lower than in the previous n-well implant steps. Arsenic is preferred over phosphorus because its larger mass amorphizes the silicon surface to create a more uniform dopant depth during implant.

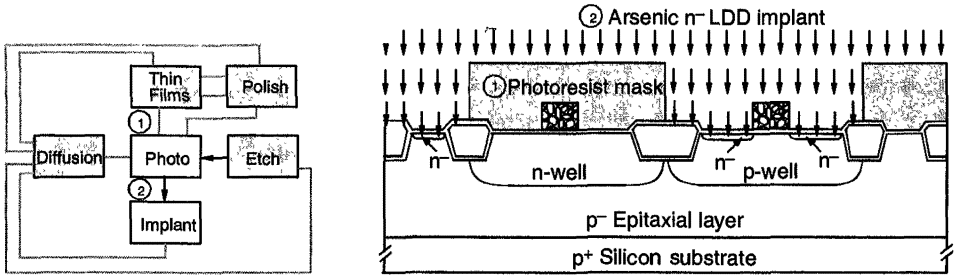


FIGURE 9.14 n^- LDD Implant

p^- LDD Implant ■ The steps for creating p^- lightly doped drain implants are described in the following process and illustrated in Figure 9.15.

Description of p- Lightly Doped Drain Implant

	Process Step	Description
1.	6th Mask, p^- LDD Implant	The purpose of this masking step is to pattern the wafer in a manner that opens windows in the photoresist where p-channel transistors can be implanted. All other areas are protected by the remaining photoresist.
2.	p^- LDD Implant (Low energy, shallow junction depth)	Boron difluoride, BF_2, is preferred over boron for this implant step. BF_2 is a heavier substance than boron. The heavier BF_2 helps amorphize the silicon surface.

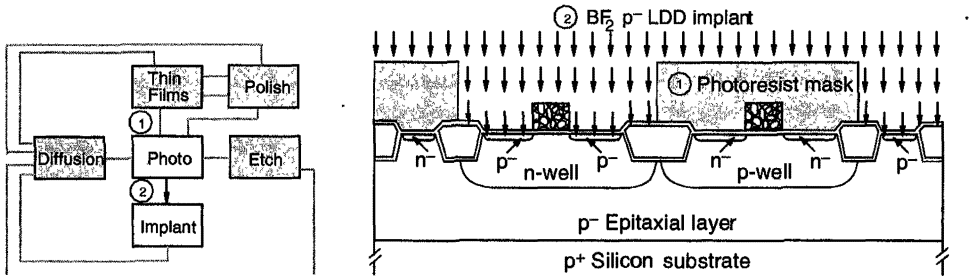


FIGURE 9.15 p^- LDD Implant

5. Sidewall Spacer Formation

Sidewall spacers will be used alongside the poly gates to prevent the higher source/drain (S/D) implant from penetrating too close to the channel where S/D punchthrough could occur. Sidewall spacer formation requires two major process steps (see Figure 9.16). First, an oxide layer is deposited across the surface of the wafer, then the oxide is etched back using a dry etch process. There is no requirement for a masking step because the anisotropic (single-direction) etch tool uses ions to sputter away most of the oxide. The etch process ends when the polysilicon is exposed. However, not all of the oxide is removed. Some oxide remains on the sidewalls of the polysilicon gates.

Description of Two Major Steps in Sidewall Spacer Formation

	Process Step	Description
1.	Spacer Oxide Deposition	This procedure is performed in the thin film bay. Approximately 1000 Å of oxide is deposited with a chemical vapor deposition (CVD) process. This layer will be used to form spacers on the sides of the polysilicon gates.
2.	Spacer Oxide Etchback	The dry plasma etcher removes most of the CVD oxide leaving behind the thicker oxide on the sidewalls of the polysilicon gates.

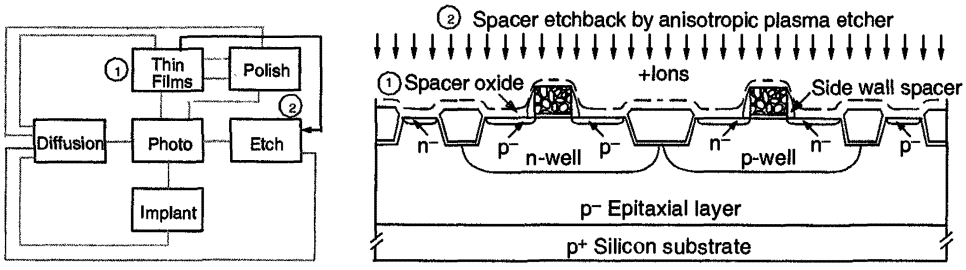


FIGURE 9.16 Sidewall Spacer Formation

6. Source/Drain (S/D) Implant Processes

To complete the retrograde implant technique, medium-dose implants are made to penetrate the silicon slightly beyond the LDD junction depth, but not as deep as the original twin-well implants (see Figure 9.17 and Figure 9.18). The spacer oxide from the previous step will protect the channel from the dopant atoms during the implant process.

n^+ S/D Implant ■ The steps for creating n^+ source/drain implants are described in the following process and illustrated in Figure 9.17.

Description of n+ Source/Drain Implant

	Process Step	Description
1.	7th Mask, n^+ S/D Implant	This masking step defines areas of the nMOS transistors that are to be implanted.
2.	n^+ S/D Implant (Medium energy)	This is a medium-energy implant step that penetrates the silicon deeper than the LDD junction depth. Spacer oxide prevents the arsenic dopant from encroaching into the narrow channel.

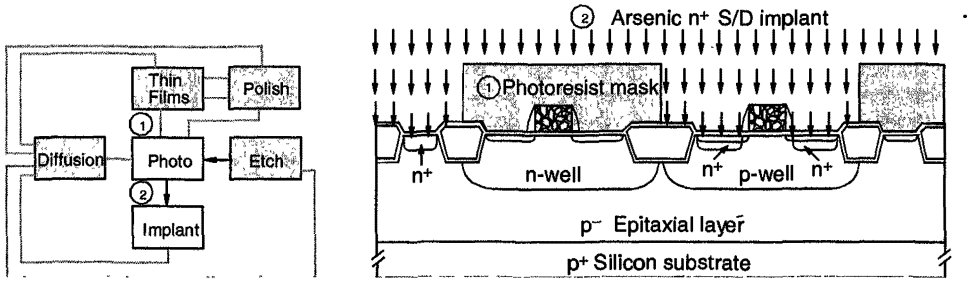


FIGURE 9.17 n^+ S/D Implant

p^+ S/D Implant ■ The steps for creating the p^+ source/drain regions are similar to n^+ S/D formation. The steps for p^+ S/D are described in the following process and illustrated in Figure 9.18.

Description of p+ Source/Drain Implant

	Process Step	Description
1.	8th Mask, p^+ S/D Implant	This masking step defines areas of the pMOS transistors that are to be implanted.
2.	p^+ S/D Implant (Medium energy)	This implant step penetrates the silicon slightly beyond the LDD junction depth. Spacer oxide prevents boron dopant from encroaching into the narrow channel.
3.	Anneal	The implanted wafer is annealed in a *rapid thermal process* (*RTP*) tool. RTP tools have the ability to quickly reach temperatures of ~1000 °C and maintain the setpoint for several seconds. This condition is important to prevent structures from spreading and to control the diffusion of dopants in the S/D regions.

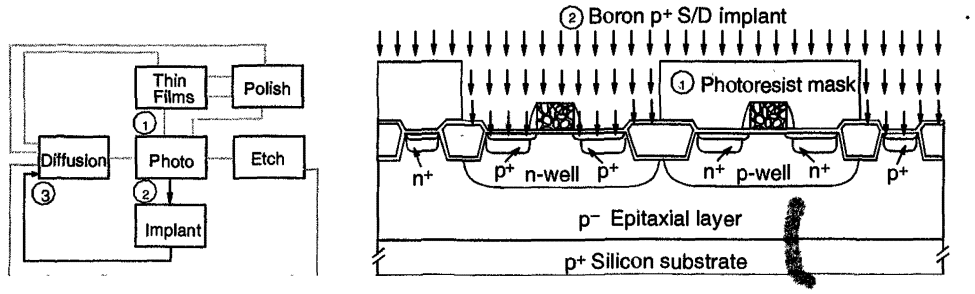


FIGURE 9.18 p^+ S/D Implant

7. Contact Formation

The purpose of the contact formation process steps is to form *metal contacts* on all active areas of silicon. These metal contacts serve to promote adhesion between the silicon and the metal conductor material that will be deposited later (see Figure 9.19). Titanium is a good choice for the metal contact material, although cobalt can also be used. Titanium has low resistivity characteristics and reacts very well with silicon. When raised to a temperature >700 °C, titanium bonds with silicon to form a *titanium silicide* ($TiSi_2$) compound, or *tisilicide*. Titanium and silicon dioxide do not react; therefore, no chemical bonding or physical connection is made between these two materials. Thus, titanium can be easily etched off the oxide without the need for a masking step. Tisilicide remains at all locations where active silicon exists (e.g. source, drain, and gate).

Description of Major Steps in Titanium Contact Formation

	Process Step	Description
1.	Titanium Deposition	The wafer is cleaned thoroughly to remove contaminants and oxides from the silicon. Titanium (Ti) is deposited on wafers using a sputtering process. *Sputtering* is a *physical vapor deposition (PVD)* process that is performed in a plasma process chamber where energetic argon ions bombard a metal target to release metal atoms, which deposit onto a wafer.
2.	Anneal	The wafer is inserted in the RTP tool. The high temperature triggers a chemical reaction between titanium and silicon that forms $TiSi_2$ (tisilicide).
3.	Titanium Etch	Chemicals etch away the unreacted titanium, leaving behind tisilicide over the active silicon areas.

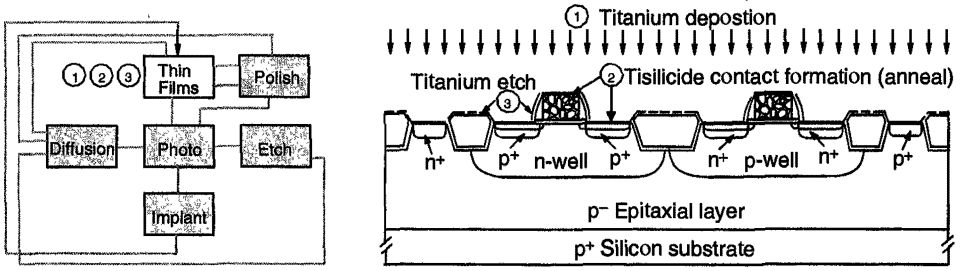


FIGURE 9.19 Contact Formation

8. Local Interconnect (LI) Process

The next step in the IC manufacturing process is to form metal connecting lines between transistors and other tisilicide contacts. The method used in this process flow is referred to as *local interconnect (LI)*. The steps leading up to the formation of LI are as complicated as the formation of the STI. The process begins with the deposition of dielectric films, followed by CMP, patterning, etch, and tungsten metal deposition and finished off with a metal polish (see Figure 9.20 and Figure 9.21). This process is referred to as *damascene*—a name adopted from a practice that began thousands of years ago by artisans in Damascus, Syria.[3] The result of these steps produces a top surface that resembles intricate inlaid jewelry or artwork. The graphic in Figure 9.22 illustrates how the metal lines are formed within the walls of the etched oxide.

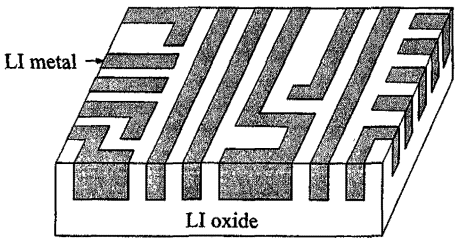


FIGURE 9.20 LI Oxide as a Dielectric for Inlaid LI Metal

Description of LI Oxide Dielectric Formation

	Process Step	Description
1.	Nitride (SiN_3) CVD	A barrier layer of silicon nitride is deposited using a CVD process. The nitride provides protection for the active regions from the dopants in the next film layer.
2.	Doped Oxide CVD	The LI dielectric component of the local interconnect structure is formed from CVD oxide (SiO_2). This oxide is lightly doped with phosphorus or boron. The addition of dopants to the oxide improves the dielectric qualities of the glass. An additional RTP step allows the glass to flow and smooth the surface.
3.	Oxide Polish (CMP)	CMP is used to planarize the LI oxide. The resulting thickness of the oxide after polish is ~8000 Å.
4.	9th Mask, LI Etch	The wafer is patterned in photolithography and then etched in the etch bay. Narrow trenches are created in the LI oxide that will serve as the forms for defining the paths for the LI metal.

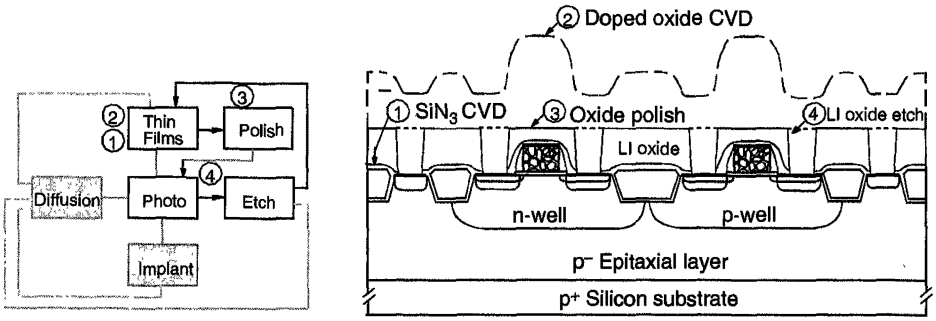


FIGURE 9.21 LI Oxide Dielectric Formation

Description of LI Metal Formation

	Process Step	Description
1.	Titanium Deposition (PVD process)	A thin barrier layer of titanium (Ti) lines the bottom and the inside of the LI walls. The Ti serves as a "double-adhesive tape" to hold the tungsten (W) to the SiO_2.
2.	Titanium Nitride Deposition	Titanium nitride (TiN) is deposited immediately over the Ti to serve as a diffusion barrier for the tungsten metal.
3.	Tungsten Deposition (CVD process polish)	Tungsten fills the LI trenches and coats the entire wafer. Tungsten is preferred over aluminum for LI metal due to its ability to fill holes without leaving voids in the formed metal plug. Another reason is for its good polishing characteristics.
4.	Tungsten polish	The tungsten is polished down to the upper surface of the LI oxide.

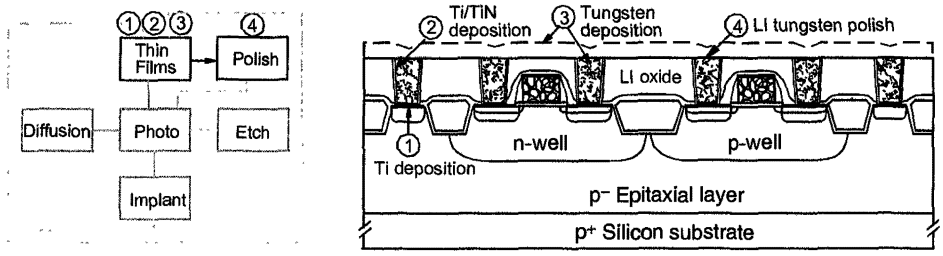


FIGURE 9.22 LI Metal Formation

9. Via-1 and Plug-1 Formation

The *interlayer dielectric* (*ILD*) serves as an insulator material between each metal layer or between the first metal layer and silicon. The ILD will have many small *vias,* which are small openings in the ILD that provide an electrical pathway from one metal layer to adjacent metal layers. Vias are filled with a conductive metal (usually tungsten, and referred to as a *tungsten plug*) and are placed at the appropriate locations to form the electrical circuit between metal layers (see Figure 9.23 and Figure 9.24). ILD-1 is the first of a series of interconnection process to be described.

Description of Major Steps in Via-1 Formation

	Process Step	Description
1.	ILD-1 Oxide Deposition (CVD)	Oxide is deposited across the surface of the wafer with a CVD tool in thin films. This layer (ILD-1) will serve as the dielectric material from which vias are to be formed.
2.	Oxide Polish	CMP is used to planarize the ILD-1 oxide. The resulting thickness of the oxide after polish is approximately 8000 Å. The wafer is scrubbed and cleaned to remove particles generated during the polish step.
3.	10th Mask, ILD-1 Etch	The wafer is patterned in photolithography and then etched in the etch bay. Small via holes, less than a quarter micron in diameter, are etched into the ILD-1 oxide. This step is critical in terms of CD, OL, and defects.

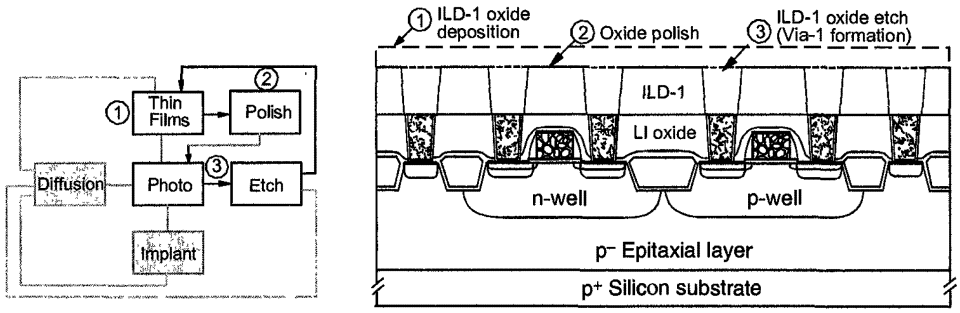


FIGURE 9.23 Via-1 Formation

Description of Major Steps in Plug-1 Formation

	Process Step	Description
1.	Titanium Barrier Metal Deposition (PVD)	A thin layer of titanium is deposited across the surface of the wafer with a PVD tool in thin films. Titanium lines the bottom and the walls of the via holes. The Ti serves as the glue that holds the tungsten plug inside the hole.
2.	Titanium Nitride Deposition (CVD)	TiN is thinly deposited on top of the Ti layer. The TiN serves as a diffusion barrier for the tungsten in the next deposition step.
3.	Tungsten Deposition (CVD)	Another CVD tool is used to deposit tungsten on the wafer. Tungsten fills the small openings to form the plug.
4.	Tungsten Polish	The tungsten-coated wafer is polished down to the upper surface of the ILD-1 oxide.

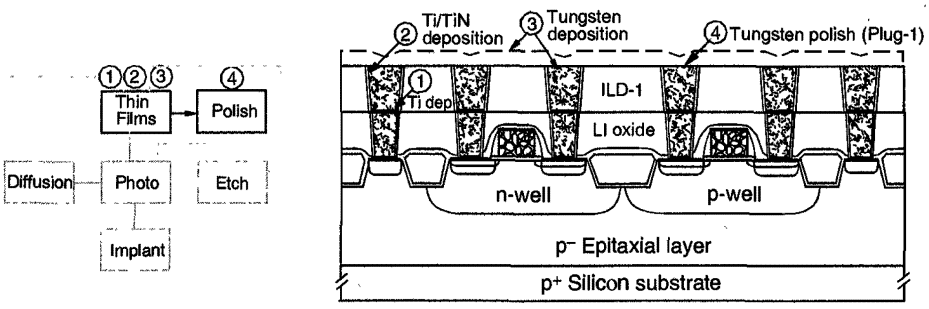
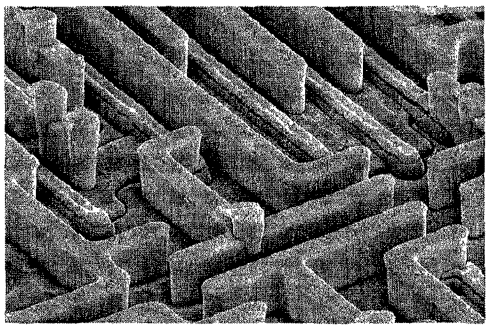


FIGURE 9.24 Plug-1 Formation



SEM Micrograph of Polysilicon, Tungsten LI and Tungsten Plugs (Polysilicon is represented by the thin lines between the taller LI lines). (Micrograph courtesy of Integrated Circuit Engineering)

10. Metal-1 Interconnect Formation

The next series of operations has to do with depositing a three-layer metal film, referred to as a *metal stack* or *sandwich,* on the wafer, followed by a masking and etching step (see Figure 9.25). The multilayer metal stack consists of different refractory metals—titanium, aluminum/ copper, and titanium nitride. When finished, the wafer will have the first of five metal stacks to be constructed on the device. The number of metal layers on a device varies depending on die complexity, with advanced die currently having around eight metal layers. Including LI metal, the device covered in this process flow has a total of six metal layers.

Description of Steps in Metal-1 Interconnect Formation

	Process Step	Description
1.	Titanium Barrier Metal Deposition (PVD)	As with other metal processes, titanium is the first metal to be deposited on the entire wafer. It provides a good bond between the tungsten plugs and the next metal, aluminum. It also bonds well to the interlayer dielectric material and improves the reliability of the metal stack.
2.	Aluminum-Copper Deposition (PVD)	The aluminum-copper (99% Al, 1% Cu) metal is sputtered onto the titanium-coated wafer by a PVD tool in thin films. The 1% Cu is added to the aluminum to improve the reliability of the aluminum.
3.	Titanium Nitride Deposition (PVD)	TiN is thinly deposited on top of the aluminum-copper layer to serve as an antireflective coating for the next photomasking step.
4.	11th Mask, Metal Etch	The wafer is patterned with photoresist, then the three-layer metal stack is etched using a plasma etcher.

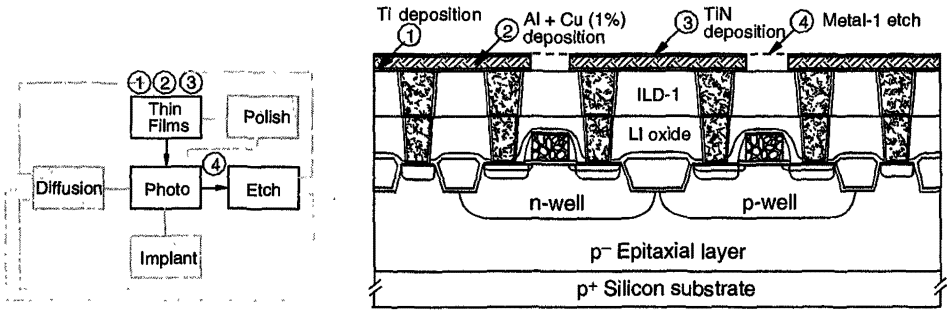


FIGURE 9.25 Metal-1 Interconnect Formation

11. Via-2 and Plug-1 Formation

The next four operations will result in the formation of the second interlayer dielectric (ILD-2) and the via openings (see Figure 9.26 and Figure 9.27). ILD-2 is similar to ILD-1 except for the very first step of filling small and large gaps that were etched into Metal-1. *Gap fill* is the filling of these gaps with a dielectric material that can be placed into the narrow spaces without creating voids or other defects that could affect electrical performance. There are two common methods for filling gaps—spin-on-glass (SOG) with etchback and *high-density plasma chemical vapor deposition* (HDPCVD). HDPCVD is the preferred method for sub-quarter micron processes. Once the gaps are filled, a plasma-enhanced CVD (PECVD) system finishes depositing the remainder of the ILD-2 oxide. Following the deposition of the ILD-2, the oxide is polished, patterned, and then etched to open the vias for the tungsten plugs to be formed.

Description of Major Steps in Via-2 Formation

	Process Step	Description
1.	ILD-2 Gap Fill	The latest gap-fill method for ULSI devices utilizes a high-density plasma process to alternately deposit and etch the interlayer oxide as its being deposited. The result is a dense oxide with few or no voids in the metal gaps.
2.	ILD-2 Oxide Deposition	A PECVD system is used to deposit the remainder of the ILD-2 oxide layer.
3.	ILD-2 Oxide Polish	The wafer is polished to planarize the surface prior to the next patterning step.
4.	12th Mask, ILD-2 Etch	The wafer is patterned with photoresist, then the ILD-2 oxide is etched using a plasma etcher.

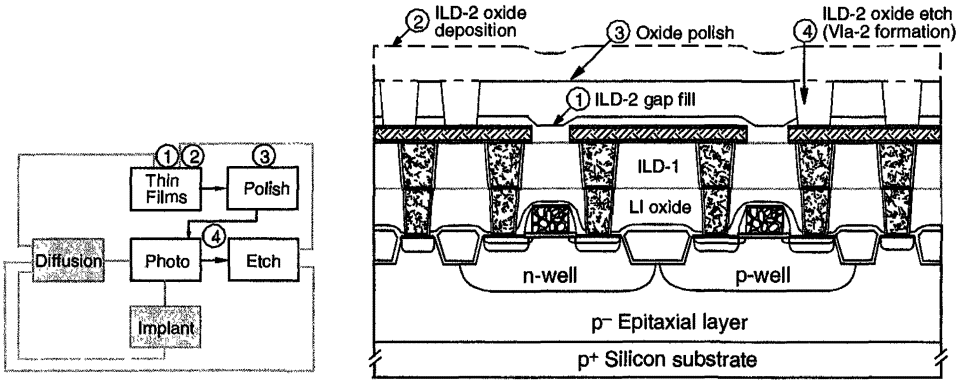


FIGURE 9.26 Via-2 Formation

Description of Major Steps in Plug-2 Formation

	Process Step	Description
1.	Titanium Barrier Metal Deposition (PVD)	As with other metal processes, titanium is the first metal to be deposited on the wafer. It provides a good bond between the tungsten plugs and the next metal, aluminum. It also bonds well to the interlayer dielectric material and improves the reliability of the metal stack.
2.	Titanium Nitride Deposition (CVD)	TiN is thinly deposited on top of the Ti layer. The TiN serves as a diffusion barrier for the tungsten in the deposition step.
3.	Tungsten Deposition (CVD)	Another CVD tool is used to deposit tungsten on the wafer. Tungsten fills the small via openings to form the plug.
4.	Tungsten Polish	The tungsten-coated wafer is polished down to the upper surface of the ILD-2 oxide, leaving the tungsten plugs in the vias.

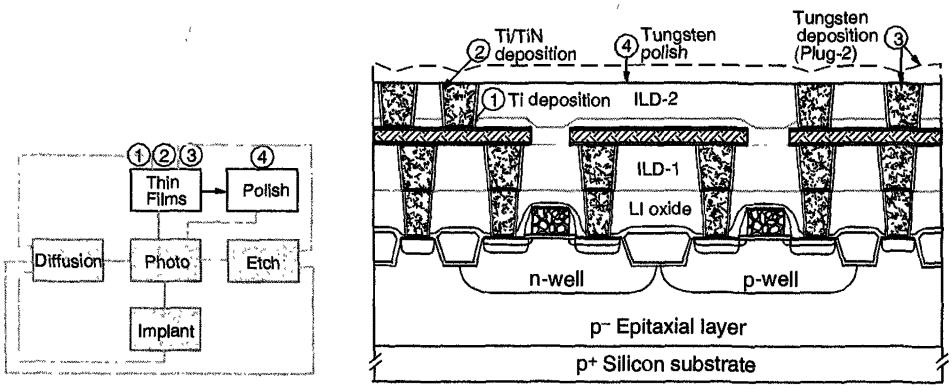
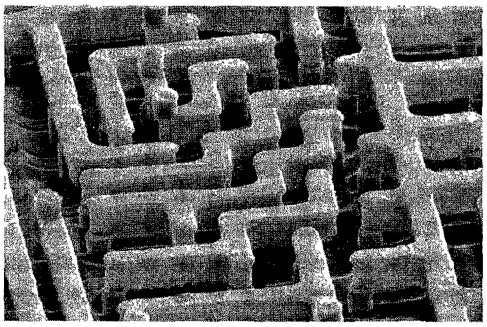


FIGURE 9.27 Plug-2 Formation



SEM Micrograph of First Metal Layer over First Set of Tungsten Vias (Note TiN Cap over Aluminum)
(Micrograph courtesy of Integrated Circuit Engineering)

12. Metal-2 Interconnect Formation

The next series of operations describes the process of forming the interconnects between layers (see Figure 9.28 on page 220). The process is repeated in the same manner for all remaining metal stacks.

Description of Major Steps in Metal-2 Interconnect Formation

Process Step		Description
1.	Metal-2 Deposition to Etch	The second metal stack is deposited exactly as the first metal stack. The stack is a composite of three layers—Ti, Al/Cu, and TiN. A plasma etcher is used to etch the Metal-2 lines through windows in the patterned resist.
2.	ILD-3 Gap Fill	Following the Metal-2 etch, HDPCVD is used to fill the metal gaps with dense oxide.
3.	ILD-3 Oxide Deposition to Polish	ILD-3 oxide is deposited using a PECVD method. This action is followed by polishing the oxide to attain surface planarity.
4.	Via-3 Etch, Ti/TiN Deposition, Tungsten Deposition, Polish	Via-3 openings are etched, then these are filled with barrier metal (Ti/TiN). Tungsten is deposited across the wafer surface. The tungsten is polished off the wafer until the ILD-3 oxide layer is reached. Tungsten plugs are left in the vias to provide the interconnect between Metal-2 and Metal-3.

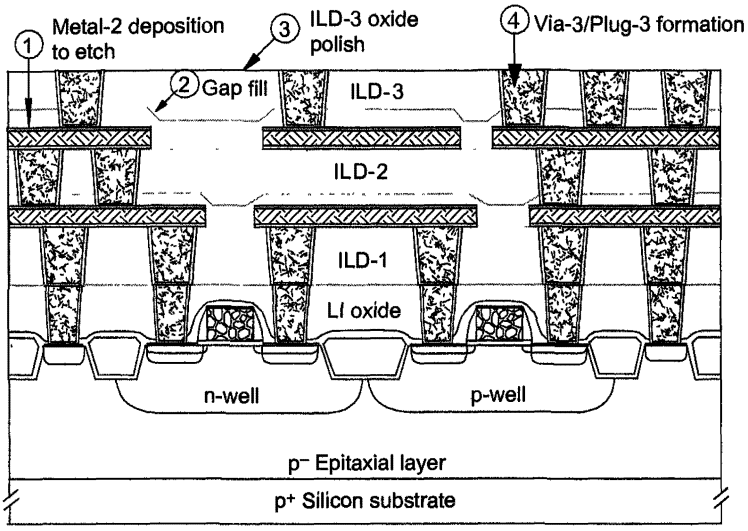


FIGURE 9.28 Metal-2 Interconnect Formation

13. Metal-3 to Pad Etch and Alloy

After repeating the layering process for layers 3 and 4 and at the completion of the Metal-4 etch, ILD-5 oxide is deposited using the thin film process (Figure 9.29). CMP is not required for this dielectric layer because the structures to be patterned are much larger than the quarter micron dimensions formed earlier in the process. The ILD-5 layer is etched to allow metal filling by the Metal-5 deposition. The Metal-5 layer is deposited thicker than previous metal stacks. This metal layer is etched as necessary to form bonding pads and to remove metal from areas where it is not needed.

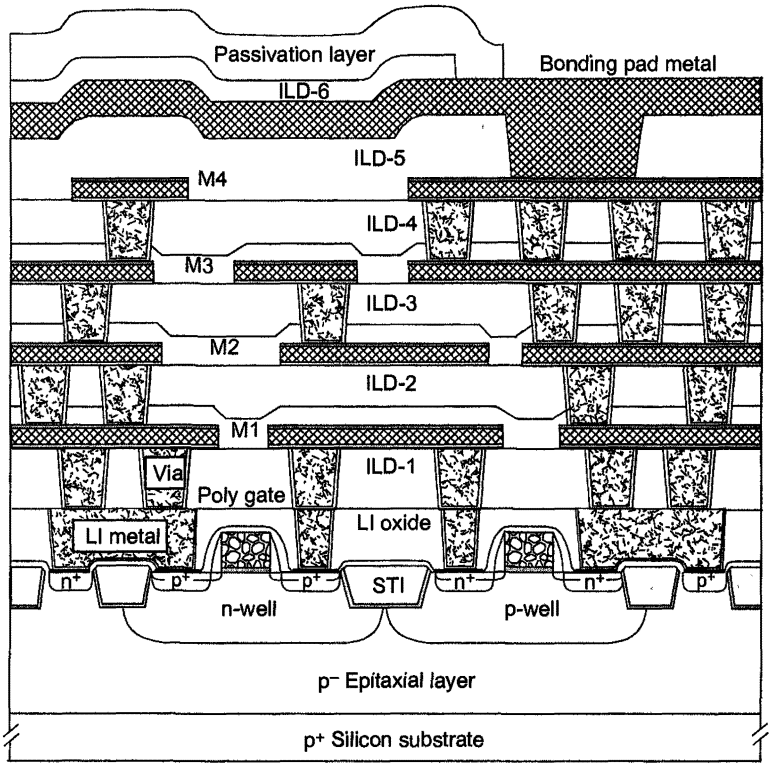
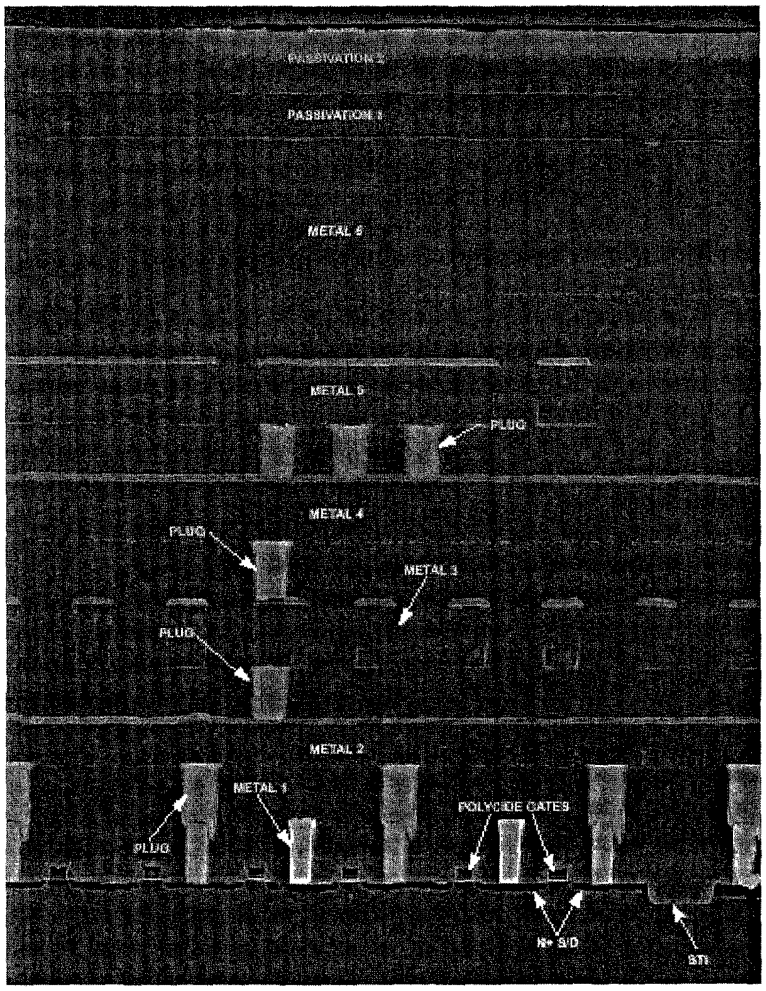


FIGURE 9.29 Full 0.18 μm CMOS Cross Section

This product incorporates 0.18 μm processing technology, including twin-tub technology, retrograde implants, shallow trench isolation, shallow S/D diffused regions, sidewall spacers, titanium silicide contacts, titanium barrier metal, tungsten local interconnect, tungsten plugs, three-layer metal stacks, HDP oxide gap fill, and PECVD oxide as the interlayer dielectric.

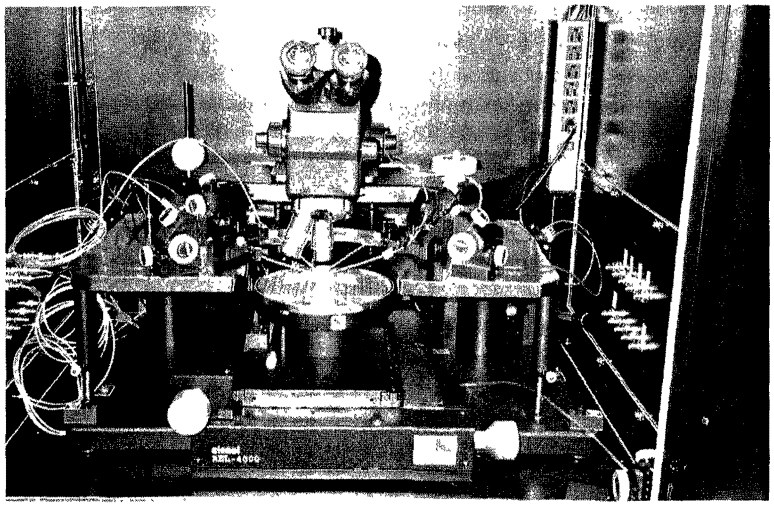
The final steps of the process include one more oxide layer (ILD-6) followed by a top layer of silicon nitride (~2000 Å). This layer is referred to as the *passivation layer*. Its purpose is to protect the product from moisture, scratches, and contamination. A final low-temperature alloy step in a diffusion furnace is performed. This heat treatment helps improve the metallurgical bonds between metal connections, thus improving electrical performance and reliability. Care must be taken with this alloy step to prevent overheating the product, which can cause permanent structural defects.



SEM Micrograph of Cross-section of Microprocessor (Micrograph courtesy of Integrated Circuit Engineering)

14. Parametric Testing

The wafer is tested twice to determine its product-worthiness—once when the wafer completes the first metal etch and again after the completion of the last wafer fab process step. Following metal etch, special microprobes connected to electronic test instruments are used to measure certain electrical parameters of specific device test structures on the wafer. This in-line parametric testing procedure is also referred to as *wafer electrical test,* or WET.



Wafer Electrical Test Using a Micromanipulator Prober (Photo courtesy of Advanced Micro Devices)

The last process step for the wafer is at wafer test/sort just outside the wafer fab. The wafer is probed and tested automatically in the electrical test and sort department (also referred to as *E-sort*). Each die on the wafer is tested for electrical functionality. Those die that fail are tracked through software using x-y position coordinates (previously marked with an ink spot) so that the locations of good and bad die are known for later processing. This data is used to determine die yield; that is, the percentage of good die versus bad die on a wafer. The wafer is now ready to be shipped to the assembly department. In some companies, the assembly plants are located in other parts of the world.

If the wafer passes the wafer sort, the wafer is sent to the backgrind department. Here the backgrinder grinds the backside of the wafer. This process makes the wafer thinner and easier to slice for eventual separation into individual die.

SUMMARY

An overview of the wafer fab process defines six major process areas: diffusion, photo, etch, ion implant, thin films, and polish, plus wafer test/sort at the end of the process. Diffusion is for high-temperature processing; photolithography patterns the wafer with photoresist; etch replicates the photoresist pattern in the wafer; ion implant dopes the wafer; thin films deposit dielectric and metal layers; and polish planarizes the top surface of the wafer.

A simplified CMOS process is defined with 14 manufacturing steps. (1) Twin-well implants create the n-well and p-well in the silicon. (2) Shallow-trench isolation isolates the silicon-active areas for sub-0.25 μm technology. (3) The gate structure is formed through gate oxide growth, polysilicon deposition, and patterning. (4) Lightly doped drain (LDD) implant forms a shallow implant in the source and drain. (5) Sidewall spacer formation protects the channel from the subsequent source/drain implant. (6) Source/drain implant is a medium-energy implant that forms a junction depth deeper than the LDD implant. (7) Contact formation forms a silicide contact for tungsten metal to alloy with the silicon. (8) Local interconnect (LI) forms the first metal connecting lines between transistors and contacts. (9) First interlayer dielectric (ILD) to Via-1 deposits the dielectric and creates vias between LI and first metal layer. (10) First metal layer to first metal etch deposits the metal sandwich and patterns the metal. (11) Second ILD to Via-2 deposits the second interlayer dielectric and creates the via openings. (12) Second metal layer to Via-3 deposits the second metal stack and deposits and etches ILD-3. (13) Metal-3 to pad etch and alloy repeats the layering process until the metal-5 bonding pad is deposited, followed by ILD-6 and a passivation layer. (14) The final process step is parametric testing to verify acceptability of each die on the wafer.

KEY TERMS

diffusion
high-temperature diffusion furace
wet cleaning station
photolithography
coater/developer track
stepper
etch
plasma etcher
plasma resist stripper
ion implanter
thin films
CMP (chemical mechanical planarization)
wafer scrubbers
twin-well (twin-tub)
retrograde implant
latchup
epitaxial silicon
native oxide
alignment-and-exposure system
linewidths
critical dimension (CD)
junction depth
ion implanter
anneal
shallow trench isolation (STI)
LPCVD (low-pressure chemical vapor deposition)
sidewall liner
overlay registration (OL)
defect inspection (DI)
visual inspection (VI)
radical chemicals
dry plasma etcher
isolation regions
spin-on-glass (SOG)
etchback
polish
gate oxide
polysilicon gate
anti-reflective coating (ARC)
silane
anisotropic plasma etcher
channel length
punchthrough
lightly doped drain (LDD) implant
source/drain implant
amorphous
sidewall spacers
rapid thermal process (RTP)
metal contacts
titanium silicide (tisilicide)
local interconnect (LI)
damascene
interlayer dielectric (ILD)
vias
tungsten plug
metal stack
sandwich
gap fill
high-density plasma chemical vapor deposition (HDPCVD)
plasma enhanced CVD (PECVD)
passivation layer
wafer electrical test (WET)

REVIEW QUESTIONS

1. List the six distinct production areas in a wafer fab and give a short description of each area.
2. What activity is performed in the diffusion area?
3. List five processes that are done in a high-temperature furnace.
4. What is the purpose of photolithography?
5. What is a coater/developer track used for? List five operations performed using this tool.
6. What is the major concern for particulate contamination in photolithography?
7. Identify the three production areas where photoresist-coated wafers can be found.
8. What is the purpose of the etch process? Name the most common tools used in this area?
9. What is the ion implanter used for?
10. What is the purpose of the thin films bay?
11. List four different tools used or processes done in the thin films area.
12. What does CMP stand for, and what is its purpose? What is another name for CMP?
13. List the 14 manufacturing steps outlined for a typical CMOS process.
14. Describe the epitaxial layer on the silicon wafer.
15. What is a twin-well?

16. Explain the retrograde implant technique. What problem does it help resolve?
17. What are the reasons for the thermal anneal process after ion implantation?
18. Why are the retrograde p-well implant energy levels lower than the n-well implant energy levels?
19. What is shallow trench isolation (STI)? What process did it replace?
20. What tool is used to etch silicon for STI, and why?
21. What tools are used to deposit the oxide in STI oxide fill?
22. Describe the most effective wafer planarization technique.
23. Why is the formation of the transistor gate structure a critical process step?
24. What problem can occur due to smaller gate widths?
25. How do lightly doped drain (LDD) implants reduce the channel-current leakage effects?
26. Explain the purpose of sidewall spacers.
27. What are medium-dose implants after the LDDs used for?
28. What is the purpose of the metal contact?
29. What is a local interconnect (LI)?
30. What is the purpose of the ILD?
31. What is a via? What is a tungsten plug?
32. Describe the materials used in a metal stack.
33. What is gap fill?
34. Why is CMP not required for the final ILD layer?
35. What is the purpose of the passivation layer?
36. Describe what happens in the wafer test/sort area.

REFERENCES

1. S. Ghandhi, Chap. 11 in *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., (New York: Wiley, 1994), pp. 704–800.
2. S. Wolf, *Process Integration,* vol. 2, *Silicon Processing for the VLSI Era,* (Sunset Beach: Lattice Press, 1990), pp. 17–19.
3. P. Singer, "Making the Move to Dual-Damascene Processing," *Semiconductor International* (January 1999): pp. 70–82.

CHAPTER 10

OXIDATION

A cornerstone of silicon IC wafer fabrication is the ability to thermally grow an oxide layer on the surface of a wafer. A major development from the 1950s was *oxide masking,* which patterned and etched openings in a thermally grown oxide layer to diffuse dopants into the silicon substrate. This development was a key factor in developing a process to fabricate large quantities of transistors.[1] In this way, oxidation played a major role in the development of silicon planar technology and remains viable today, which explains the widespread usage of silicon in wafer fabrication.

With the proper manufacturing control, oxide layers have high quality, stability, and desirable dielectric properties. Because of these qualities, oxidation is critical, especially for the thin oxide that is an essential part of the gate structure for MOS technology. Thermal oxide is used as a dielectric material, as well as for device isolation, oxide screens for implants, stress-relief oxides (pad oxides), and reoxidizing nitride and polysilicon surfaces for photoresist adhesion and stress reduction.

Oxides can be either deposited or grown. The focus in this chapter will be on thermally grown oxides. We will review the nature of oxide films and how they are grown, including detailed information on the growth mechanism and the high-temperature chamber where growth occurs.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe an oxide film for semiconductor manufacturing, including its atomic structure, its various uses, and its benefits.
2. State the chemical reaction for oxidation, and describe how oxide grows on silicon.
3. Explain selective oxidation and give two examples.
4. Identify the three types of thermal processing equipment, describe the five parts of a vertical furnace, and discuss the attributes of a fast ramp vertical furnace.
5. Explain what is a rapid thermal processor, its usage and design.
6. Describe the critical aspects of the oxidation process, including its quality measures and some common troubleshooting problems.

INTRODUCTION

Oxide on a silicon wafer is created by either the grown or deposited method. A *grown oxide layer* occurs on a wafer by providing externally supplied high-purity oxygen in an elevated-temperature environment to react with the silicon substrate. The high-temperature oxidation process occurs in the diffusion area of the wafer fab. This is the first process area when a silicon wafer enters the fabrication process (see Figure 10.1 on page 226). A *deposited oxide layer* is generated by using an external silicon source and O_2 and reacting these materials in a chamber to form a thin film on the wafer surface. This chapter will cover the basic technology of oxidation and explain the high-temperature process of thermal oxidation. Chapter 11 will cover the deposition of oxide and other materials.

Since a silicon wafer is a flat surface (planar), oxide is grown or deposited essentially as a layer. Because of this, it is referred to as a *thin film.* Once a layer is on the wafer surface, it is then modified during subsequent processing to attain the necessary three-dimensional shape used in the formation of circuit components, such as a trench capacitor or interconnect conductor line. This

modification produces three-dimensional shapes on the wafer surface referred to as *topography* or *surface topology*.

Oxide growth is a natural phenomenon that occurs by exposing the silicon wafer to oxygen at an elevated temperature. The ability to grow an oxide film on silicon is often listed as a significant reason for the use of silicon as the most common semiconductor substrate material (another major reason is silicon's relatively high melting temperature). We use the term "grow" to indicate that temperature is used to cause the oxide to grow out of the silicon semiconductor material, actually consuming silicon in the process.

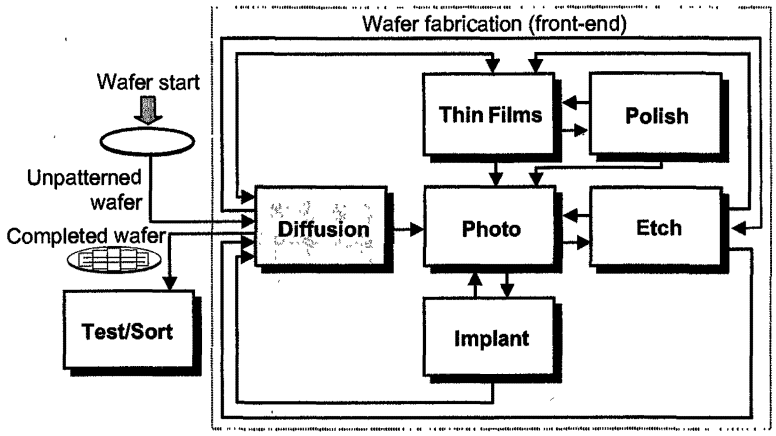


FIGURE 10.1 Diffusion Area of Wafer Fabrication
(Used with permission from Advanced Micro Devices)

The amount of thermal exposure (i.e., temperature multiplied by time) for a wafer during processing is referred to as the *thermal budget*. The thermal budget requirement for wafer fabrication is rapidly dropping.[2] A goal throughout semiconductor processing is to minimize the amount of thermal (heat) exposure to the wafer. As the critical dimension on device structures decreases to 0.18 μm and beyond, scaling requirements dictate shallower junction depths. To minimize the unacceptable diffusion of dopants out of a shallow junction region, the thermal budget must decrease accordingly. Another problem caused by excessive thermal budget is the increase in ohmic contact resistance of metal-layer interconnects (see Chapter 12), which increases the overall resistance of the conductive paths. A factor in determining the process conditions for many wafer fabrication process steps is the ability to minimize the thermal budget, either by reduced temperature or by minimizing the time at temperature.

OXIDE FILM

An oxide layer is grown on silicon typically at thermal oxidation temperatures between 750°C to 1100°C. An oxide layer grown on silicon is referred to as *thermal oxide* or *thermal silicon dioxide* (SiO_2). Since silicon dioxide is an oxide material, the two terms are used interchangeably. Another term for silicon dioxide is *glass*. Silicon dioxide is a dielectric material and will not conduct electricity.

Nature of Oxide Film

When a silicon surface is exposed to oxygen, an amorphous silicon dioxide film grows immediately. Although undoped silicon is a semiconductor material, SiO_2 is an insulator. The atomic structure of this SiO_2 film consists of a silicon atom surrounded by four oxygen atoms (see Figure 10.2). We can refer to this atomic structure as a silicon dioxide *tetrahedron cell*. Amorphous SiO_2 has no long-range periodic crystal order at the atomic level. Long-range order is absent because the tetrahedra are not arranged in a regular, three-dimensional array as in a crystal.[3]

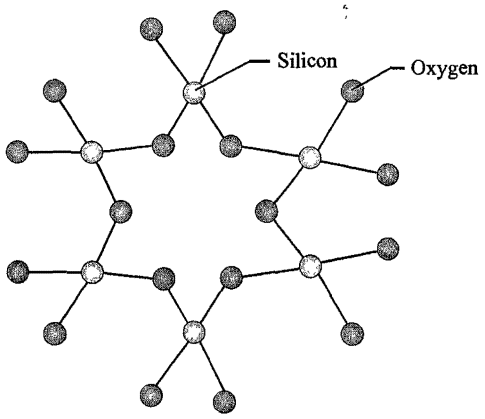


FIGURE 10.2 Atomic Structure of Silicon Dioxide
(Used with permission from International SEMATECH)

Silicon dioxide is a form of intrinsic (pure) glass with a melting temperature of 1732°C.[4] Thermally grown SiO_2 has strong adhesion to silicon and exhibits excellent dielectric properties. A layer of SiO_2 always covers the wafer surface because silicon oxidizes immediately upon exposure to air with a few monolayers of native oxide (a monolayer is one layer of atoms on a surface). Over time, the native oxide layer will thicken to an upper limit of about 40 Å, even at 25°C room temperature.[5] This type of oxide is nonuniform and usually considered a contaminant, with little use in semiconductor fabrication. There are some applications for use of native oxide in wafer fabrication, such as part of a dielectric film stack in memory cells. Native oxide grown from air is contaminated, but native oxide grown in chemical baths using high-purity chemicals is very clean.

Uses of Oxide Film

Oxide is important for silicon semiconductor fabrication because of its ease of formation and its excellent interface with the underlying silicon material. It is the most common layer used in semiconductor fabrication. Different ways in which an oxide layer is used to fabricate a microchip are:

- Device scratch protection and contaminant isolation
- Field isolation to confine charged carriers (termed surface passivation)
- Dielectric material in the gate oxide or memory cell structures
- Impurity-mask barrier during doping
- Dielectric layer between metal conductor layers

Device Protection and Isolation ■ Silicon dioxide grown on the surface of the wafer serves as an effective barrier to isolate and protect sensitive devices in the silicon. SiO_2 physically protects devices because it is a very hard and nonporous (dense) material that effectively insulates active devices in the silicon surface. The hard SiO_2 layer will protect the silicon from scratches and processing damage that might occur during fabrication. Transistors have traditionally been electrically isolated by thermally growing thick SiO_2 in the region between them using the LOCOS process (discussed later in this chapter). However, this process is unacceptable for sub-0.25 μm technology and has been replaced by shallow trench isolation (STI). The STI process uses deposited oxide as the main dielectric (see Chapters 9 and 11).

Surface Passivation ■ A major benefit from thermally grown SiO_2 is the reduction in the surface-state density of silicon that occurs by tying up dangling silicon bonds. This action is referred to as *surface passivation,* which prevents the deterioration of the electronic properties and reduces current-leakage paths of the semiconductor caused by moisture, ions or other external contaminants.[6] The hard SiO_2 layer will protect the silicon from scratches and processing damage that might occur during backend final fabrication. SiO_2 growth on the silicon surface can serve to tie up electrically active contaminants (mobile ionic contaminants) in the oxide layer away from the active surface of the silicon. Passivation is also important for controlling the leakage current of

junction devices and growing a stable gate oxide.[7] For the oxide layer to function as a good passivation layer, it has a quality requirement to be a uniform thickness and not have pinholes or voids in the material.

Another factor in using the oxide layer to passify the wafer surface is the oxide thickness. Adequate oxide thickness is necessary to prevent electrical charging in a metal layer caused by a charge buildup in the silicon wafer surface, much like charge storage and breakdown characteristics of ordinary capacitors. This charging would lead to electrical shorting and other undesirable electrical effects. A thick layer of oxide that inhibits charge buildup from metal layers is the *field oxide layer,* which typically is between 2,500 Å to 15,000 Å thick (see Figure 10.3).

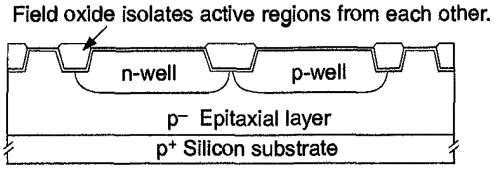


FIGURE 10.3 Field Oxide Layer

Silicon dioxide also has a coefficient of thermal expansion very similar to that of silicon. During high-temperature processing, the wafer will expand and then contract during cooling. The SiO_2 will expand and contract at close to the same rate as silicon, thus minimizing wafer warpage during the thermal excursions. This action also prevents film stress from separating the oxide film from the silicon substrate.

Gate Oxide Dielectric ■ An extremely thin layer of oxide is used as the dielectric material for the important *gate oxide structure* common in MOS technology (see Figure 10.4). The gate oxide is thermally grown because of its high quality and stability with the underlying semiconductor silicon. SiO_2 has a high dielectric strength of 10^7 volts/cm and high resistivity of approximately 10^{17} ohms-cm.[8]

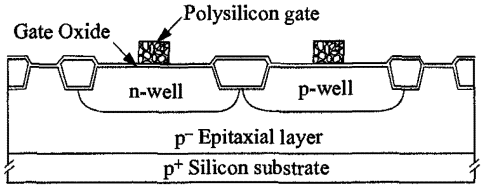


FIGURE 10.4 Gate Oxide Dielectric

The thickness of the gate oxide is chosen specifically for the scaling requirements of the device technology. For 0.18-μm generation technology, a typical gate oxide thickness is 20 ± 1.5 Å. The gate oxide has a specific thickness so that it scales properly with the entire gate structure to permit induction of a charge in the silicon wafer under the oxide. The widespread use of MOS technology in the ULSI era has made the formation of the gate oxide a primary concern in process development. A critical aspect of device reliability is *gate oxide integrity.* The gate structure in a MOS device is the part that controls the flow of electrical current through the device. Because this oxide is fundamental to the proper function of microchips based on field-effect technology, it is essential that it have high quality, with excellent film-thickness uniformity and no contaminants. Any contamination that degrades the proper functioning (integrity) of the gate oxide structure must be scrupulously controlled.

Dopant Barrier ■ Silicon dioxide can be used as an effective barrier to selectively introduce dopants into the silicon surface (see Figure 10.5). Once an oxide layer is grown on the silicon surface, then mask openings are etched into the SiO_2 to create windows where dopant materials can enter the wafer. The oxide protects the surface of silicon from impurity diffusion where there are

no openings and therefore allows selective impurity doping. Dopants have a slow rate of movement through SiO_2 when compared to silicon, requiring only a thin oxide layer to block dopants (note that this rate is temperature dependent).

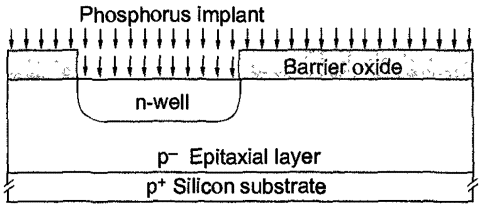


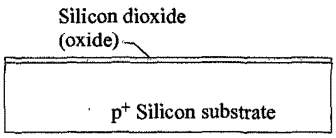
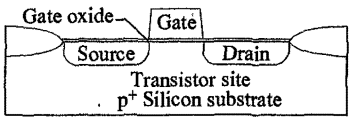
FIGURE 10.5 Oxide Layer Dopant Barrier

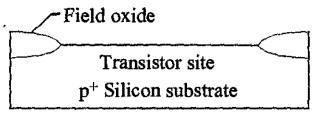
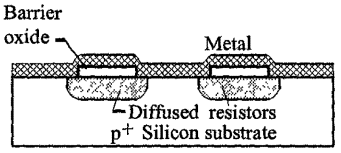
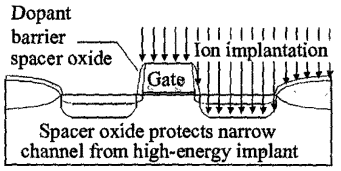
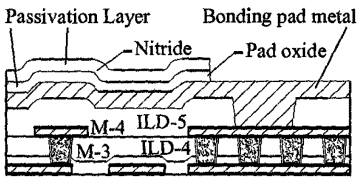
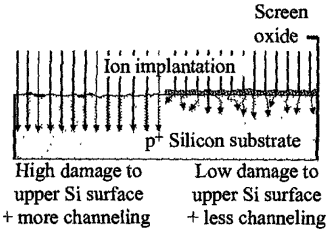
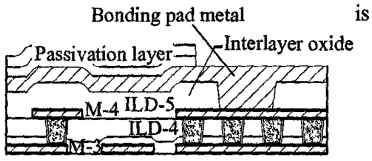
A thin oxide layer (e.g., 150 Å) is also grown in areas where ion implant occurs. This oxide screen serves to reduce damage to the silicon wafer and obtain better control over the depth that the dopant is implanted into the silicon by reducing channeling effects (see Chapter 17). Once implantation is done, HF (hydrofluoric acid) can be used to selectively remove the oxide, resulting in a planar silicon surface again.

Dielectric Between Metal Layers ■ Under normal conditions, silicon dioxide will not conduct electricity. For this reason, SiO_2 is an effective insulator between metal layers on the microchip. Silicon dioxide prevents shorting of an upper metal layer to the underlying metal layer, just as the insulator material on an electrical cord prevents an electrical short. The oxide quality is critical with no holes or voids allowed. This oxide is often doped to produce more effective flow properties and to better minimize the diffusion of contaminants (i.e., it acts as a getter). It is typically deposited using chemical vapor deposition (not thermally grown) and will be covered in more detail in Chapter 11.

A summary of different types of oxides and their uses in semiconductor fabrication is shown in Table 10.1.[9]

TABLE 10.1 Oxide Applications

Application	Purpose	Structure	Comments
Native Oxide	This oxide is a contaminant and is generally undesirable. Sometimes used in memory storage or film passivation.		Growth rate at room temperature is 15 Å per hour up to a maximum thickness of about 40 Å.
Gate Oxide	Serves as a dielectric between the gate and source-drain parts of MOS transistor.		Common gate oxide film thickness range from about 20 Å to several hundred Å. Dry thermal oxidation is the preferred growth method.

Application	Purpose	Structure	Comments
Field Oxide	Serves as an isolation barrier between individual transistors to isolate them from each other.		Common field oxide film thickness ranges from 2,500 Å to 15,000 Å. Wet oxidation is the preferred growth method.
Barrier Oxide	Protects active devices and silicon from follow-on processing.		Thermally grown to several hundred angstroms thickness.
Dopant Barrier	Masking material for depositing or implanting dopants into wafer.		Dopants diffuse into unmasked areas of silicon by selective diffusion.
Pad Oxide	Provides stress reduction for silicon nitride (Si_3N_4).		Thermally grown and very thin.
Implant Screen Oxide	Used to reduce implant channeling and damage.		Thermally grown.
Insulating Barrier Between Metal Layers	Serves as a protective layer between metal lines.		This oxide is not thermally grown, but is deposited.

THERMAL OXIDATION GROWTH

The various applications for thermally grown oxide have different thickness requirements. Silicon dioxide thickness ranges for different requirements are summarized in Table 10.2. The color chart in Appendix D shows silicon dioxide color as a function of the film thickness, which was used in the early days of wafer fabrication to estimate the oxide thickness. Other critical quality parameters for oxide layers in semiconductor fabrication are thickness, uniformity, pinholes and voids.

TABLE 10.2 Oxide Thickness Ranges for Various Requirements

Semiconductor Application	Typical Oxide Thickness, Å
Gate Oxide (0.18 μm generation)	20–60
Capacitor Dielectrics	5–100
Dopant Masking Oxide	400–1,200 (Varies depending on dopant, implant energy, time, and temperature)
STI Barrier Oxide	150
LOCOS Pad Oxide	200–500
Field Oxide	2,500–15,000

Chemical Reaction for Oxidation

Thermal oxide is grown by a chemical reaction between silicon and oxygen. Uniform oxide growth is achieved by exposing silicon to elevated temperature in the presence of high-purity oxygen. If this growth occurs with *dry oxygen,* meaning no moisture, then the following chemical equation describes the reaction:

$$Si\ (solid) + O_2\ (gas) \rightarrow SiO_2\ (solid)$$

The time and quality of this reaction varies and is affected by the purity of oxygen gas supplied to the silicon wafer surface and the reaction temperature. It occurs naturally when silicon is exposed to air at room temperature. The reaction rate is increased with an increase in temperature. The typical temperature for the oxidation of silicon during wafer fabrication is normally between 750°C to 1100°C and can vary for different oxidation process steps. The furnace temperature at any one operation is precisely controlled. The rate of oxide thickness for dry oxidation versus temperature and time is shown in Figure 10.6 on page 232.[10]

Wet Oxidation ■ When water vapor is introduced into the reaction, known as *wet oxidation*, the rate of theoxidation reaction is increased further. The chemical reaction for wet oxidation is:

$$Si\ (solid) + 2H_2O\ (vapor) \rightarrow SiO_2\ (solid) + 2H_2\ (gas)$$

For wet oxidation, oxygen saturated with water vapor is used instead of dry oxygen as the oxidizing gas. The water vapor could also be supplied as steam, referred to as*pyrogenic steam* (see Figure 10.7 on page 232). A wet oxidation reaction produces a silicon dioxide film and hydrogen gas during oxidation growth. The faster growth rate in a wet atmosphere is due to the faster diffusion and higher solubility of water vapor than oxygen in silicon dioxide.[11] However, the hydrogen molecules produced in the reaction are trapped in the solid silicon dioxide layer, making the layer less dense than oxide grown in dry oxygen. This condition can be improved to achieve a similar oxide in structure and properties such as that produced in dry oxidation by heating the oxide in an inert atmosphere.

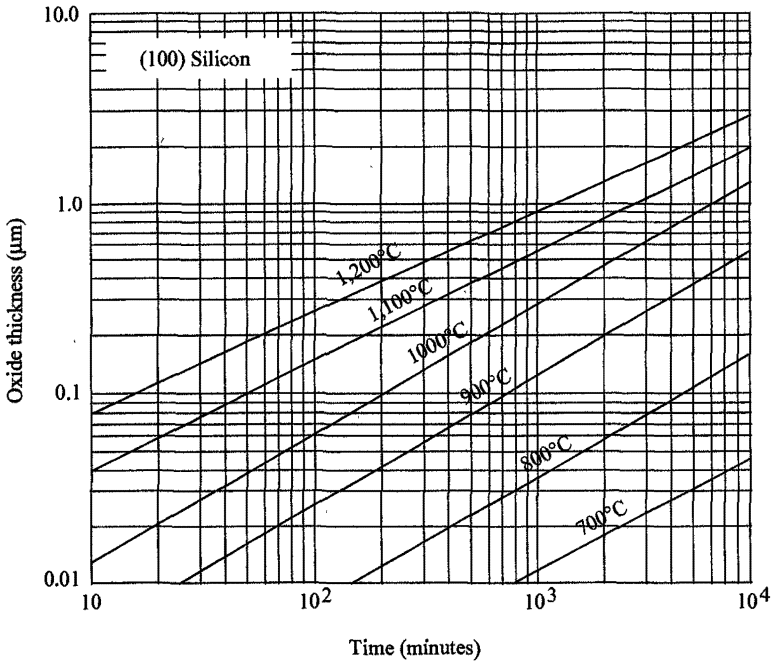


FIGURE 10.6 Dry Oxidation Time (Minutes)

Oxidation Growth Model

Silicon dioxide grows by consuming silicon, which is the case for either the dry or wet oxidation process. This is shown in Figure 10.8. The thickness of the silicon consumed is 0.46 of the total oxide thickness.[12] This means that for every 1,000 Å of oxide thickness, 460 Å of Si is consumed.

The growth of the oxidation layer is controlled and limited by the movement of oxygen through the oxide at the oxide-silicon interface. For the oxide layer to continue growing, the oxygen must come in contact with the silicon. However, the SiO_2 separates the oxygen gas from the silicon wafer. Oxide growth occurs when the oxygen gas molecules move through the existing silicon dioxide layer to the silicon wafer. This movement is referred to as diffusion (more precisely, gas diffusivity through a solid barrier). *Diffusion* is the movement of one material through another.

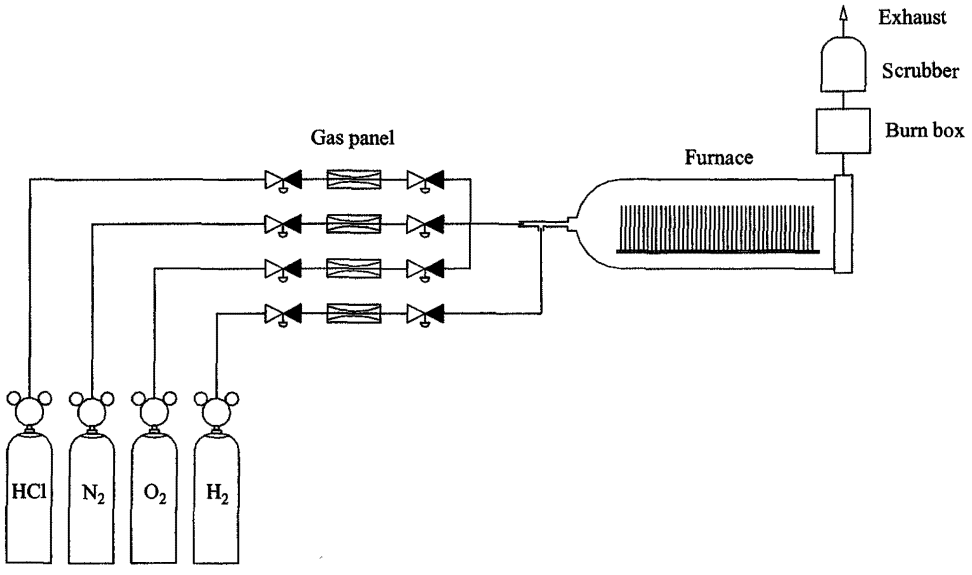


FIGURE 10.7 Wet Oxidation

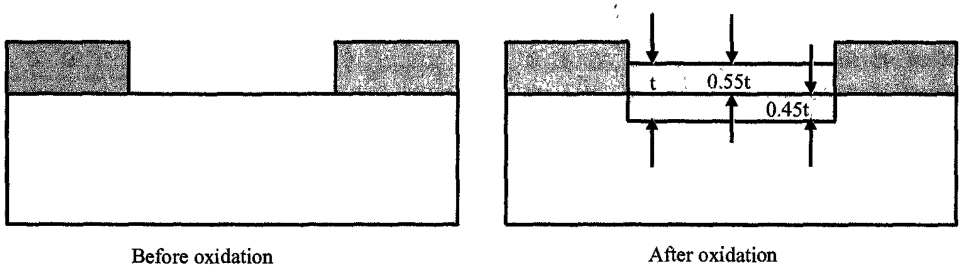


FIGURE 10.8 Consumption of Silicon During Oxidation

Atoms diffuse from regions of high concentration to regions of low concentration for solid, liquid, or gas states, and this diffusion is accelerated by thermal energy. It is a natural physical process that occurs regularly in everyday life. An example of liquid-state diffusion is putting a drop of red food colorant into a glass of water (see Figure 10.9). The colorant is initially concentrated when it enters the water, and then it will diffuse into the water until it is uniformly red. If the water is heated, then the diffusion will spread throughout the water more quickly.

The workbay in the wafer fab where oxidation occurs is still referred to as diffusion or the diffusion bay. In the early days of wafer fabrication, diffusion was an important process for creating pn junctions. Dopants were supplied to the silicon by chemical sources and then diffused to the desired junction depth by subjecting the wafers to elevated temperatures. Diffusion of dopants to create pn junctions is no longer done in wafer fabrication, although the use of the term diffusion bay is still common today. It has been replaced by ion implantation (see Chapter 17). Nevertheless, diffusion of materials (in this case oxygen) occurs in processes such as oxidation. Oxidation is well defined and governed by certain laws of diffusion that are based on a set of mathematical relationships referred to as Fick's laws. *Fick's laws* describe the rate of movement of diffusing materials based on temperature, concentration, and the energy necessary to drive diffusion. The detailed study of Fick's laws is not covered in this text.

Oxide-Silicon Interface ■ There is an abrupt transition at the oxide-silicon (Si/SiO_2) interface between single crystal silicon to amorphous SiO_2. Recall that for the SiO_2 molecule, each silicon is bonded to four oxygen atoms and each oxygen atom is bonded to two silicon atoms. At the Si/SiO_2

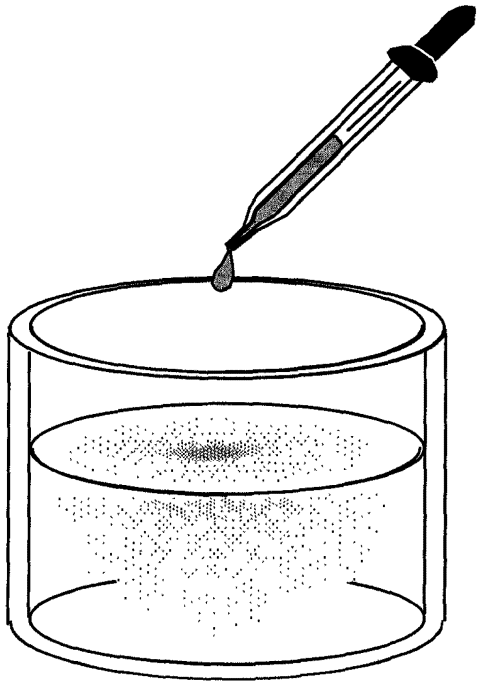


FIGURE 10.9 Liquid-State Diffusion

interface, some silicon atoms in the structure remain unbonded (see Figure 10.10). This incomplete oxidation of silicon less than 2 nm away from the Si/SiO_2 interface is the origin of a positive *fixed oxide charge.* Other charges built up at the interface include an *interface-trapped charge* consisting of positive or negative charges that result from structural defects, oxidation-induced defects, or metal impurities, and a *mobile oxide charge* due to mobile ionic contaminants (MICs). There is also the possibility of an *oxide-trapped charge* that is positive or negative and trapped in the bulk of the oxide away from the interface.[13] This accumulation of charge at the Si/SiO_2 interface is undesirable for normal device operation and can cause the threshold voltage of a MOS device to shift to unacceptable values.[14] Some of these undesirable charges can be minimized through a low-temperature (~ 450°C) anneal step in hydrogen or forming gas (hydrogen-nitrogen mixture).[15]

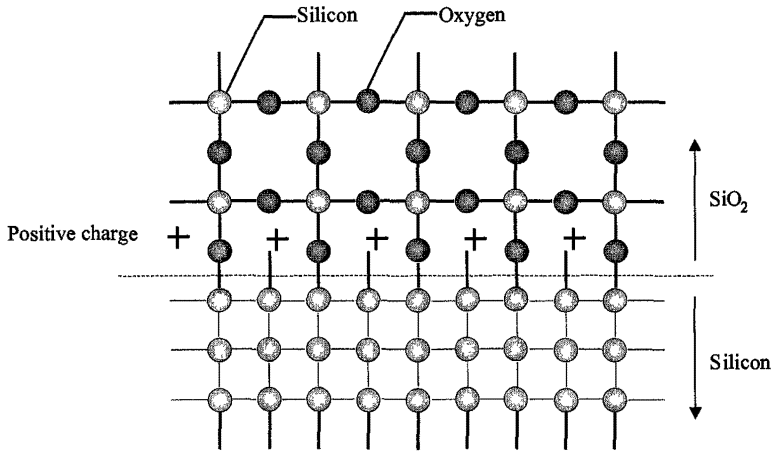


FIGURE 10.10 Charge Buildup at Si/SiO_2 Interface
(Used with permission from International SEMATECH)

Use of Chlorinated Agents in Oxidation. Using a chlorine-containing gas during the oxidation process can serve to neutralize the charge accumulation at the interface. The chlorine ions will diffuse to the positively charged layer and form a neutral layer. The chloride concentration is kept below 3% because excessive chloride ions will cause device instability. Another important advantage of chloride ions added to the thermal oxidation process is that they tend to increase the oxidation growth rate by about 10–15%. Furthermore, the presence of chlorine can actually immobilize (referred to as gettering) mobile ionic contaminants (MICs) that come from the furnace equipment, processing materials, and handling.

Early work used chlorine or vapor-phase hydrogen chloride (HCl). Chlorine is extremely toxic and HCl vapor in the presence of water vapor is corrosive. A common chlorine compound often used to transport the chlorine into the furnace is dichloroethylene (DCE) or variants of this compound.[16] Other less-corrosive sources of chlorine are trichloroethylene (TCE) and 1,1,1-trichloroethane (TCA).[17] TCE is carcinogenic and not used anymore, while TCA has fallen out of favor because it is an ozone-depleting chemical. Also, some semiconductor firms have switched back to HCl because of significant improvements in high-purity piping and fittings used to transport this chemical.

Rate of Oxide Growth ■ The *rate of oxide growth* describes how fast the oxide grows on the wafer. It depends on parameters such as temperature, pressure, oxidizing condition (dry or wet), silicon crystal orientation, and doping levels. The growth rate is of interest because the wafer processing time can be reduced if diffusion occurs quickly, which serves to reduce the thermal budget. The model for oxide growth on silicon is referred to as a linear-parabolic model developed by Deal and Grove and accurately represents oxide growth over a wide range of thicknesses (most optimally from 300 to 20,000 Å).[18] Oxide growth is described by two growth stages: the linear stage and the parabolic stage.

The initial growth of silicon dioxide is referred to as the *linear stage* and consumes silicon on the wafer surface as a linear function of time. This means the oxide layer is growing into the wafer

at a linear rate over time. The linear stage of oxide growth is valid up to about 150 Å of oxide thickness.[19] It is described by the linear equation:[20]

$$X = \left(\frac{B}{A\,hr}\right)t$$

Where, X = the thickness of the growing oxide
(B/A) = the linear rate constant
t = the time it takes to grow the oxide

In the linear stage, oxidation varies linearly with time. Oxidation is *reaction-rate controlled* in the linear region because the limiting factor for oxidation growth is the reaction occurring at the Si/SiO_2 interface. You will note that the linear rate constant, B/A, is the slope of this linear relationship, and therefore it controls the rate of the reaction. As temperature increases, the values for B/A increase, which means that the rate of oxidation also increases.

The *parabolic stage* of oxidation growth is the second phase of oxidation growth and starts after about 150 Å of oxide thickness. The equation that describes the parabolic stage is:[21]

$$X = (Bt)^{\frac{1}{2}}$$

Where, X = the thickness of the growing oxide
B = parabolic rate constant
t = the time it takes to grow the oxide

Note that this equation generates the shape of a parabola. Oxide growth in the parabolic stage is much slower than the linear stage. This is because as the oxide layer becomes thicker, the oxygen must diffuse through a larger distance to arrive at the Si/SiO_2 interface (see Figure 10.11). The reaction is thus limited by the rate at which the oxygen diffuses through the oxide. For this reason, the parabolic stage of oxide growth is said to be *diffusion controlled.* As the parabolic rate constant increases, the rate of oxide growth will increase. For instance, the parabolic rate constant, B, for wet oxidation is found to be much larger than that for dry oxidation.[22] Thus the rate of oxidation increases with wet oxidation.

A general curve shows both the linear and parabolic stage, as illustrated in Figure 10.12 on page 236.[23] This is a simplified curve from the original curve published by Deal and Grove in 1965 used to describe the thermal oxidation of silicon.[24]

Factors Affecting Oxide Growth ■ There are other factors that can affect the rate of oxide growth besides temperature and the presence of H_2O. We will review some of these other factors.

Dopant Effects. Heavily doped silicon oxidizes at a faster rate than lightly doped material. In the parabolic stage, boron doping will oxidize faster than phosphorus. Boron tends to become incorporated in the oxide film, which weakens its bond structure and leads to a subsequent increase in the diffusivity of the oxygen through it.[25] There is little difference in the linear rate constant between boron and phosphorus doping.

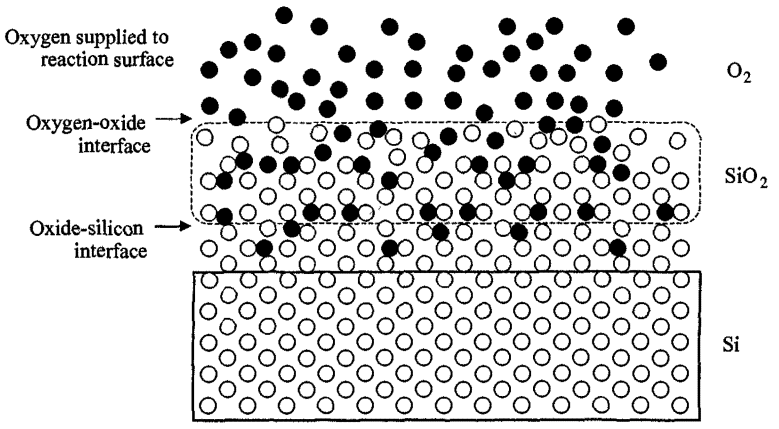


FIGURE 10.11 Diffusion of Oxygen Through Oxide Layer
(Used with permission from International SEMATECH)

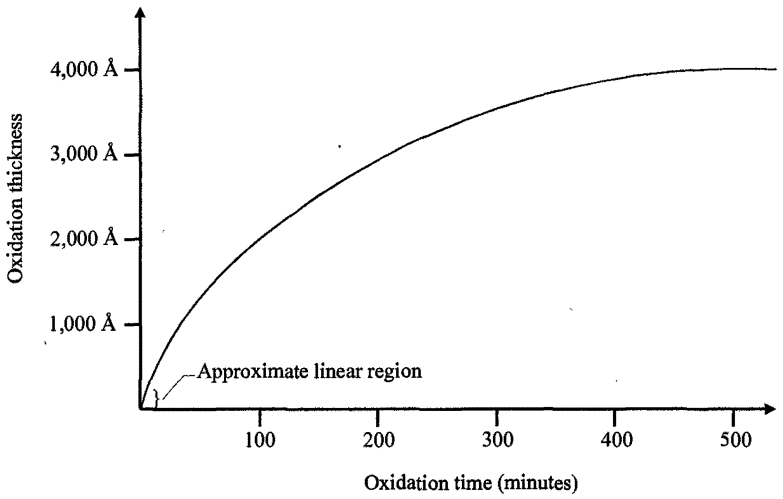


FIGURE 10.12 Linear and Parabolic Stages for Dry Oxidation Growth at 1100ºC (Used with permission from International SEMATECH)

Crystal Orientation. The linear oxidation rate is dependent on the orientation of the crystal because the density of silicon atoms on the (111) plane is greater than on the (100) plane. Therefore, (111) silicon crystal will oxidize in the linear stage at a faster rate than (100) silicon crystal, but (111) also has a greater charge buildup.

During the parabolic stage, the parabolic rate constant, B, is independent of the crystal orientation of the silicon substrate. There is no difference in oxidation growth rate for (111) vs. (100) in the parabolic stage. This is logical since the parabolic rate of oxidation growth is based on the diffusivity of the oxygen through the existing SiO_2 and is not affected by the reaction at the Si/SiO_2 interface.[26] However, charge buildup remains high because this factor is a function of atom density at the surface.

Pressure Effect. Since the growth rate of an oxide layer depends on the movement of the oxidizer from the gas phase to the silicon interface, the growth rate will increase with pressure. Higher pressure significantly increases the linear and parabolic rate constants, forcing the oxygen atoms to penetrate through the growing oxide more rapidly. This condition allows the temperature to be reduced to achieve equivalent growth rates or a more rapid oxide growth at the same temperature. A rule of thumb for the increased oxide growth versus temperature reduction is that for each 1-atmosphere increase in pressure, the furnace temperature can be decreased 30°C.[27] This ratio can be used to reduce the thermal budget. A high-pressure oxidation process could be used, for example, to grow a thick field oxide layer.

Plasma Enhancement. Plasma-enhanced oxidation is another method that increases the oxide growth rate at low temperature, thus reducing the thermal budget.[28] The technique is usually carried out in an oxygen-plasma discharge generated by an RF source. The silicon is biased to be at a potential less than the plasma potential, which collects the charged oxygen in the plasma on the wafer. This activity results in rapid oxidation of the silicon and allows oxides to be grown at temperatures less than 600°C. The drawbacks to this technique are problems associated with particle generation, higher film stress that is different from thermally grown oxides, and inferior film quality compared to thermally grown oxides.[29] For these reasons, this method is not widely used in wafer fabrication.

Initial Growth Phase ■ The Deal and Grove linear-parabolic model accurately predicts oxidation growth for thickness above 300 Å. However, below this thickness it is found that dry oxidation is faster than predicted. This is an important oxidation growth range because the gate oxide thickness for subquarter micron MOS technology is currently around 20 to 60 Å. The manufacturing process must be capable of producing these oxide layers with high yield and long-term reliability.

Because the gate oxide is becoming so thin, there is not one model that accurately predicts the rate of oxidation. One particular model has detected the presence of pores, about 10 Å in diameter, in very thin oxides grown in dry oxygen. These pores allow the oxidant to remain in direct

contact with the silicon in the early phases of growth, thus causing the rapid initial growth of oxide.[30] The investigation into thin gate oxides is an area of ongoing research.

Selective Oxidation ■ The oxidation of selective areas on a wafer uses SiO_2 to electrically isolate adjacent devices on the silicon surface. The traditional method for isolating devices with greater than 0.25 μm feature sizes has been *local oxidation of silicon* (*LOCOS*). A deposited layer of silicon nitride (Si_3N_4) serves as an oxidation barrier. It is etched to allow selective thermal oxide growth, since oxide will not grow where the nitride covers the silicon (see Figure 10.13). After thermal oxidation, the nitride and any underlying barrier oxide are removed to expose bare silicon surface regions ready for device formation.

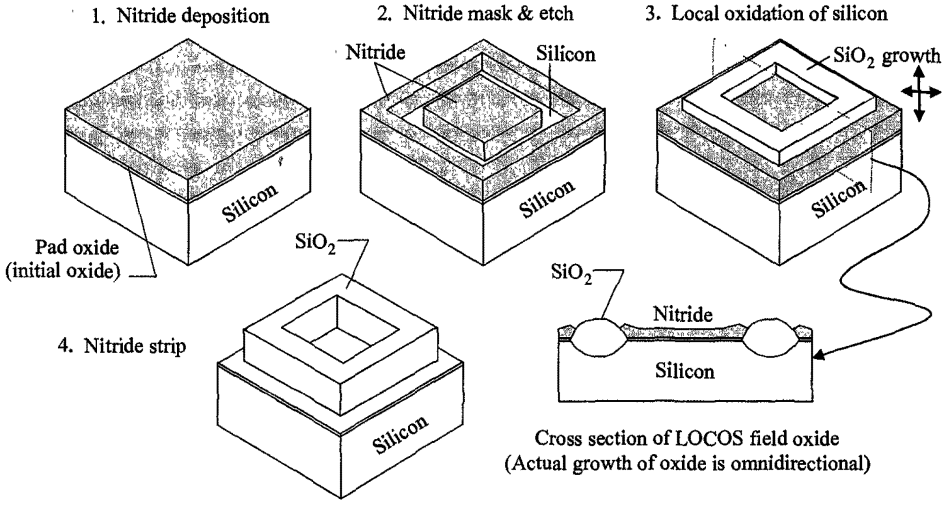


FIGURE 10.13 LOCOS Process

As the oxygen diffuses through the grown oxide, it moves in all directions. Some of the oxygen moves down into the silicon and other oxygen atoms move sideways. This means there is a slight lateral growth of the oxide under the nitride mask. Since the oxide is thicker than any silicon consumed, growth under the nitride mask serves to push up the nitride edges. This action is referred to as the *bird's beak effect.* This phenomenon is an undesirable by-product of LOCOS type oxidation processes (see Figure 10.14). The bird's beak effect is most pronounced when the oxide is relatively thick. To reduce stress between the nitride mask and silicon, a thin barrier layer of thermal oxide is used between them, which is termed a *pad oxide.*

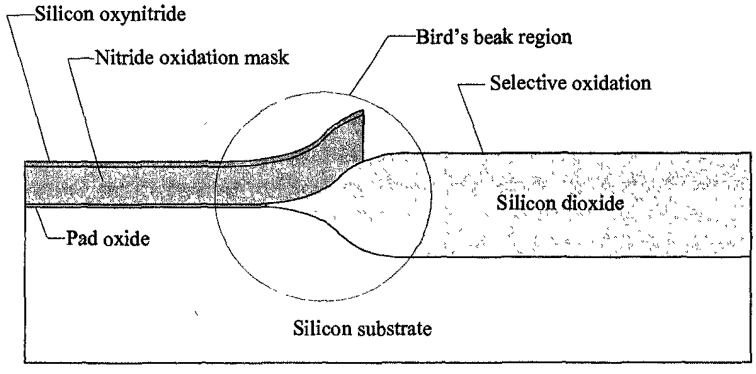


FIGURE 10.14 Selective Oxidation and Bird's Beak Effect
(Used with permission from International SEMATECH)

Shallow Trench Isolation (STI). The primary form of selective oxidation used for sub-0.25-μm technology is *shallow trench isolation* (*STI*). The main dielectric material in STI is a

deposited oxide (see Chapter 11). Selective oxidation is accomplished by creating a mask, usually of silicon nitride (Si_3N_4). The mask is deposited and patterned, followed by etching of the silicon to form a trench. The etching of silicon trenches is followed by thermal oxidation of a 150–200 Å thick oxide layer in areas exposed through the mask (see Figure 10.15). This thermally grown oxide passivates the silicon surface and serves as a barrier between the silicon and deposited trench-fill oxide. It is also an effective barrier that prevents sidewall current leakage in finished devices.

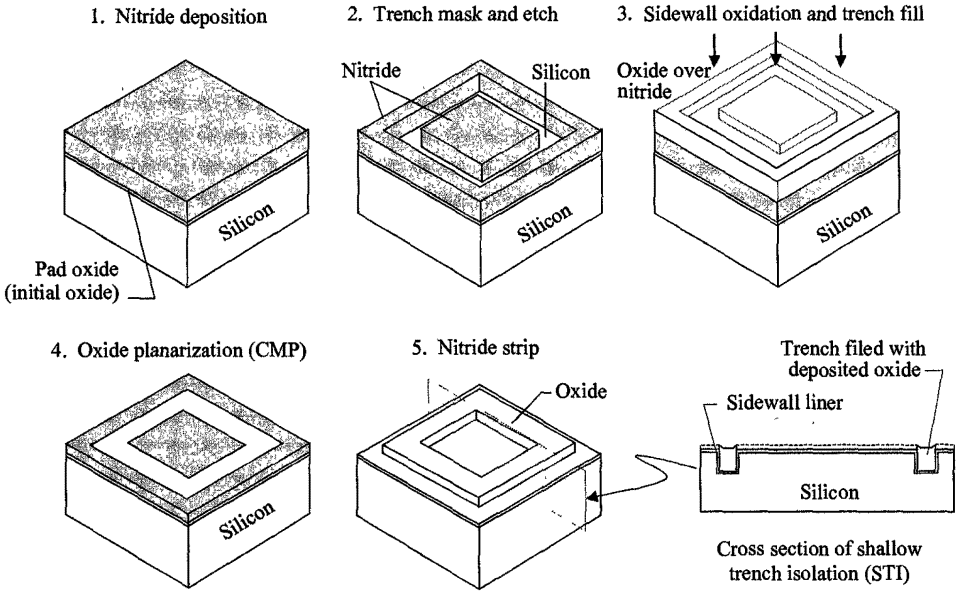
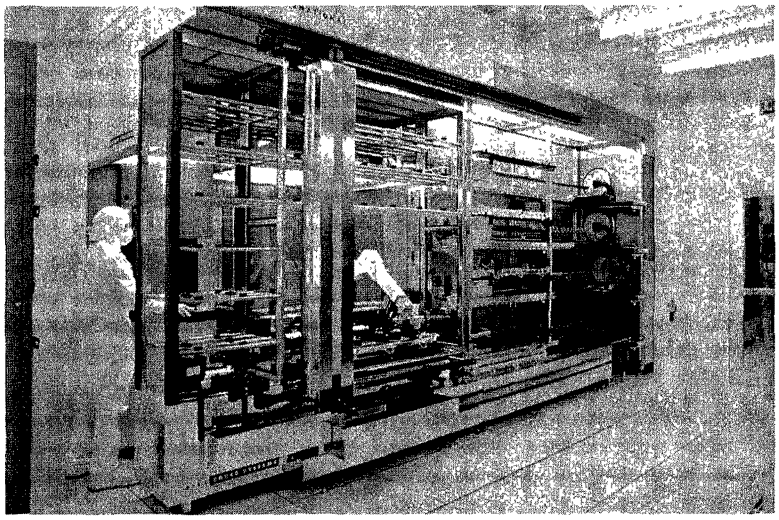


FIGURE 10.15 STI Oxide Liner

Silicon nitride is usually deposited by chemical vapor deposition (see Chapter 11), followed by etching to open up the selective areas for thermal oxidation. The silicon surfaces on the wafer covered by the silicon nitride layer do not oxidize as long as the nitride mask is sufficiently thick (the nitride mask covering the silicon will oxidize slightly). Silicon nitride can be grown on silicon by reacting nitrogen with surface silicon atoms at elevated temperatures, usually above 900°C. However, nitride grown using this method can only be grown to a very thin layer (about 5 nm), which limits its use to only special applications that take advantage of its high dielectric constant.[31]

Stress in Silicon Dioxide ■ Silicon dioxide stress is undesirable because it can contribute to wafer warpage and possibly to defect generation in the form of a slip in the silicon wafer. Measured stress in a thermally grown SiO_2 film is found to be compressive and have a relatively small magnitude. The stress in an oxide film results from the difference between the coefficients of thermal expansion of Si and SiO_2.[32] Stress from an oxide film layer can cause the wafer to bow so that the oxidized surface becomes convex. Optical measurements of the curvature can be used to quantify stress.

Oxidation-Induced Stacking Faults ■ Wet or dry thermal oxidation can cause *oxidation-induced stacking faults* (*OISFs*) at the interface between Si and SiO_2. Recall from Chapter 4 that stacking faults are a form of unit-cell dislocation due to layer stacking errors. It is believed that OISFs are caused by incomplete oxidation at the Si/SiO_2 interface, which leads to excess interstitial silicon in this region.[33] Stacking faults can cause increased leakage current if the fault is in the vicinity of a pn junction. OISF formation is greatly reduced by performing thermal oxidation with chlorine because chlorine promotes vacancy formation at the silicon surface, providing a means to remove the excess interstitial silicon atoms.



Horizontal Diffusion Furnace
(Photo courtesy of International SEMATECH)

FURNACE EQUIPMENT

In this section we will review basic furnace equipment technology. Furnace equipment serves a variety of purposes during the wafer fabrication process. Thermal oxide growth, including the formation of the critical gate oxide, is a major reason for the use of furnaces. Other applications include thermal anneal of the wafer surface after ion implantation (Chapter 17); the deposition of a variety of films, such as doped and undoped polysilicon, silicon nitride, and silicon dioxide (Chapter 11); reflow of glass (Chapter 11); and the formation of silicide films (Chapter 12). We will focus on the equipment used for thermal oxidation processes.

There are three basic types of thermal processing equipment used:

- Horizontal furnace
- Vertical furnace
- Rapid thermal processor (RTP): single-wafer

The *horizontal furnace* was the workhorse in thermal wafer processing since the early days of the semiconductor industry. Its name derives from the horizontal position of the quartz tube where wafers are located and heated. This furnace was largely replaced in the early 1990s by the



Vertical Diffusion Furnace
(Photo courtesy of International SEMATECH)

vertical furnace, mainly because the vertical furnace is easier to automate, improves operator safety, and reduces particulate contamination.[34] The vertical furnace, also called a *vertical diffusion furnace,* or VDF, has better temperature control and uniformity than a horizontal furnace. Both the horizontal and vertical furnaces are considered conventional *hot wall* batch furnaces because both the wafer and the furnace walls are heated, and they generally process large quantities or batches of wafers (100 to 200 wafers). Conventional furnaces ramp up and down the temperature of the wafers at about 20°C/minute or less.

The *rapid thermal processor* (*RTP*) is a small, fast-heating system that typically processes a single wafer at a time with a radiant heat source and cooling source. The RTP is also referred to as a rapid thermal anneal (RTA) system when used to anneal the silicon substrate (see Chapter 17). Due to its extremely fast localized heating times, the RTP only heats the wafer (not the wall of the furnace). The typical RTP is able to achieve up and down ramp rates in the tens of degrees per second range, with models available that use dual-sided wafer heating to attain ramp-up rates up to 250°C/second.[35] The RTP has been in use since the late 1980s and is used in areas such as barrier-layer formation (see Chapter 12) and oxide reflow.

HORIZONTAL VERSUS VERTICAL FURNACES

A comparison of performance factors for the conventional horizontal and vertical furnaces is given in Table 10.3.[36]

TABLE 10.3 Performance Comparison of Horizontal and Vertical Furnace Systems

Performance Factor	Performance Objective	Horizontal Furnace	Vertical Furnace
Typical wafer loading size	Small, for process flexibility	200 wafers/batch	100 wafers/batch
Cleanroom footprint	Small, to use less space	Larger, but has four process tubes	Smaller (single process tube)
Parallel processing	Ideal for process flexibility	Not capable	Capable of loading/unloading wafers during process, which increases throughput
Gas flow dynamics (GFD)	Optimize for uniformity	Worse due to paddle and boat hardware. Buoyancy and gravity effects cause nonuniform radial gas distribution.	Superior GFD and symmetric/uniform gas distribution
Boat rotation for improved film uniformity	Ideal condition	Impossible to design	Easy to include
Temperature gradient across wafer	Ideally small	Large, due to radiant shadow of paddle	Small
Particle control during loading/unloading	Minimum particles	Relatively poor	Improved particle control from top-down loading scheme
Quartz change	Easily done in short time	More involved and slow	Easier and quicker, leading to reduced downtime
Wafer-loading technique	Ideally automated	Difficult to automate in a successful fashion	Easily automated with robotics
Pre- and post-process control of furnace ambient	Control is desirable	Relatively difficult to control	Excellent control, with options of either vacuum or neutral ambient

Horizontal furnaces are still in use and have undergone advances in technology, leading to some renewed interest in their use in fabs. Their low cost relative to vertical furnaces makes them

appealing to wafers with geometries larger than 0.5 μm. This capacity permits a mix-and-match approach, with horizontal furnaces used for certain less-demanding applications and vertical furnaces used for critical applications.

Both horizontal and vertical furnaces have been configured for atmospheric oxidation and diffusions, as well as low-pressure chemical vapor deposition (LPCVD) applications. Examples of materials LPCVD furnaces are used to deposit as thin layers are SiO_2, Si_3N_4, and polycrystalline silicon (poly) on silicon wafers. The topics of LPCVD and diffusion will be discussed in Chapter 11.

Vertical Furnace

The vertical furnace was first introduced in the early 1990s. The change from horizontal furnaces was driven primarily by the reduced cleanroom footprint achieved by vertical furnaces and the improvement in automated handling. To understand basic furnace technology, we will analyze a conventional vertical furnace. There are five major control systems to a vertical furnace system (these same five systems also apply to a horizontal furnace):

- Process chamber
- Wafer transfer system
- Gas distribution system
- Exhaust system
- Temperature control system

A block diagram of a vertical furnace system showing the five major systems is shown in Figure 10.16. Note that these are all under the control of a single microcontroller.

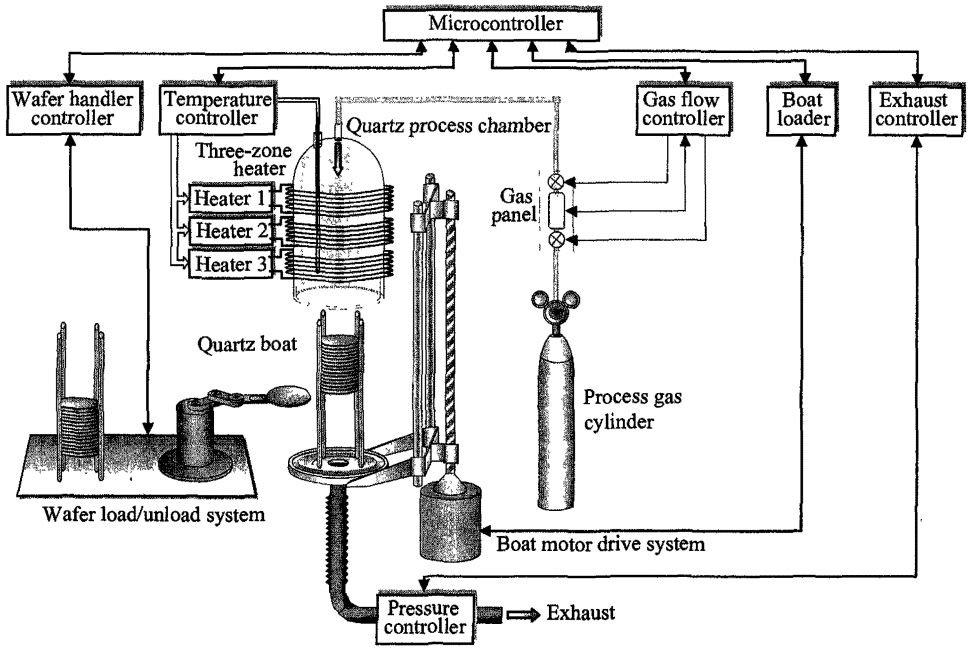


FIGURE 10.16 Block Diagram of Vertical Furnace System

Process Chamber ■ The *process chamber,* or *furnace tube,* is where wafers are heated in the furnace. A vertical furnace consists of a vertical quartz bell-jar style furnace tube that surrounds the wafers for thermal processing in the furnace, a heating element with multiple heat zones, and a heating jacket (see Figure 10.17 on page 242). Furnace tube is also a term used to describe the heated chamber on horizontal furnaces. The furnace tube must be easy to remove so that it can be cleaned when required.

Furnace Tube Materials. The wafers are placed in a vertical *wafer boat* that holds the wafers horizontal in the furnace tube. This boat and the other high-temperature components for the

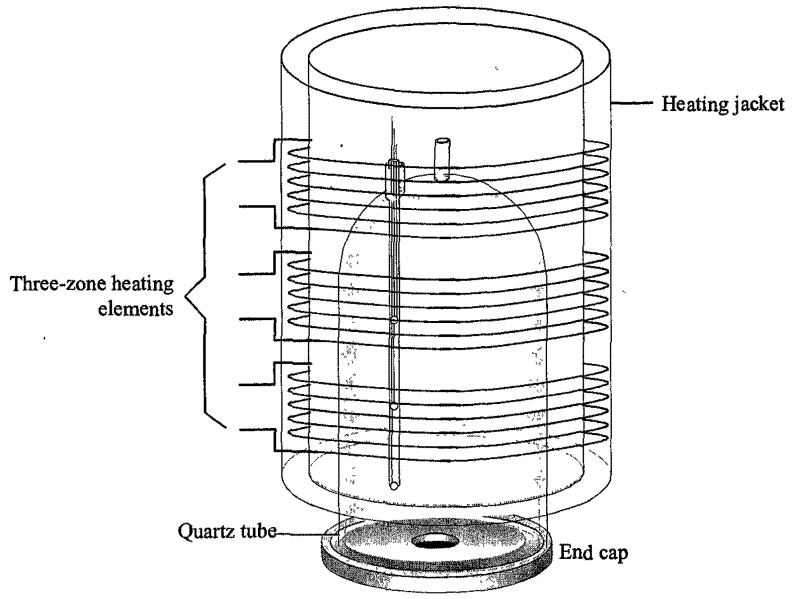


FIGURE 10.17 Vertical Furnace Process Tube

furnace tube are made of amorphous quartz that is resistant to high temperature. Quartz is single crystal SiO_2, while amorphous quartz used for high-temperature applications is fused silica. Furnace containers are commonly called *quartzware,* and the supporting structure that holds wafers in the furnace tube is called a *quartz carrier.*

Films are deposited on the vertical wafer boat and also the inside wall of the furnace tube during processing, especially when the furnace is used for CVD deposition (see Chapter 11). The deposited films will crack and flake, sometimes after only a few cycles. The particles can become airborne and settle on the wafers causing defects and yield loss. This action requires furnace components such as the process tube to be removed and cleaned to reduce particles. Another option to reduce flaking of films is to manufacture the furnace components with silicon carbide (SiC), since the adhesion of the deposited thin film improves with SiC.[37] The drawback to SiC material for furnace components is that it is substantially more expensive than quartz.

Heat Zones. Each quartz process tube is surrounded by a heating element that can be controlled to produce multiple heating zones. It is common to have anywhere from three to seven zones, with 300-mm furnaces having up to nine zones.[38] The number of zones is important because it facilitates control of the furnace tube to attain a *flat zone* near the middle of the tube where thermal processing takes place. The heating zones on each side of the flat zone serve to optimize the wafer ramp up and ramp down to process temperature. This action allows the temperature in the flat zone to be controlled to less than 0.25°C, even at temperatures of over 1000°C.[39]

Wafers are loaded horizontally into a quartz holder called a tower or boat, which is positioned on a quartz mount referred to as the pedestal. The pedestal and furnace tube rest on a water-cooled base plate. Some furnaces rotate the tower during processing for improved heating uniformity.

An important feature of an advanced vertical furnace is its control over the atmosphere around the wafers in the furnace tube. This atmosphere is referred to as the *ambient.* Some furnaces use load locks to keep the ambient in the furnace tube from being exposed to atmospheric gases, while others have nitrogen purge to remove any remaining gases after processing a batch of wafers. Some designs also include an extra quartz enclosure called a liner or inner tube around the wafers to improve the ambient temperature control.[40]

Heating Element. The heating element in a vertical furnace is a metal resistance wire wrapped around the outside of the process tube to provide uniform heat throughout the zone it is heating. Heating elements for three zones are shown in Figure 10.18. The heater elements are switched on and off in response to signals from the temperature controller using a switching system consisting of silicon-controlled rectifiers (SCRs). This system permits control of the amount of power delivered to the heaters (e.g., 50% power or 100% power).

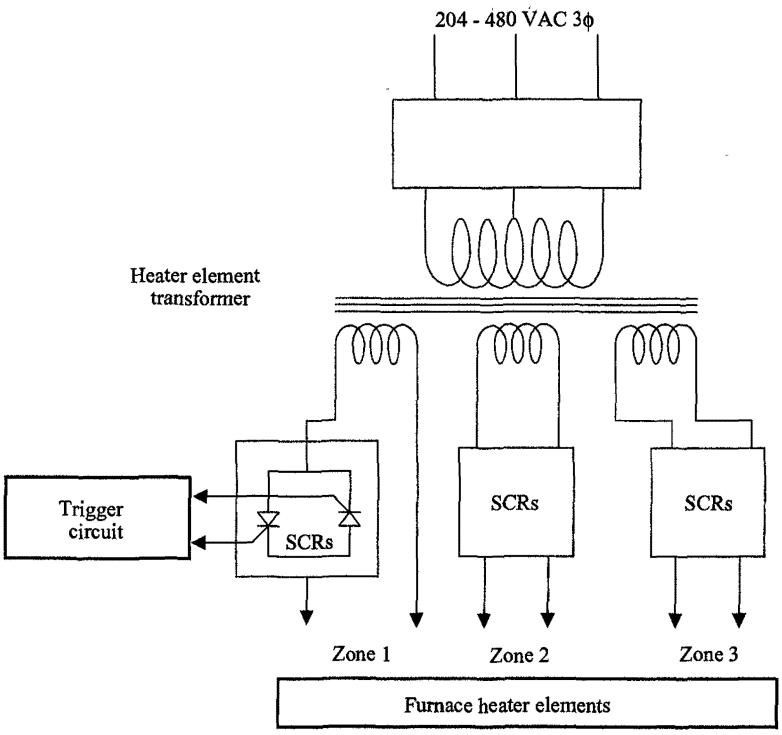


FIGURE 10.18 Heater Element Power Distribution
(Used with permission from International SEMATECH)

Temperature Control. The ability to precisely control temperature in the furnace tube is critical in a furnace. An important part of this temperature control is the sensor known as a *thermocouple* (*TC*) that detects the temperature and provides a corresponding millivolt signal to the furnace controller. Thermocouples are often used because they are rugged, accurate, inexpensive, and work over a wide range of temperatures.

There are multiple thermocouples for each zone of the process tube (see Figure 10.19). The *profile thermocouples* are positioned inside the process chamber in close proximity to the wafer stack, with one in each temperature zone. The profile thermocouples approximate the wafer surface temperature. There are *control thermocouples* (also referred to as spike thermocouples) positioned outside the process tube near the heater element windings in each temperature control zone. These

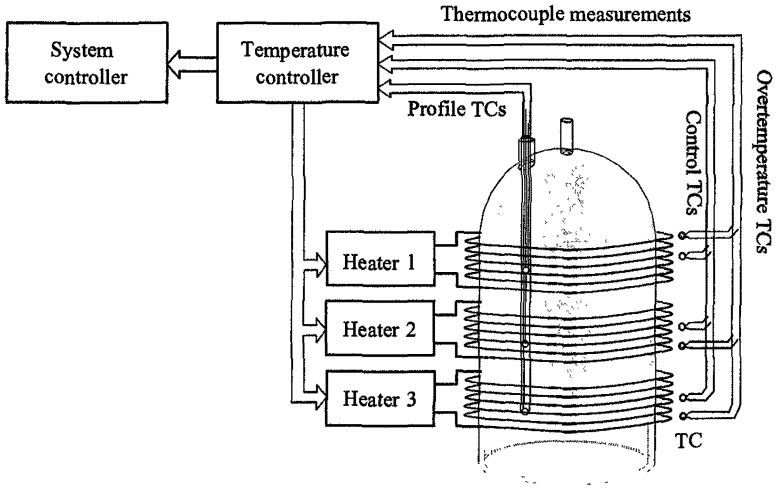


FIGURE 10.19 Locations of Thermocouples in the Furnace Chamber

thermocouples measure the heater element temperature.[41] In addition, there are typically *over-temperature thermocouples* near the control thermocouples to monitor the maximum heater temperature to ensure the furnace never operates at an excessive temperature.

Wafer Transfer System ■ The primary function of the *wafer transfer system* in a vertical furnace is to load and unload wafers from the process chamber. All wafer loading and unloading during the wafer transfer is done using a wafer transfer robot. The robot moves wafers between four stations: cassette station, furnace station, load station, and cooling station. The load station during wafer transfer is maintained at a class 10 or better.

Gas Distribution System ■ The gas distribution system maintains the ambient inside the furnace by delivering the correct flow of gases to the process tube. Depending on the process, there are different bulk and specialty gases that are delivered to the furnace through the distribution system. Some of these typical furnace gases used for oxidation and other furnace processes are listed in Table 10.4.[42]

TABLE 10.4 Common Gases Used in Furnace Processes

Gases	Classifications	Examples
Bulk	Inert gas	Argon (Ar), Nitrogen (N_2)
	Reducing gas	Hydrogen (H_2)
	Oxidizing gas	Oxygen (O_2)
Specialty	Silicon-precursor gas	Silane (SiH_4), dichlorosilane (DCS) or (H_2SiCl_2)
	Dopant gas	Arsine (AsH_3), phosphine (PH_3) Diborane (B_2H_6)
	Reactant gas	Ammonia (NH_3), hydrogen chloride (HCl)
	Atmospheric/purge gas	Nitrogen (N_2), helium (He)
	Other specialty gas	Tungsten hexafluoride (WF_6)

It is important to properly remove gases and their by-products. In a vertical furnace, this removal is done through a port located at one end of the vertical process tube. The gases move into an exhaust manifold to control the direction of flow for each gas. Flammable gases, such as silane (SiH_4), phosphine, and hydrogen use a chamber known as a *burn box* to actually combust the gas in the presence of air far downstream of the process chamber, reducing the flammable gas to less harmful by-products (see Figure 10.20). Special particle filters are then used to remove the solids. The gases pass through a plant *scrubber* where the toxic gases are absorbed. Most furnace exhausts use wet scrubbers that use water solutions to absorb the gases.

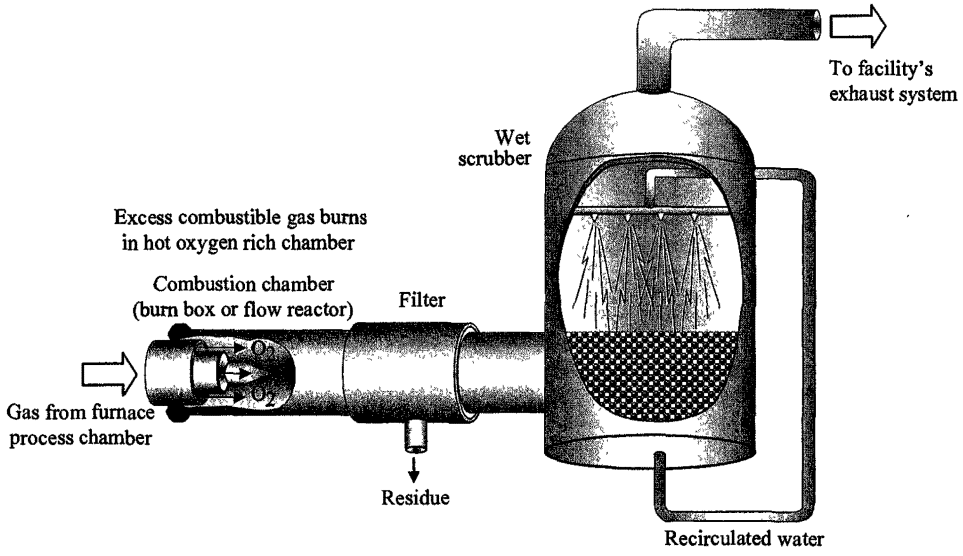


FIGURE 10.20 Burn Box to Combust Exhaust
(Used with permission from International SEMATECH)

Furnace Control System ■ A furnace microcontroller regulates all operations of the furnace, such as process time and temperature control, sequence of process steps, type of gas, the gas flow rates, temperature ramp rates, and wafer loading and unloading. The *temperature ramp rates* are the rate of temperature change that the wafers are exposed to during heating and cooling and are measured in °C/minute. A typical ramp rate for a vertical furnace is 10°C/minute while maintaining the flat zone temperature to within 0.6°C.[43] Other functions such as diagnostics and data collection are also performed by the microcontroller.

Each furnace microcontroller is typically interfaced into a host computer. The host computer is capable of downloading a specific wafer process recipe, which contains all the necessary data for the microcontroller. The host computer will also perform functions such as wafer lot tracking, recipe programming, and automatic scheduling of lots.

Fast Ramp Vertical Furnace

Key issues in furnace performance (and the resulting wafer throughput) are the heating and cooling times required by the furnace. New vertical diffusion furnaces, called *fast ramp furnaces,* are able to quickly raise the temperature of a batch of wafers to the processing temperature (ramp-up time), reducing the time needed to stabilize at the process temperature, followed by a rapid cooldown after processing (ramp-down time). The development of fast ramp furnaces makes it possible to process a wafer batch size of 100 wafers with a ramp-up rate of 100°C/minute and a cooling ramp-down of 60°C/minute.[44] This compares with a ramp-up rate of a few degrees per minute for a conventional vertical furnace.

This fast ramp performance becomes critical with large-diameter wafers due to the larger wafer mass required for heating and cooling. Figure 10.21 shows a typical thermal profile of a fast ramp vertical furnace versus a conventional vertical furnace. Another important aspect of fast ramp furnaces is control of the ambient atmosphere in the furnace. For instance, nitrogen ambient control during loading will inhibit uncontrolled oxide growth and produce more uniform thickness for oxide films.

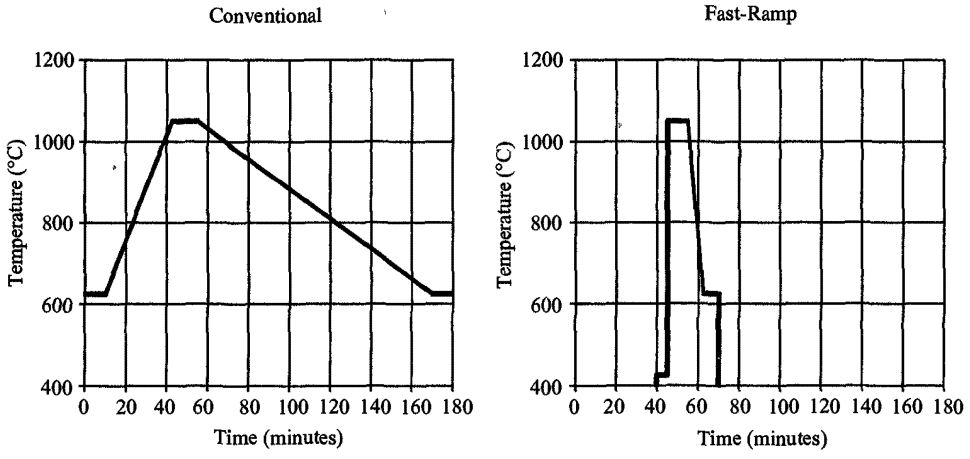


FIGURE 10.21 Thermal Profile of Conventional Versus Fast-Ramp Vertical Furnace
Printed from the June 1996 edition of *Solid State Technology,* copyright 1996 by PennWell Publishing Company

Advanced Temperature Control ■ Improved temperature control and thermal uniformity across the wafer during thermal processes are critical factors for reducing the thermal budget during wafer fabrication. A typical fast-ramp furnace can control the temperature uniformity to within ± 0.1°C at process temperature, while maintaining ± 5°C on all wafers during an 80°C/minute ramp cycle.[45] The key differences in a fast-ramp furnace are fast heating elements, special wafer carriers that increase the wafer-to-wafer gap for more uniform heating and cooling across the wafer, forced air cooling, and properly tuned temperature controllers.

Temperature in conventional furnaces is measured with thermocouples (see previous section), with a thermocouple controlling each furnace zone. A new technique of temperature control is

called *model-based temperature control,* which allows individual wafers to be controlled for heating and cooling rather than merely controlling the ambient of the furnace. This is done by optimizing the control software for each individual furnace with the size of the wafer load. The furnace controller is able to monitor each zone relative to one another to optimize the temperature at the wafer.

Another factor for fast ramp furnaces is the size of the load. There is a trade-off between a large batch size of 150 to 200 wafers and the ramp rate. Increasing the ramp rate for a large batch of wafers fast will create stress on the wafers because the edge circumference of the wafers will heat up as quickly as the furnace does, while the temperature at the center of the wafer will lag by several hundred degrees. Fast ramp furnaces will typically use a smaller batch size of 50–100 wafers to assist in increasing the ramp rate. The smaller batch size also improves the flow of parts through the furnace process because there are fewer wafers to batch together for a run.

Rapid Thermal Processor

The rapid thermal processor (RTP) is a method of heating a single wafer to a temperature range of 400 to 1300°C in a very short time (often fractions of a second). The main advantages for an RTP over a conventional vertical furnace are:

- Reduced thermal budget
- Minimized dopant movement in the silicon
- Ease of clustering multiple tools
- Reduced contamination due to cold wall heating
- Cleaner ambient because of the smaller chamber volume
- Shorter time to process a wafer (referred to as cycle time)

The comparison of a conventional vertical furnace to a rapid thermal processor is outlined in Table 10.5.[46]

TABLE 10.5 Comparison of Conventional Vertical Furnace and RTP

Vertical Furnace	RTP
Batch	Single-wafer
Hot wall	Cold wall
Long time to heat and cool batch	Short time to heat and cool wafer
Small thermal gradient across wafer	Large thermal gradient across wafer
Long cycle time	Short cycle time
Ambient temperature measurement	Wafer temperature measurement
Issues:	Issues:
Large thermal budget	Temperature uniformity
Particles	Minimize dopant movement
Ambient control	Repeatability from wafer to wafer
	Throughput
	Wafer stress due to rapid heating
	Absolute temperature measurement

RTP Design ■ A schematic of a rapid thermal processor (RTP) is shown in Figure 10.22. A single wafer is rapidly heated in a chamber, commonly called a *reactor,* under atmospheric conditions or at low pressure. The RTP has a gas-handling system and a computer that controls system operation.

Most RTPs use multiple tungsten halogen lamps organized into zones as the heat source. The lamps are usually located on the top and bottom of the wafer and can range from 25 lamps to over 150 lamps. They are configured in multiple zones, such as from 4 to 14 zones to permit temperature contouring on the wafer. This contouring can compensate for nonuniform heating and cooling that may occur during ramp up and ramp down in cold wall systems. The silicon wafers are heated by selectively absorbing radiation from the lamps, which produce short-wavelength radiation. In this manner, RTP transfers energy between a radiant heat source and a wafer and does not heat the

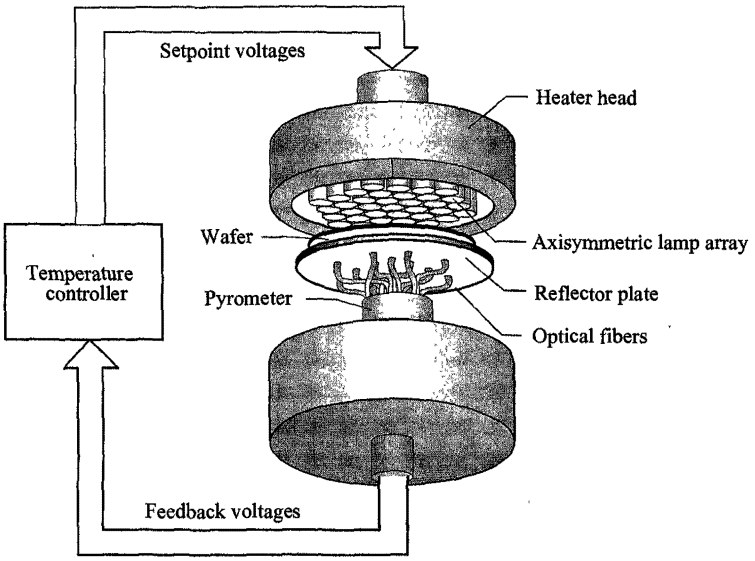
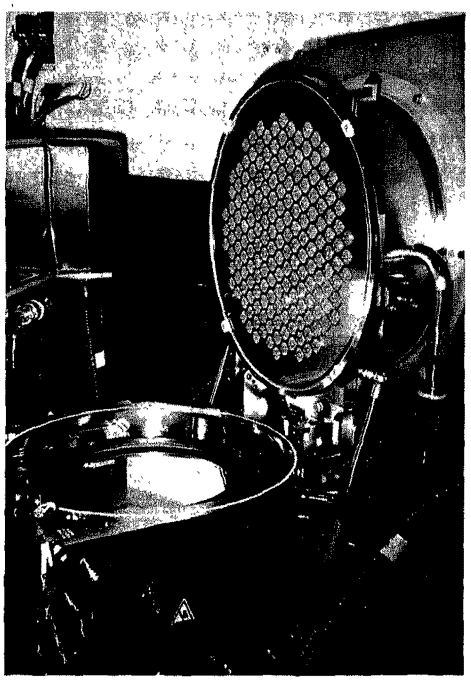


FIGURE 10.22 Rapid Thermal Processor (RTP)

surrounding reactor wall. This is why the term *cold wall* is used. A cold wall reactor is usually polished metal (such as stainless steel to increase reflectivity), with a quartz window to pass the radiation from the tungsten halogen lamp heat sources. Note that there are some systems that use lamps or resistance heaters to heat a susceptor, which then heats the wafer through thermal contact. In this case, the RTP would be classified as a warm or hot wall heater.

Temperature control in an RTP is done with thermocouple or optical pyrometers. Thermocouples are used for physical contact with the wafers to establish the actual wafer temperature. While relatively reliable, thermocouples have a slow response time and suffer from a shortened life span at higher temperatures.

An *optical pyrometer* can measure temperature from a distance with a fast response time. This measurement is done by sensing the emitted infrared radiation of the wafer while it is being heated. However, the emissivity (amount of emitted radiation) of the wafer surface can vary depending on the type of film on the top or bottom surface of the wafer. Recent advances in optical pyrometry



Rapid Thermal Processor
(Photo courtesy of Advanced Micro Devices)

have made emissivity less of a factor by utilizing solid-state pyrometers for better reliability and temperature measurement control.

The RTP has been hampered in the past with poor temperature uniformity across the wafer and nonequilibrium heating.[47] These problems lead to crystal slip and wafer warpage. There have been significant advances in RTP systems in the past few years, resulting in improved temperature control.

Equipment Integration ■ An RTP introduces flexibility into thermal processing because it can be integrated with other process steps into a multichamber cluster tool. This is beneficial because all processing steps occur in a vacuum environment to eliminate concerns with native oxide and contamination. The movement of single wafers through a cluster tool also reduces the time it takes to process wafers by reducing the waiting times associated with large wafer batches. Single wafers can also be processed in parallel with one another. These benefits occur in conjunction with the important action of reducing the thermal budget of the wafer.

RTP Applications ■ The RTP has become widely used in many processes throughout wafer fabrication. Advances in RTP chamber design and temperature uniformity are making this equipment acceptable with better temperature uniformity across wafers. Examples of operations that often use the RTP in wafer fabrication are:

- Anneal of implants to remove defects and activate and diffuse dopants
- Densification of deposited films, such as deposited oxide layers
- Borophosphosilicate glass (BPSG) reflow
- Anneal of barrier layers, such as titanium nitride (TiN)
- Silicide formation, such as titanium silicide ($TiSi_2$)
- Contact alloying

The first widely used application for RTP processing was annealing after ion implant. The advantage of using an RTP over a conventional diffusion furnace is that the wafer spends a much shorter time at temperature, thus reducing the thermal budget. However, the uniform temperature control of conventional furnaces makes this approach competitive.

OXIDATION PROCESS

The goal of thermal oxidation is the growth of a defect-free, uniform layer of SiO_2 at the thickness required. The type of oxidation process conditions used for a particular wafer fabrication step depends on the thickness of the oxidation layer and the properties required. Thin oxides, such as gate oxides, are usually grown in dry oxygen. Because sodium ion contamination is a concern, HCl is added to the O_2 supply during the oxidation of high-quality oxides. Where thick oxides are required, such as the field oxidation layer, then steam is used (HCl is not used in steam oxidation processes). Higher pressure during the growth process allows thick oxide growth to be achieved at a reduced temperature in reasonable time periods. The flow illustrated in Figure 10.23 shows the typical steps for thermal oxidation.

Pre Oxidation Cleaning

Cleanliness of the wafer is critical to achieve high quality oxidation (see Chapter 6). Contaminants such as particulates and mobile ionic contaminants (MICs) will have serious impact on device performance and yields. For example, the presence of MICs in the thermally grown gate oxide structure can result in long-term changes in the device threshold voltage as the MICs drift from the gate oxide to the Si/SiO_2 interface.[48] This condition is detrimental to device electrical performance. For a thermal oxidation process, the ability to keep MICs and particulate matter out of the system depends on maintaining the system to a high state of cleanliness control. The following areas are critical for minimizing contamination:

- Maintenance of the furnace and associated equipment (especially quartz components) for cleanliness
- Purity of processing chemicals
- Purity of oxidizing ambient (the source of oxygen in the furnace)
- Wafer cleaning and handling practices

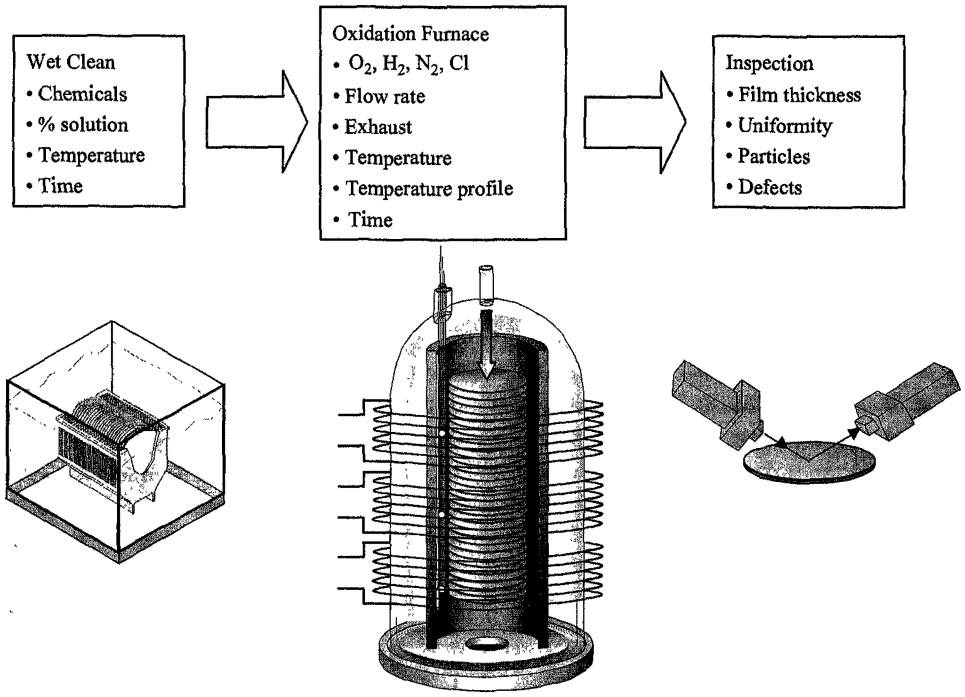


FIGURE 10.23 Thermal Oxidation Process Flow Chart

The basic wet-cleaning chemistries that have withstood the test of time are the RCA SC-1 and SC-2 clean systems and the piranha clean (sulfuric acid, hydrogen peroxide, and water mixture). These wafer cleaning methods and their modifications were discussed in Chapter 6. A wide selection of cleaning equipment is found in the oxidation area of the fab, including manual and automated wet sinks, ultrasonic systems (megasonics), acid spray processors, and rinser/dryer systems.

Oxidation Process Recipe

For thermal oxidation to occur, certain process conditions are followed in a specific format in the furnace equipment. This is known as a *process recipe.* For wafer fabrication, it is common that the parameters of a process recipe are stored in a software database and downloaded into the tool software for a particular wafer lot. We will review a sample process recipe for dry oxidation in a batch furnace.

Dry Thermal Oxidation ■ The most common reason for dry thermal oxidation is to grow the very thin layer of oxide used as the gate oxide.[49] As the devices become smaller in order to increase the component density on a chip (following Moore's law from Chapter 1), then the device dimensions and electrical parameters must also be scaled down proportionally. Gate oxide layers are considered the most critical oxidation application in wafer fabrication, with a gate oxide thickness on the order of 20–40 Å for 0.18-μm devices and 20–30 Å for 0.15-μm devices.[50] Dry thermal oxidation produces a high-quality oxide that is uniformly dense, free of pinholes, and reproducible. It is important for gate oxides to be produced in a clean, sodium-free system that is truly dry, since as little as 25 ppm of water will alter the oxide growth rate and properties of the oxide.[51] There is ongoing research to evaluate other replacement materials that could function as a dielectric in the gate structure, but there is no suitable alternative in production.

Process Recipe ■ Table 10.6 shows a sample process recipe for dry oxidation of a gate oxide.[52] During idle conditions, the process recipe shows there is a nitrogen (N_2) purge gas introduced into the chamber. It flows continuously until the process recipe is started and then the purge is turned off and N_2 process gas is turned on. Once the furnace tube is loaded, the temperature is raised from idle at 850°C to 1,000°C at a ramp rate of 20°C per minute. The wafers have a five-minute stabilization period. Since this is dry oxidation, oxygen at a flow rate of 2.5 slm is introduced into the process chamber. HCl flows in at 67 sccm and is used to reduce the interface charges and to getter MICs. During the anneal step, which is done to minimize charge buildup, the O_2 and HCl are turned off and N_2 is turned on. Oxygen still diffuses through the oxide even though the oxide flow has been cut off, which is taken into account to produce stoichiometric (uniform composition) silicon dioxide. A five-minute period is allowed to unload the furnace tube.

TABLE 10.6 Process Recipe for Dry Oxidation Process

Step	Time (min)	Temp (°C)	N_2 Purge Gas (slm)	Process Gas N_2 (slm)	Process Gas O_2 (slm)	Process Gas HCl (sccm)	Comments
0		850	8.0	0	0	0	Idle condition
1	5	850		8.0	0	0	Load furnace tube
2	7.5	Ramp 20°C/min		8.0	0	0	Ramp temperature up
3	5	1000		8.0	0	0	Temperature stabilization
4	30	1000		0	2.5	67	Dry oxidation
5	30	1000		8.0	0	0	Anneal
6	30	Ramp –5°C/min		8.0	0	0	Ramp temperature down
7	5	850		8.0	0	0	Unload furnace tube
8		850	8.0	0	0	0	Idle

Note: Gas flow units are slm (standard liters per minute) and sccm (standard cubic centimeters per minute).

QUALITY MEASUREMENTS

Representative quality measures for thermal oxide growth are listed in Table 10.7.[53]

TABLE 10.7 Quality Measures for Thermal Oxide Layer

Quality Parameter	Types of Defects	Remarks
1. Oxide thickness.	A. Thickness of gate oxide outside of specification (representative gate oxide specification is 20 ± 1.5 Å).	Possible reasons for this problem are: • Incorrect process gas flow (e.g., MFC calibrated improperly). Since HCl enhances oxidation rate, verify HCl:O_2 ratio is correct. • Verify O_2 leak integrity of the furnace with a bare silicon test wafer. • Check the metrology equipment by measuring the oxide thickness against a standard thickness wafer. • Excess native oxide growth due to overexposure of wafer to air either before or after normal furnace oxidation.

Quality Parameter	Types of Defects	Remarks
2. Gate oxide integrity (GOI).	A. Gate oxide breakdown. B. Mobile ionic contamination (MIC) in film.	Gate oxide defects are often related to processing conditions: • Perform a C-V test to demonstrate gate oxide integrity using an unpatterned test wafer. • Perform an oxide-charge analysis on a test wafer with a surface-charge analyzer. • Review the preoxidation clean steps to assess sources of contamination (e.g., particles or MICs). • Verify no contamination from an incoming gas line or defective filter.
3. Particles in the oxide film.	A. Contaminated quartzware. B. Wafer broken inside furnace. C. Contaminated carrier. D. Contaminated gas filter or line.	Actions to correct these sources of particles added during the oxidation process are: • Check cleanliness of quartzware and carrier. • Verify the alignment of the robotic handling systems. • Check incoming gas filters.
4. Particles under the oxide film.	A. Contaminated preoxidation clean.	Source of particles prior to the oxidation step are: • Verify the preoxidation clean steps are properly set up and performed. • Check cleanliness of quartzware and carrier.

When a batch of wafers is loaded into a furnace tube for oxidation, a number of test wafers with bare surfaces (also referred to as unpatterned wafers, as described in Chapter 19) are placed at strategic locations in the tube (see Figure 10.24). These test wafers are used to perform the various evaluations after the oxidation step to ensure the oxide has acceptable quality.

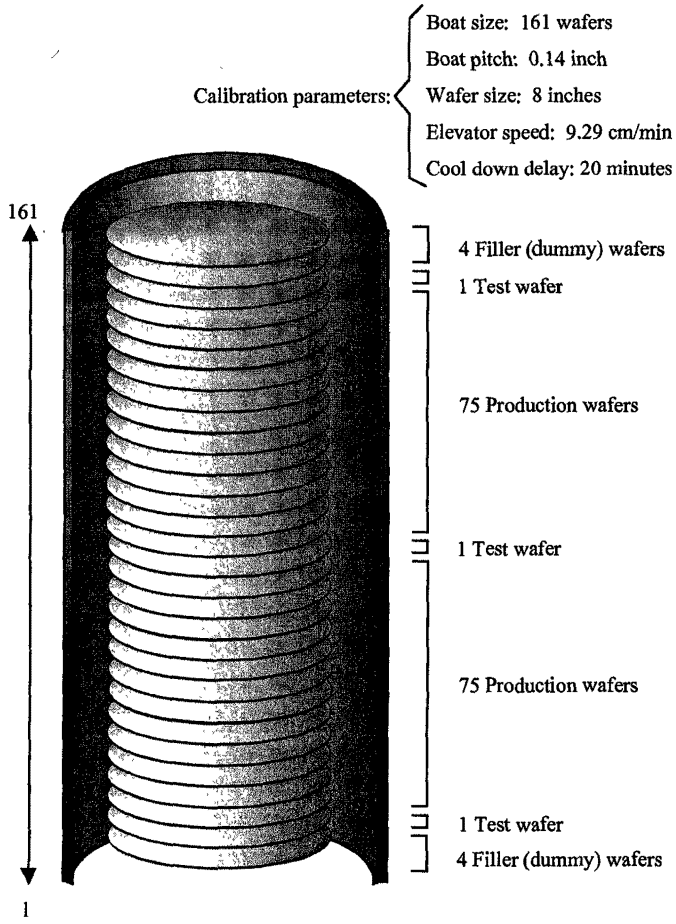


FIGURE 10.24 Wafer Loading Pattern in Vertical Furnace (Used with permission from International SEMATECH)

OXIDATION TROUBLESHOOTING

Common problems and corrective actions for thermal oxidation are described in Table 10.8.

TABLE 10.8 Common Oxidation Troubleshooting Problems

Problem	Probable Causes	Corrective Actions
1. Incorrect gas flow into furnace tube.	A. Incorrect process recipe. B. Malfunctioning MFC. C. Incorrect H_2:O_2 ratio for steam process (O_2 starved).	Possible corrective actions for this problem are: • Verify that correct process recipe is downloaded into furnace software for for the wafer being processed. • Check process gas MFCs (O_2, N_2, H_2, Cl) to verify calibration was performed and operating properly. • Verify gas valves are functioning properly (e.g., no leaks, and so on). • Check that no room air is leaking into the furnace tube from outside the quartz or furnace door seal.
2. Incorrect temperature uniformity in vertical furnace chamber.	A. Wrong process recipe for product. B. Incorrect operation of thermocouples.	Possible corrective actions are: • Verify that the correct process recipe is loaded for the wafer being processed. • Assess whether the temperature nonuniformity is during the ramp up, flat zone, and ramp down for corrective action. • Check all thermocouples for proper operation. Verify no degradation due to excessive heat or corrosion. Replace defective thermocouples. Verify no drift in thermocouple reference temperature.
3. Inadequate temperature uniformity in rapid thermal processor (RTP).	A. Malfunctioning heating system (e.g., tungsten halogen lamps). B. Verify correct operation of of temperature measurement sensor.	Possible corrective actions are: • Verify heating lamps are operating and have the correct lamp intensity. • Verify correct calibration and temperature measurement of the optical pyrometer temperature sensor. Check there is no variation in wafer emissivity by performing reflectivity measurement of the wafer surface.

SUMMARY

Silicon dioxide (or oxide) can be thermally grown or deposited on a wafer. SiO_2 has a tetrahedron cell atomic structure of a silicon atom surrounded by four oxygen atoms and is amorphous. Five uses of oxide film in wafer fabrication are (1) protection and isolation, (2) passivation, (3) dielectric, (4) dopant mask and (5) deposited insulator between metal layers. The thickness of an oxide layer depends on its applications. The chemical reaction for thermally grown dry oxide requires silicon and an oxidizer gas, consumes silicon during the reaction, and is speeded up with the use of wet oxidation. There is an undesirable oxide charge at the Si/SiO_2 interface that can be neutralized by the use of HCl. The growth of thermal oxide follows a linear growth stage up to about 150 Å, followed by a slower parabolic growth stage. Factors affecting the oxide growth rate are dopants, crystal orientation, pressure, and temperature. SiO_2 can be selectively grown or deposited.

The three basic types of thermal processing equipment are a horizontal furnace, vertical furnace, and rapid thermal processor (RTP). The vertical furnace has become the primary batch furnace in fabrication. It consists of a process chamber, wafer transfer system, gas distribution system, exhaust system, and temperature control system. Chamber materials are made of quartz or silicon carbide. The temperature is precisely controlled with the use of multiple thermocouples. The goal in thermal processing is to minimize the thermal budget. Fast-ramp vertical furnaces quickly ramp the temperature up or down with a smaller batch of wafers by using advanced temperature control. The rapid thermal processor (RTP) can reduce the thermal budget by heating wafers with ramp rates of up to hundreds of degrees per second and is often integrated with other process steps. Preoxidation cleaning is critical for attaining a high-quality oxide.

KEY TERMS

grown oxide layer
deposited oxide layer
topography (surface topology)
thermal budget
thermal oxide (thermal silicon dioxide [SiO_2])
tetrahedron cell
surface passivation
field oxide layer
gate oxide structure
gate oxide integrity
dry oxygen
wet oxidation
pyrogenic steam
diffusion
Fick's laws
fixed oxide charge
interface-trapped charge
mobile oxide charge
oxide-trapped charge
rate of oxide growth
linear stage
reaction-rate controlled
parabolic stage
diffusion controlled
local oxidation of silicon (LOCOS)
bird's beak effect
pad oxide
shallow trench isolation (STI)
oxidation-induced stacking faults (OISFs)
horizontal furnace
vertical furnace
hot wall batch furnaces
rapid thermal processor (RTP)
process chamber or furnace tube
wafer boat
quartzware
quartz carrier
flat zone
ambient
thermocouple (TC)
profile thermocouples
spike thermocouples
over-temperature thermocouples
wafer transfer system
burn box
scrubber
temperature ramp rates
fast ramp furnaces
model-based temperature control
reactor
cold wall
optical pyrometer
process recipe

REVIEW QUESTIONS

1. What is the difference between a grown and a deposited oxide layer?
2. Describe what topography is on the wafer surface.
3. Define thermal budget and explain why it is undesirable.
4. State what thermal silicon dioxide is and give another name for it.
5. Describe the atomic structure for silicon dioxide.
6. What is surface passivation and why is it beneficial?
7. Describe the field oxide layer, and state its range of thickness.
8. Why is the gate oxide thermally grown?
9. Explain gate oxide integrity.
10. Describe how SiO_2 can be used as a dopant barrier.
11. List six applications for thermal oxides in wafer fabrication and give a purpose for each application.
12. State the chemical reaction for dry oxidation. At what temperature range does this reaction usually take place?
13. State the chemical reaction for wet oxidation. Is this faster or slower than dry oxidation? Why?
14. If an oxide layer is thermally grown to be 2,000 Å thick, how much silicon is consumed?
15. What is diffusion? How does this process occur in thermal oxidation?
16. List the four types of oxide charges that can occur at the Si/SiO_2 interface. Are oxide charges desirable? Why or why not?
17. Give two advantages to using chlorinated agents during oxidation.
18. Describe rate of oxide growth. What parameters influence this rate?
19. Describe the linear growth stage for thermal oxidation. What is the thickness range where it occurs? State the equation that describes this growth.
20. What does it mean to be reaction-rate controlled for thermal oxidation growth?

21. Describe the parabolic growth stage for thermal oxidation. What is the thickness range where it occurs? State the equation that describes this growth.
22. What does it mean to be diffusion controlled for thermal oxidation growth?
23. What effect does doping have on oxide growth?
24. Explain the effect from crystal orientation on oxide growth.
25. Pressure has what effect on oxide growth?
26. What is the effect of plasma on oxide growth?
27. What is LOCOS and how does this process use thermal oxidation? What is a bird's beak in forming an oxidation layer, and why is this condition undesirable?
28. Explain shallow trench isolation (STI).
29. What causes stress in an oxide layer?
30. What are oxidation-induced stacking faults (OISFs)?
31. List the three basic types of thermal processing equipment.
32. What is a hot wall furnace?
33. List five performance factors for horizontal and vertical furnaces, and stipulate which type of furnace is the optimum.
34. What are the five components in a vertical furnace system?
35. Describe the process chamber (or furnace tube).
36. What two common materials are used for process tube materials?
37. How many zones are there typically in a furnace, and why is this number important?
38. What is a wafer boat?
39. What is meant by the term *ambient* when used in a high-temperature furnace?
40. What are the three locations for thermocouples in a vertical furnace? Describe the purpose of each thermocouple.
41. Explain the purpose of the wafer transfer system in a vertical furnace.
42. A gas delivery system in a vertical furnace serves what purpose?
43. Explain what a burn box is.
44. Describe the purpose of the furnace microcontroller.
45. Explain what a fast ramp vertical furnace is. What ramp-up and ramp-down rates can this furnace achieve?
46. Describe the advanced way to control temperature in a fast ramp vertical furnace.
47. What is a rapid thermal processor (RTP)? What are six advantages it has over the conventional furnace?
48. Describe how an RTP heats a wafer. Is an RTP typically a hot wall or cold wall heating system?
49. Describe the purpose of an optical pyrometer in an RTP and how it functions.
50. Why is preoxidation cleaning important?

FURNACE AND RTP EQUIPMENT SUPPLIERS' WEB SITES

Amtech Systems Inc.	http:/www.amtechsystems.com/
Applied Materials	http:/www.appliedmaterials.com/products/
Asahi Glass Electronic Materials	http:/www.agem-usa.com/
ASM	http:/www.asm.com/
Axcelis (formerly Eaton)	http:/www.axcelis.com/
CVD Equipment Corporation	http:/www.cvdequipment.com/
GaSonics International	http:/www.gasonics.com/
Eaton Corporation	http:/www.semiconductor.eaton.com/
Heraeus Amersil Inc.	http:/www.heraeus-amersil.com/
Kokusai Semiconductor Equipment	http:/www.ksec.com/
MRL Industries	http:/www.mrlind.com/
Omega Engineering Inc.	http:/www.omega.com/
Semitool Inc.	http:/www.semitool.com/
Silicon Valley Group	http:/www.svg.com/
TEL, Tokyo Electron Ltd.	http:/www.teainet.com
Tystar Corporation	http:/www.tystar.com/

REFERENCES

1. G. Moore, "The Role of Fairchild in Silicon Technology in the Early Days of 'Silicon Valley,'" *Proceedings of the IEEE* 86–1 (January 1998): p. 62.
2. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics for Future ICs," *Solid State Technology* (September 1998): p. 70.
3. G. Anner, *Planar Processing Primer,* (New York: Van Nostrand Reinhold, 1990), p. 146.
4. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., (New York: Wiley, 1994), p. 452.
5. Ibid.
6. SEMATECH, *Semiconductor Dictionary,* (Austin, TX: SEMATECH, 1995).
7. S. Ghandhi, *VLSI Fabrication,* p. 452.
8. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* (Boston: Kluwer Academic, 1995), p. 40.
9. SEMATECH, "Furnace Equipment Overview," Module 3 in *Furnace Processes and Related Topics,* (Austin, TX: SEMATECH), p. 3–2.
10. Ibid.
11. B. El-Kareh, *Fundamentals,* p. 47.
12. B. El-Kareh, *Fundamentals,* p. 45.
13. D. Schroder, *Semiconductor Material and Device Characterization,* 2nd. ed., (New York: Wiley, 1998), pp. 337–339.
14. J. Mayer and S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs,* (New York: Macmillan, 1990), p. 265.
15. D. Schroder, *Semiconductor Material,* pp. 337–339.
16. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics,* p. 8.
17. G. Anner, *Planar Processing Primer,* p. 176.
18. S. Wolf and R. Tauber, *Process Technology,* vol. 1, *Silicon Processing for the VLSI Era,* 2nd ed. (Sunset Beach: Lattice Press, 2000), p. 269.
19. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics,* p. 6.
20. S. Ghandhi, *VLSI Fabrication,* p. 460.
21. Ibid.
22. Ibid.
23. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics,* p. 7.
24. B. Deal and A. Grove, "General Relationship for the Thermal Oxidation in Silicon," *Journal of Applied Physics* (1965): pp. 3770–78.
25. S. Ghandhi, *VLSI Fabrication,* p. 463.
26. S. Wolf and R. Tauber, *Process Technology,* p. 278.
27. S. Wolf and R. Tauber, *Process Technology,* p. 283.
28. B. El-Kareh, *Fundamentals,* p. 60.
29. S. Wolf and R. Tauber, *Process Technology,* p. 218.
30. S. Ghandhi, *VLSI Fabrication,* p. 462.
31. B. El-Kareh, *Fundamentals,* p. 69.
32. S. Wolf and R. Tauber, *Process Technology,* p. 296.
33. S. Ghandhi, *VLSI Fabrication,* p. 473.
34. P. Singer, "Furnaces Evolving to Meet Diverse Thermal Processing Needs," *Semiconductor International,* (March 1997), p. 85.
35. Industry Watch, "Thermal Processing: Meeting the Challenges of 300 mm," *Semiconductor International,* (April 1998), p. 19.
36. SEMATECH, "Furnace Equipment Overview," module 3 in *Furnace Processes and Related Topics,* pp. 3–6.
37. J. Tomanovich, "LPCVD Components Trend Toward SiC," *Solid State Technology,* (June 1997): p. 135.
38. P. Singer, "Furnaces Evolving," p. 86.
39. A. Helms et al., "Status and Future of Batch, Hot-Wall Furnaces," *Solid State Technology,* (November 1999): p. 83.
40. P. Singer, "Furnaces Evolving," p. 86.
41. C. Porter et al., "Improving Furnaces with Model-Based Temperature Control," *Solid State Technology,* (November 1996): p. 120.
42. SEMATECH, "Gas Delivery Systems," module 8 in *Furnace Processes and Related Topics,* (Austin, TX: SEMATECH, 1994), pp. 8–3.
43. A. Helms et al., "Status and Future," p. 83.
44. L. Peters, "Thermal Processing's Tool of Choice: Single-Wafer RTP or Fast Ramp Batch?" *Semiconductor International,* (April 1998): p. 84.
45. Ibid.
46. R. Fair, "Conventional and Rapid Thermal Processes," *ULSI Technology,* ed. C. Change and S. Sze, (New York: McGraw-Hill, 1996), p. 151.
47. K. Reid and A. Sitaram, "Rapid Thermal Processing for ULSI Applications: An Overview," *Solid State Technology,* (February 1996): p. 64.
48. S. Wolf and R. Tauber, *Process Technology,* p. 225.
49. SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics,* p. 11.
50. L. Peters, "Thermal Processing's Tool," p. 86.
51. S. Ghandhi, *VLSI Fabrication,* p. 466.
52. Adapted from SEMATECH, "Oxidation Processes," module in *Furnace Processes and Related Topics,* p. 13.
53. SEMATECH, "Troubleshooting Process and Equipment Problems," module in *Furnace Processes and Related Topics,* (Austin, TX: SEMATECH, 1994), pp. 1–16.

CHAPTER 11

DEPOSITION

Microchip fabrication is a planar process involving many different film layers on the wafer surface. Deposition is a process that places the film layers on the wafer. Conductor and insulator films are essential to the successful fabrication of semiconductor devices on silicon substrate wafers. Film layering technology is used to manufacture the circuits and to connect the different IC devices, primarily with metal conducting layers sandwiched between insulating dielectric layers.

Many different types of films are deposited on wafers during the fabrication process. In some cases, these deposited films become an integral part of the device structure, while other sacrificial films are used for particular processing steps and then removed. Films deposited during microchip fabrication are often referred to as thin films because they are so thin that their electrical and mechanical properties differ from a thicker bulk film of the same material.

This chapter will discuss the processes and equipment used to deposit thin films, with a focus on dielectric films such as silicon dioxide and silicon nitride and the deposition of polysilicon. The deposition of metal and metal compound thin films is discussed in Chapter 12.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Describe multilayer metallization. Discuss the acceptable characteristics of a thin film. State and explain the three stages of film growth.
2. Provide an overview of the different film deposition techniques.
3. List and describe the eight basic steps in a chemical vapor deposition (CVD) reaction, including the different types of chemical reactions.
4. Describe how CVD reactions are limited, and explain reaction dynamics and the effect of dopant addition to CVD films.
5. Describe the different types of CVD deposition systems, explain how the equipment functions, and discuss the benefits/limitations of a particular tool for film applications.
6. Explain the importance of dielectric materials for chip technology, giving examples of applications.
7. Discuss epitaxy and three different epilayer deposition methods.
8. Explain spin-on-dielectrics.

INTRODUCTION

The design and fabrication of early semiconductor wafers from the MSI and LSI eras was relatively straightforward. This process consisted of fabricating the semiconductor devices in silicon and interconnecting the devices to one metal conducting layer sandwiched between silicon dioxide as the dielectric material. Figure 11.1 on page 258 shows the deposited layers needed to make an early nMOS transistor. This technology was an extension of the first planar transistors made in the SSI era. The critical dimension was well over one micron. The wafer layers were not flat structures due

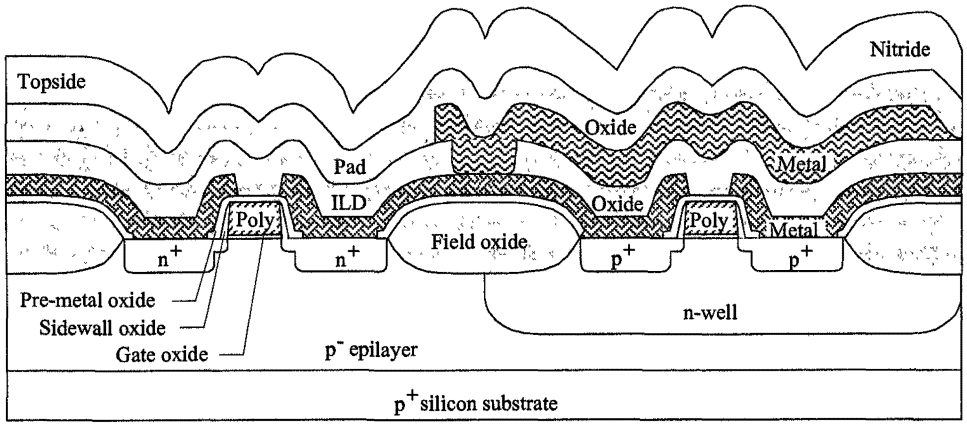


FIGURE 11.1 Film Layers for an MSI Era NMOS Transistor

to varying feature heights, which becomes a limiting factor for high-density chips with multiple metal layers needed in the ULSI era.

As wafers progress to higher density chips with shrinking geometries of 0.18 μm and below, the materials and processes used in wafer fabrication are undergoing dramatic changes. There is a concurrent scaling of all device features to maintain electrical performance. To make electrical connections in today's advanced microchips, six or more conducting metal layers are required. New metal conductor materials are needed to maintain electrical performance. Advanced dielectric materials are deposited between the metal films to provide adequate insulation protection. At the same time, each chip has literally tens of billions of electrical connections between the various metal layers and silicon devices. The ability to deposit reliable thin-film materials is a critical operation in a wafer fab. Thin films is a major process step in wafer fabrication, as shown in the process flow model in Figure 11.2.

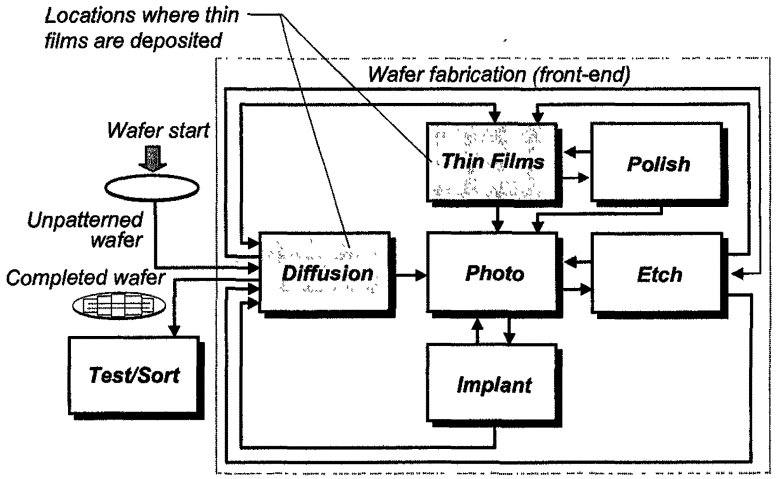


FIGURE 11.2 Process Flow in a Wafer Fab
(Used with permission of Advanced Micro Devices)

Film Layering Terminology

Multilevel metallization refers to the metal and dielectric layers needed to interconnect the densely packed devices on the wafer. An example of this process is shown in Figure 11.3. Without the dielectric insulating layers, electrical shorts would occur (similar to electrical wiring without its insulating cover). The metal layers are connected by openings in the dielectric filmreferred to as *vias.*

Adding metal levels is costly to manufacture. It is estimated that the addition of a single metal level in a CMOS process adds approximately 15% to the total wafer fabrication cost.[1] The

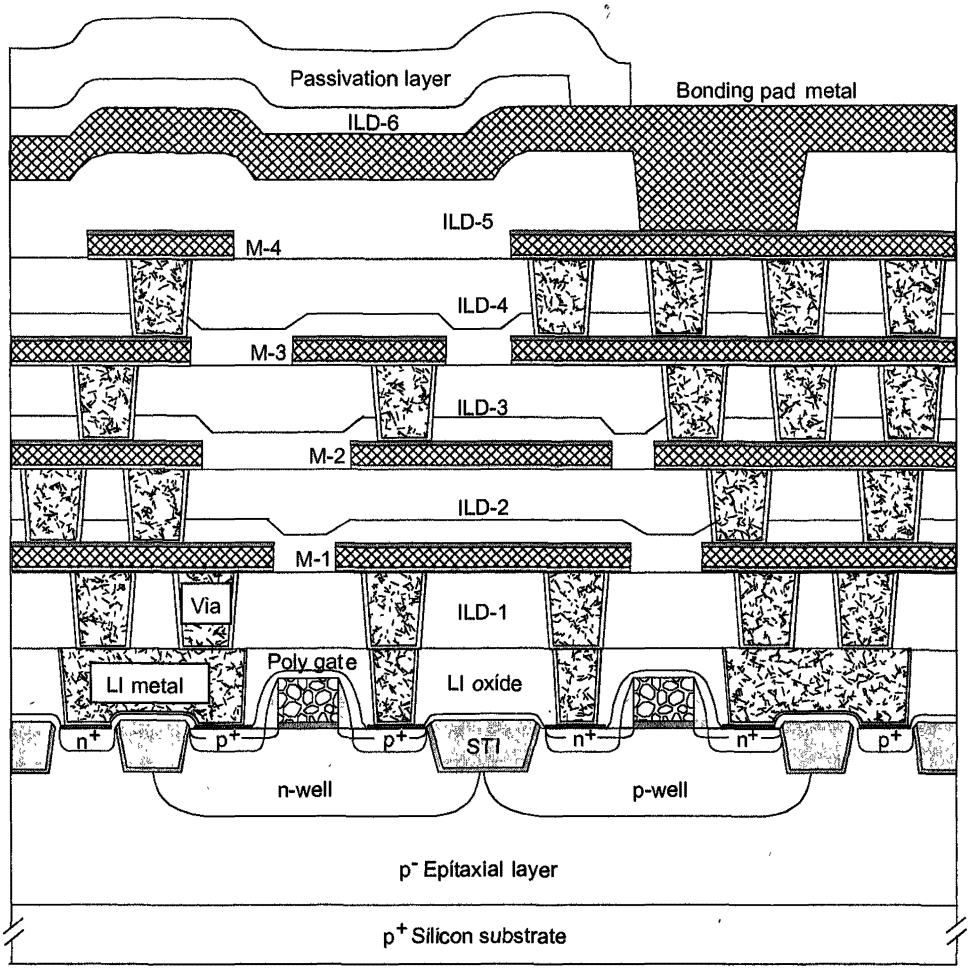
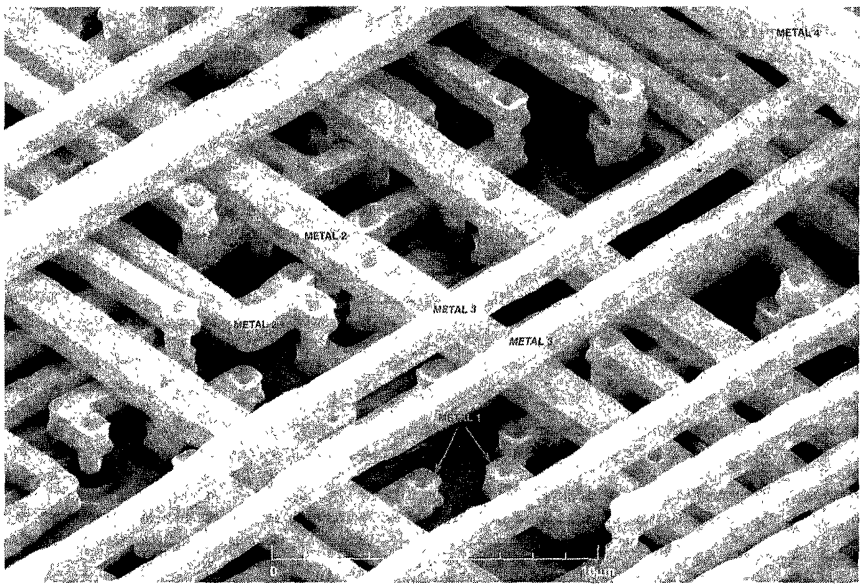


FIGURE 11.3 Multilevel Metallization on a ULSI Wafer

complexity added to wafer manufacturing by additional metal levels makes it important to reduce defects to ensure the chip yield is not impacted. The chip designer considers the tradeoffs among cost, complexity, and performance across the different wafer fab process areas for implementing additional metal levels.

Metal Layers ■ *Aluminum metallization* is the use of aluminum alloy for interconnect wiring. Aluminum alloy has been used since the beginning of semiconductor manufacturing. The Al metal is deposited on the entire surface of the wafer in a solid thin film and then etched to define the width and spacing of the interconnect lines. The industry is undergoing a transition to copper metallization to achieve increased chip speed with fewer process steps (copper metallization is reviewed in detail in Chapter 12). Each metal layer can be referred to as Metal-1, Metal-2, and so on. *Critical layers* are those metal layers with linewidths etched to the critical dimension of the device (e.g., CD of 0.15 μm). For ULSI, the CD typically occurs at the polysilicon features of the gate, the oxide structure, and the metal layers closest to the silicon surface. Critical layers are sensitive to particulate contamination (killer defects) and reliability issues such as electromigration are more pronounced in fine geometry linewidths. *Noncritical layers,* usually the upper metal levels, have much wider line widths, often 0.5 μm and larger, and are less sensitive to particulate contamination. However, factors such as long conductor lengths in the noncritical upper levels can affect chip speed and power consumption.[2]

Dielectric Layers ■ The dielectric layer between the active devices in silicon and the first metal layer is termed the *first interlayer dielectric (ILD-1)*. This layer is also called the *premetal dielectric (PMD)*. ILD-1 is typically a doped silicon dioxide, or glass (explained later in this chapter). The



Metal Layers in a Chip
(Micrograph courtesy of Integrated Circuit Engineering)

important function of the ILD-1 layer is to isolate transistor devices in two ways: electrically from the metal interconnect layers, and physically from contamination sources such as mobile ions. The ILD-1 has a restricted thermal budget in high-performance logic devices in order to avoid degrading transistor characteristics.[3]

The *interlayer dielectric* (*ILD*) is used between different metal layers in the device. The ILD serves as an insulating film between two conducting metals or adjacent metal lines. The ILD has traditionally been silicon dioxide (SiO_2) with a dielectric constant of around 3.9 to 4.0.[4] The dielectric constant is an important property of deposited insulating films because it directly affects circuit speed performance (see the following section).

FILM DEPOSITION

A *thin film* is a thin, solid layer of a material created on a substrate. If a solid material has three dimensions (thickness, width, and length), then a thin solid film has one of its dimensions (usually thickness) much smaller than the other two (see Figure 11.4). The thin film is bonded to a wafer substrate, which has a much thicker dimension than the film. The thin-film surface is so close to the substrate that it has a profound influence on the physical, mechanical, chemical, and electrical properties of the thin film material.[5] The most widely used unit to describe the thickness of a thin film in wafer fabrication is the angstrom (Å).

Thin film deposition in semiconductor fabrication is any process that physically deposits a film on the wafer substrate. The film can be either a conducting, insulating, or semiconducting material. Examples of deposited films are polysilicon, silicon dioxide (SiO_2), silicon nitride (Si_3N_4), polysilicon (silicon in a polycrystalline structure), and metals such as copper and refractory metals (e.g., tungsten).

Thin-Film Characteristics

For a thin film to be acceptable in wafer fabrication, it must have desirable film characteristics. General characteristics of an acceptable thin film for device performance are:[6]

- Good step coverage
- Ability to fill high aspect ratio gaps (conformality)
- Good thickness uniformity
- High purity and density

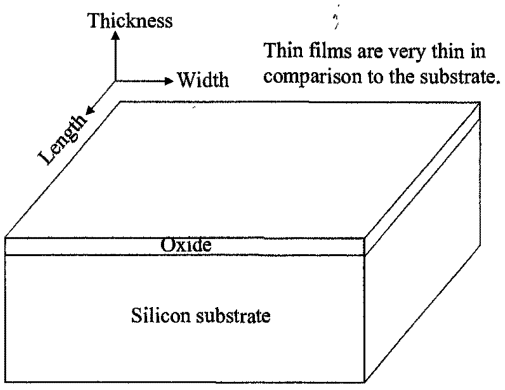


FIGURE 11.4 Solid Thin Film

- Controlled stoichiometries
- High degree of structural perfection with low film stress
- Good electrical properties
- Excellent adhesion to the substrate material and subsequent films

Film-Step Coverage ■ It is desirable for thin films to maintain a uniform thickness over surface features (see Figure 11.5). Steps on the wafer surface occur due to three-dimensional shapes from patterned features that create surface topography. If the film thins excessively at a step, this can cause high stress, electrical shorts, or undesirable induced charges in the device. Films with minimum stress are important, since stress can lead to substrate deformation in a convex or concave shape. All deposition techniques result in the formation of some stress in films.

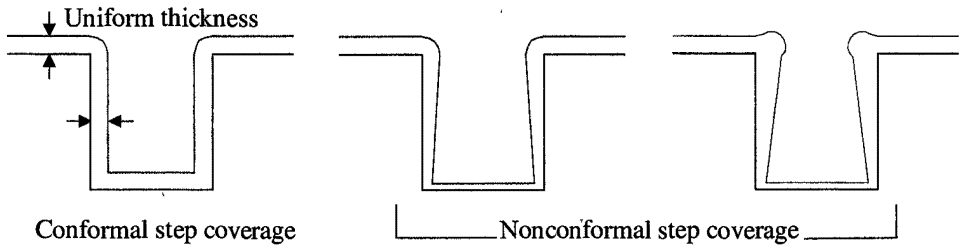


FIGURE 11.5 Film Coverage Over Steps

High Aspect Ratio Gaps ■ A small gap (trench or hole) is characterized by its *aspect ratio,* which is defined as the ratio of its depth to width (see Figure 11.6 on page 262). Aspect ratio is expressed as a ratio, such as 2:1, which in this case means the gap depth is two times the width. A gap can also be specified by its opening width, such as 0.25 μm, which may be different than the CD of other structures on the wafer. The ability to fill very small gaps and holes on the surface of the wafer has become one of the most important film characteristics for devices fabricated with sub-0.25 micron geometries. Examples of gaps that require effective gap-fill capability are vias passing through the interlayer dielectric (ILD) and trenches for shallow trench isolation (STI).[7] A high aspect ratio is typically above 3:1, with some applications having aspect ratios of 5:1 or larger. High aspect ratio gaps make it difficult to uniformly deposit the film, leading to pinch-off and voiding. Deposition processes that can produce uniform, void-free films in high aspect ratio openings are critical in the sub-0.25 μm era of high-density IC circuits with shrinking geometries.

Thickness Uniformity ■ Acceptable thin films are conformal with good thickness uniformity, which means the film follows the outline of the base material in all areas. The resistance of a material will change with varying thickness, which would be undesirable. Thinner layers of film also

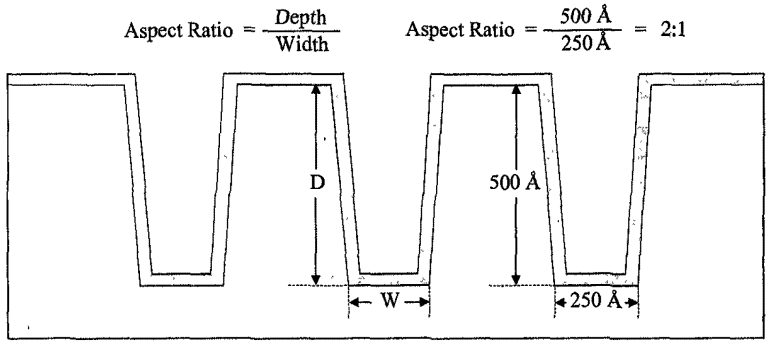
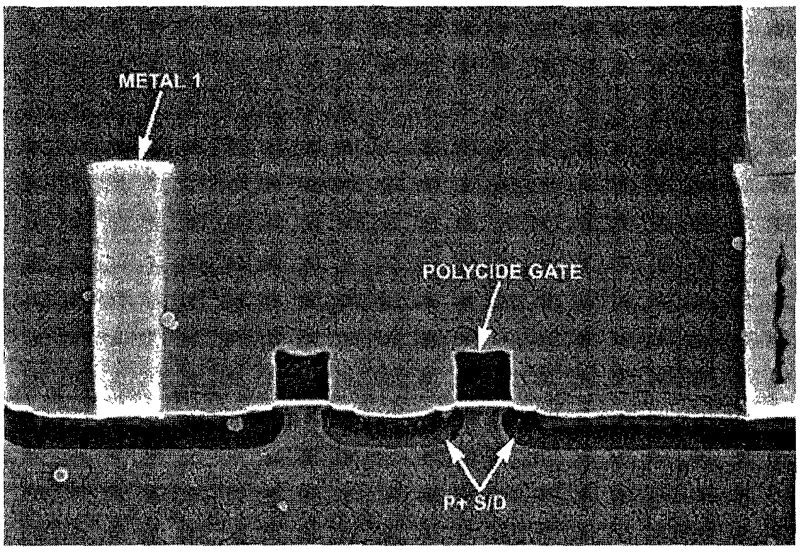


FIGURE 11.6 Aspect Ratio for Film Deposition



High Aspect Ratio Gap
(Micrograph Courtesy of Integrated Circuit Engineering)

tend to have more defects such as pinholes, which leads to less mechanical strength. It is desirable for thin films to have good surface flatness to minimize steps and cracking.

Film Purity and Density ■ High film purity means that the film does not have any unwanted chemical elements or molecules, which is important if the film is to function properly. This includes the avoidance of contaminants such as mobile ionic contaminants (MICs) and particulates. The occurrence of film contaminants such as hydrogen can degrade film properties. The density of the film is an important indicator of film quality because it means the film is free of pinholes and voids. A porous film has lower density and in some instances a lower refractive index than a nonporous film.

Stoichiometry ■ A desirable film will have uniform composition. As chemicals undergo change in a reaction, stoichiometry describes how the quantities of reactants and products are produced once the reaction ends or stabilizes. *Stoichiometry* is the ratio of one component amount to another component amount in a compound or molecule (e.g., the H_2O stoichiometric ratio is 2:1). If you know what is in a reaction, then stoichiometry tells you how much of the ingredients was needed. The chemical reactions that occur in deposition are complex and can lead to films that have a different composition than intended. A goal in deposition is to have the correct quantity of molecules in the reaction so that the deposited film approaches the ratio of the elements in the nominal chemical formula of the incoming gas.

Film Structure ■ The film structure is critical, especially related to the grain size. Materials tend to collect and grow as grains during the deposition process. If the grain size varies in a film, then it

will have varying electrical and mechanical properties that can affect the long-term reliability of the film, especially concerning electromigration (discussed in Chapter 12).[8] The growth of a film can cause undesirable stress in the film layer that deforms the wafer substrate, leading to film cracking, delamination or void formation. An example of this condition is hydrogen contamination in nitride film deposition, which leads to a compressive stress.[9] The objective during film deposition is always to minimize the amount of stress.

Film Adhesion ■ Film adhesion to the substrate material is important for thin films to avoid delamination and cracking. A cracked film can cause surface roughness and allows contamination to pass through the film. For insulating films, cracks can lead to electrical shorts or current-leakage paths. Adhesion of a film to a surface is determined by the cleanliness of the surface and also by the type of material to which the film can alloy. Metals such as chromium, titanium, and cobalt are often useful for their adhesion properties (see Chapter 12). Good adhesion of films in multilevel metallization is critical to maintaining the overall electrical and mechanical integrity of the structure. This is true in the as-deposited condition and after subsequent processing.

Film Growth

A deposited film grows in three distinct stages (see Figure 11.7).[10] The first stage is *nucleation* where clusters of stable nuclei are formed. This stage occurs when the first few atoms or molecules of the reactants combine to form isolated patches of film that attach to the wafer surface. Nucleation occurs directly on the wafer and is essential to further thin film growth. The second stage is the *nuclei coalescence* into clusters, also referred to as *island growth.* These randomly oriented island clusters grow based on surface mobility and the density of the cluster. The island clusters continue to grow and eventually develop into the third stage, a *continuous film* where they meet and form a solid sheet that spreads across the substrate surface.

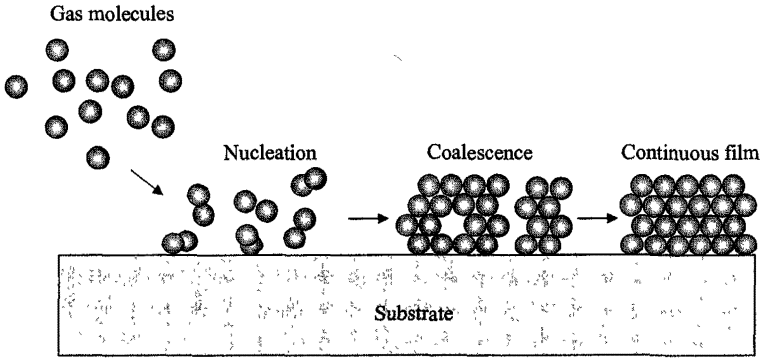


FIGURE 11.7 Stages of Film Growth

The size of the individual cluster, before meeting a neighboring cluster, is related to the surface mobility of the reacting species and the density of the nuclei. A high surface mobility and/or low nucleation rate promote the formation of relatively large clusters. This condition generally leads to polycrystalline films with short-range crystal order. On the other hand, low surface mobility combined with a high nucleation rate leads to amorphous film growth with no short-range order.[11] Low deposition temperatures usually favor amorphous films because the low thermal energy slows the surface mobility of the reacting materials.

Deposited films can be amorphous, polycrystalline or single crystalline.[12] Semiconductor films are used in all three forms. Insulator or metal films are usually amorphous or polycrystalline. Silicon deposited on an oxide layer is polycrystalline, as in the polysilicon used in the gate structure. Epitaxial single-crystal films are necessary for reliable semiconductor properties. To obtain a single-crystal deposited film, it is deposited on a single-crystal wafer substrate.

Film Deposition Techniques

Deposition of materials onto the wafer surface places a continuous thin film on the wafer. The film material derives from an external source, which may be a gas source that produces the material to be deposited through a chemical reaction or a solid target source that requires the physical removal of the material. With the increased complexity of semiconductor processing in the ULSI era, the number of deposited films on a wafer is high. As a result, the deposition of thin films is one of the most common processes in the semiconductor industry.

There are different techniques used in wafer fabrication to deposit films on a substrate. The major deposition methods shown in Table 11.1 are generally grouped into chemical or physical processes.

TABLE 11.1 Techniques of Film Deposition

Chemical Processes		Physical Processes		
Chemical Vapor Deposition (CVD)	**Plating**	**Physical-Vapor Deposition (PVD or Sputtering)**	**Evaporation**	**Spin-On Methods**
Atmospheric Pressure CVD (APCVD) or Sub-Atmospheric CVD (SACVD)	Electrochemical deposition (ECD), commonly referred to as electroplating	DC Diode	Filament and Electron Beam	Spin-on-glass (SOG)
Low Pressure CVD (LPCVD)	Electroless Plating	Radio Frequency (RF)	Molecular Beam Epitaxy (MBE)	Spin-on dielectric (SOD)
Plasma Assisted CVD: • Plasma Enhanced CVD (PECVD) • High Density Plasma CVD (HDPCVD)		DC Magnetron		
Vapor Phase Epitaxy (VPE) and Metal-organic CVD (MOCVD)		Ionized metal plasma (IMP)		

Adapted from F. Barlow III, A. Elshabini-Riad, and R. Brown, "Film Deposition Techniques and Processes," *Thin Film Technology Handbook,* eds. A. Elshabini-Riad and F. Barlow III (New York: McGraw-Hill), pp. 1–2.

This chapter reviews the dielectric thin films deposited with chemical vapor deposition (CVD), epitaxy, and spin-on-dielectric (SOD) methods. All metallization is discussed in Chapter 12, including chemical vapor deposition (CVD) for metals, sputtering, electroplating, and evaporation.

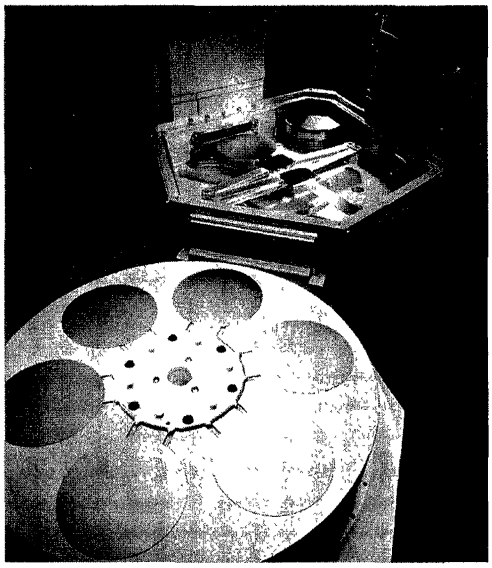
CVD is commonly used for depositing dielectric and metal films. Spin-on-dielectric (SOD) is used for applying liquid dielectric films followed by a high-temperature curing process. Sputtering, or PVD, is the most common method in wafer fabrication for applying metals such as aluminum. Electroplating, widely used in thin film deposition for magnetic recording heads in disk drives, has not been used in silicon wafer fabrication. It has recently become the most promising process for depositing copper metallization. Evaporation is the traditional method for applying a metal layer, but its poor gap-fill property led to its demise at the start of the VLSI era, when it was replaced by sputtering.

CHEMICAL VAPOR DEPOSITION

Chemical vapor deposition (*CVD*) is the process of depositing a solid film on the wafer surface through a chemical reaction of a gas mixture. The wafer surface or its vicinity is heated in order to provide additional energy to the system to drive the reactions. The essential aspects of CVD are:

1. Chemical action is involved, either through chemical reaction or by thermal decomposition (referred to as *pyrolysis*).
2. All material for the thin film is supplied by an external source.
3. The reactants in a CVD process must start out in the vapor phase (as a gas).

CVD deposition by use of a chemical reaction occurs when chemical compounds are mixed and then reacted in a deposition chamber, referred to as a *reactor.* The atoms or molecules deposit on the wafer surface and form the film.



Chemical Vapor Deposition Tool
(Photo courtesy of Novellus Systems, Inc.)

CVD Chemical Processes

There are five basic chemical reactions that can be used in CVD:[13]

1. **Pyrolosis:** a compound dissociates (breaks bonds, or decomposes) with the application of heat, usually without oxygen.
2. **Photolysis:** a compound dissociates with the application of radiant energy that breaks bonds.
3. **Reduction:** a chemical reaction occurs by reacting a molecule with hydrogen.
4. **Oxidation:** a chemical reaction of an atom or molecule with oxygen.
5. **Reduction-oxidation (redox):** a combination of reactions 3 and 4 with the formation of two new compounds.

Among these five general types of reactions, there are many specific CVD reactions used to deposit films on a wafer substrate. The choice of a particular reaction is usually defined by parameters such as deposition temperature (which must be acceptable for the wafer materials), the film properties, and manufacturing issues. An example of a specific CVD reaction is the deposition of a silicon dioxide film by an oxidation reaction of silane with oxygen. This reaction produces silicon dioxide that is deposited on the heated wafer surface, with a by-product of hydrogen.

$$\underset{\textit{silane}}{SiH_4} + \underset{\textit{oxygen}}{O_2} \xrightarrow{\text{(heat)}} \underset{\textit{silicon dioxide}}{SiO_2} + \underset{\textit{hydrogen}}{2H_2}$$

CVD Reaction

The CVD chemical reaction that deposits the solid film takes place on (or very close to) the wafer surface. This is a *heterogeneous reaction* (also referred to as *surface catalyzed*). Some reactions also occur in the gas phase above the wafer surface, which is termed a *homogeneous reaction.* Homogeneous reactions are undesirable because they form gas-phase clusters of the depositing material, which results in poorly adhering, low-density films with higher defects.[14] In CVD chemical reactions, heterogeneous reactions at the wafer surface are favored for the production of high-quality films.

CVD Reaction Steps ■ The fundamental CVD reaction has eight major steps that explain the reaction mechanism. The steps are summarized in the following list and are shown in Figure 11.8:[15]

1. **Gas transport to deposition zone:** Mass transport of gas in the main gas flow region from the reactor inlet to the deposition zone of the wafer.
2. **Formation of film precursors:** Gas-phase reactions leading to the formation of the film precursors (initial atoms and molecules that will constitute the film) and by-products.
3. **Film precursors at wafer:** Mass transport of the film precursors to the wafer growth surface.
4. **Precursor adsorption:** Adsorption (binding) of film precursors to the surface.
5. **Precursor diffusion:** Surface diffusion of film precursors to the film growth sites.
6. **Surface reactions:** Surface chemical reactions leading to film deposition and by-products.
7. **By-product removal from surface:** Desorption (removal) of the by-products of the surface reactions.
8. **By-product removal from reactor:** Mass transport of the by-products in the bulk gas-flow region away from the deposition zone and towards the reactor exit.

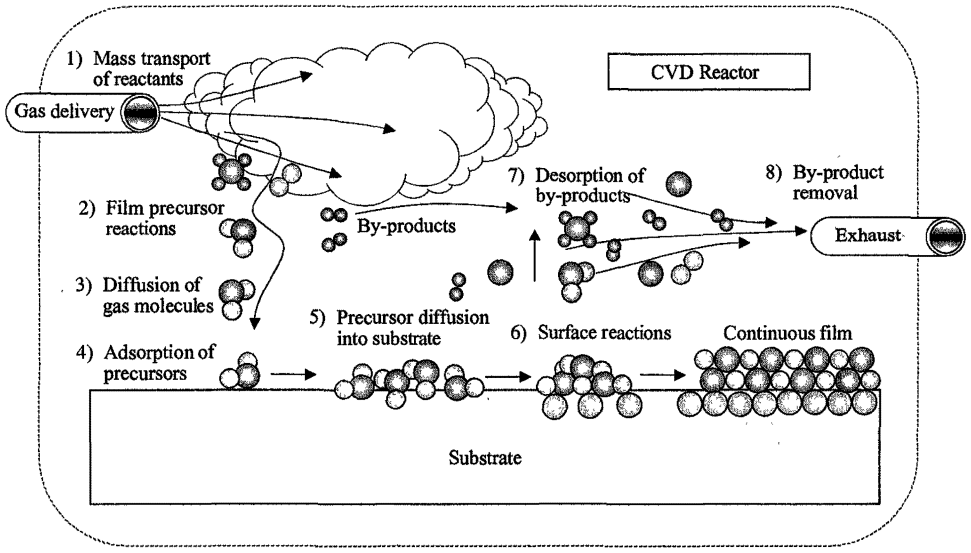


FIGURE 11.8 Schematic of CVD Transport and Reaction Steps
SEMATECH, "Deposition Processes," in *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 6.

Adsorption is the chemical binding that occurs during deposition, causing the gaseous atoms and molecules to chemically attach to the solid wafer surface. *Desorption* is the removal or release of by-products from the wafer surface reactions. In chemical reactions, *species* describes a particular chemical substance that is an atom, ion, or molecule. In gas-phase reactions, there are often intermediate reactions called *precursors* that form a gas species that does not contain the original gas components. In CVD, gas precursors are transported to the wafer surface for adsorption and reaction. For example, for the following three reactions, the first gas-phase reaction shows that when

silane decomposes, a silicon hydride precursor is produced. This precursor then reacts with silane to form disilane. During the intermediate CVD reaction, the silicon hydride is adsorbed onto the wafer surface along with silane. Disilane then decomposes to form the final silicon solid-film product.

1)	SiH_4 (gas) *silane*	→	SiH_2 (gas) *silicon hydride*	+	H_2 (gas) *hydrogen*	(pyrolysis)
2)	SiH_4 (gas) *silane*	+	SiH_2 (gas) *silicon hydride*	→	Si_2H_6 (gas) *disilane*	(Reactive intermediate formation)
3)	Si_2H_6 (gas) *disilane*	→	2Si (solid) *solid film product*	+	$3H_2$ (gas) *by-product*	(Final product formation)

Rate Limiting Step ■ The length of time it takes for the CVD reaction to occur is important for productivity in manufacturing. The surface reaction rate increases with increasing temperature. Since CVD reaction steps are sequential, the slowest step defines the bottleneck in the process. In other words, the step that occurs at the slowest rate will determine the rate of deposition.[16]

The rate of a CVD reaction cannot proceed more rapidly than the mass-transport rate at which reactant gases are supplied from the main gas stream to the wafer substrate. This is true no matter how high the temperature is increased. This situation is referred to as a *mass-transport limited* deposition process. A mass-transport limited process is only weakly dependent on temperature.[17] This means that regardless of the temperature, there is insufficient reactant gas being supplied to the wafer surface to speed up the reaction. Because of this, a high-temperature, higher pressure CVD process that wants to proceed rapidly is usually mass-transport limited—it cannot supply enough reactant gas to the substrate because of diffusion or physical limitations.

At lower reaction temperatures and pressure, the surface reaction rate is reduced because there is less energy available to drive the surface reaction. Eventually the arrival rate of reactants will exceed the rate at which they are consumed by the chemical reaction process at the wafer surface. In this case, the deposition rate is *reaction-rate limited.*[18] This condition is similar to pouring too much liquid through a funnel—the liquid will not flow faster than the funnel permits. Even though more reactant gas is supplied, the reaction will not speed up because the low temperature provides insufficient energy. A reaction-rate limited CVD process is also referred to as *kinetically controlled,* meaning the diffusivity of the reactants is not as critical as the kinetic energy of the reactants at the surface. For reaction-rate limited processes, it is important to maintain a uniform temperature in order to achieve a uniform deposition rate across the wafer surface.

CVD Gas-Flow Dynamics ■ The gas-flow dynamics for CVD are important for uniform film deposition. Gas flow refers to the physical transport of reactant gases to the CVD sites on the wafer (see Figure 11.9 on page 268). At the molecular level, this requires having a sufficient number of molecules at the right place and the right time. This condition is necessary for a stoichiometric reaction to occur between the reactants and products.

The principal factors for CVD gas flow are the mass transport of the reactant gases from the main gas stream to the wafer surface and the chemical reaction rate at the surface. It is assumed that the dominant transfer mechanism from the gas phase to the surface is diffusion.[19] Considering the gas flow right at the wafer surface during a CVD reaction process, there is zero or near-zero gas flow. This condition creates a *boundary layer* of gas flow that increases from the zero gas-flow velocity at the wafer surface (due to frictional forces) to some typical gas flow farther from the surface.[20] These dynamics are shown in Figure 11.10 on page 268. The gas flow farther from the surface reaches some average gas-flow velocity that represents the average velocity of the main gas stream in the CVD reactor. If the boundary layer is narrow, then it can be treated as not moving at all near the wafer surface. In this case, it is referred to as a *stagnant layer.* The size of the stagnant layer will influence the design of the different types of CVD reactors considered later in this chapter.

Pressure in CVD ■ If a CVD reaction occurs at low pressure, then the diffusivity of the gas species increases significantly through the boundary layer to reach the surface. This diffusivity

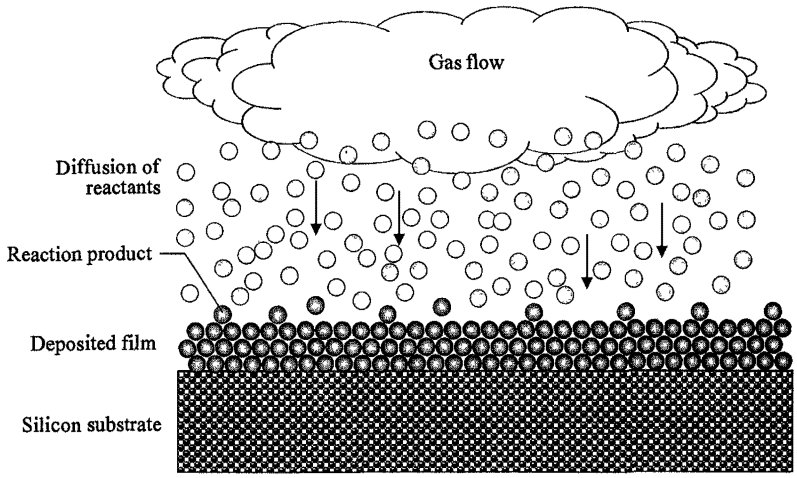


FIGURE 11.9 Gas Flow in CVD

increases the transport of the reactants to the substrate surface (and also increases the removal of by-products away from the substrate). The net effect of lower pressure in a CVD reactor is that the reactants reach the surface more quickly. In this case, the rate-limiting step becomes the surface reaction. At lower pressures, the CVD process is reaction-rate limited. This means that wafers can be stacked vertically and at very close spacing in the reactor since the transport of the reactants from the gas stream to the wafer is not limiting the process.

Doping During CVD ■ The introduction of dopants into SiO_2 during CVD deposition is desirable because it leads to important benefits for wafer fabrication. For instance, the introduction of phosphine (PH_3) to the gas stream during silicon dioxide deposition will lead to the formation of *phosphosilicate glass* (*PSG*). The chemical reaction is:

$$\underset{\textit{silane}}{SiH_4\,(gas)} + \underset{\textit{phosphine}}{2PH_3\,(gas)} + \underset{\textit{oxygen}}{O_2\,(gas)} \rightarrow \underset{\textit{silicon dioxide}}{SiO_2\,(solid)} + \underset{\textit{phosphorus}}{2P\,(solid)} + \underset{\textit{hydrogen}}{5H_2\,(gas)}$$

Within the glass, the phosphorous is in the form of a phosphorus pentoxide (P_2O_5), making the glass a dual compound as it codeposits with silicon dioxide. The P_2O_5 content in films is limited to about 4% by weight for films that are left permanently on the wafer because PSG becomes increasingly hygroscopic (water-absorbing).

PSG can be deposited at 600 to 650°C using high-density plasma CVD (HDPCVD) and has recently become popular for the first interlayer dielectric (ILD-1) layer because of its low deposition temperature, relatively planar surface, and excellent gap-fill characteristics. Incorporation of P_2O_5 into silicon dioxide films leads to a reduction in the film stress, which improves the film integrity. The dopant increases the moisture-barrier property of glass. Another benefit is that PSG layers are effective in serving as a getter to immobilize ionic contaminants. The ions become attached to the phosphorus and are thus restrained from diffusing through the film to the wafer surface.

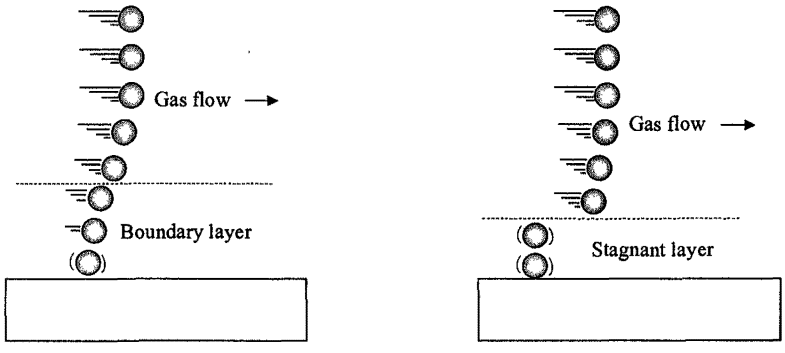


FIGURE 11.10 Gas-Flow Dynamics at the Wafer Surface

Borosilicate Glass. If diborane (B_2H_6) is used instead of phosphine (PH_3), then the result is *borosilicate glass (BSG)*. BSG requires reflow at a high temperature (e.g., 1000°C) to smooth out sharp steps in the wafer surface and densify the film. However, a high-temperature reflow is undesirable for the wafer thermal budget. BSG also has the disadvantage of not being a good barrier to mobile ionic contaminants (MICs).[21]

Borophosphosilicate Glass. Another method for doping oxide is to incorporate about 2 to 6 % weight each of boric oxide (B_2O_3) and phosphorus pentoxide (P_2O_5) into the silicon dioxide deposition to form *borophosphosilicate glass (BPSG)*. To achieve a dense oxide with good step coverage after deposition, the oxide is reflowed at a high temperature until it softens and flows. For BPSG, the reflow temperature is 800 to 1000°C for one hour. Reflow also improves its gettering characteristics to block mobile ionic contaminants (MICs). BPSG was commonly deposited on the wafer as the ILD-1 but has been recently replaced by PSG deposited by HDPCVD to reduce the thermal budget.[22]

High temperatures adversely affect the wafer thermal budget in sub-0.25 μm devices. For example, refractory metals (e.g., titaniun and cobalt) used to reduce contact resistance at junctions undergo material transformations at about 600°C that increase their contact resistance to unacceptable levels. In addition, the oxide planarity achieved by a high-temperature reflow is not sufficient for reduced depth of field requirements in ULSI photolithography. For this reason, reflow of doped oxide for planarity and gap fill has been largely replaced by high-density plasma CVD to achieve a stable film with good gap fill, followed by chemical mechanical planarization (CMP) to attain a flat surface (see Chapter 18).

Fluorosilicate Glass. *Fluorosilicate glass,* or *FSG,* is a fluorinated oxide that is being investigated as a first generation low-*k* dielectric for ILD deposition on 0.18 μm generation devices. By adding fluorine to silicon dioxide, there is a decrease in the dielectric constant of the material from about 3.9 for SiO_2 to about 3.5 for FSG.

To form an FSG film, silicon tetrafluoride (SiF_4) is added to the silane (SiH_4) and O_2 gases during the CVD reaction. A potential problem with using FSG is that the fluorine bonds are not always stable and can lead to corrosion, limiting the fluorine content to about 6%. If fluorine ions encounter moisture, hydrofluoric acid (HF), which etches oxide, is created. A method to avoid this problem is to remove weakly bonded fluorine atoms through the introduction of hydrogen during the deposition reaction.[23]

Oxide Versus Silicon Doping. Note that the doping of silicon dioxide is not the same as doping of silicon. When silicon is doped, there is a donation or acceptance of electrons between the dopant and silicon, within a single crystal lattice structure. Deposited oxide is an amorphous crystal structure and the dopants do not act as donors or acceptors of electrons. Doping is done to modify a physical characteristic of SiO_2, such as improving its ability to getter mobile ions.

CVD DEPOSITION SYSTEMS

There are different types of production systems for the deposition of dielectric and metal film layers. Some systems use large batch processing to perform CVD deposition on a large quantity of wafers at one time. However, the most advanced deposition systems currently in use for wafer manufacturing employ single-wafer processing in a cluster tool approach.

Table A.3 in Appendix A summarizes different gases used for CVD reactions. An example of a CVD gas is pure silane, which is used to deposit Si or SiO_2 in CVD systems. Silane reacts with air to form solid particles, causing particle contamination with the silane gas lines and leading to contaminated pipes or combustion. The gas line piping must be carefully installed to avoid any leaks. Many of the gases used in CVD film deposition are toxic. These gases are classified in Table A.3 based on four different safety hazards: pyrophoric (ignites spontaneously in air at or below 54.5°C), poisonous, corrosive, and dangerous combinations of gases, such as silane with hydrogen.

CVD Equipment Design

CVD has a range of equipment reactor designs, with each producing slightly different types of film quality. CVD reactors are broadly categorized based on the reaction chamber pressure regime used during operation: atmospheric-pressure CVD (APCVD) reactors and reduced-pressure reactors. The reduced-pressure CVD reactors have two general types. First, there are low-pressure CVD (LPCVD) reactors where the energy input is thermal. Second, there are plasma-assisted CVD reactors, either plasma-enhanced CVD (PECVD) or high-density plasma CVD (HDPCVD), where the energy is partially supplied by a plasma as well as thermally. Figure 11.11 depicts the different CVD reactor types.

CVD Reactor Types	Atmospheric	Low-pressure	Batch	Single-wafer
Hot-wall	✓	✓	✓	
Cold-wall	✓	✓	✓	✓
Continuous motion	✓		✓	
Epitaxial	✓		✓	
Plenum	✓		✓	
Nozzle	✓		✓	
Barrel	✓		✓	
Cold-wall planar		✓	✓	✓
Plasma-assisted		✓	✓	✓
Vertical-flow Isothermal		✓	✓	✓

FIGURE 11.11 CVD Reactor Types

CVD Reactor Heating ■ A major distinction in CVD reactors is whether they are hot-wall or cold-wall reactors. A *hot-wall reactor* employs a heating method that heats not only the wafer but also the wafer holder (referred to as a *susceptor*) and the walls of the reactor. Hot-wall reactors have film-forming reactions on both the reactor chamber walls and substrate, requiring frequent cleaning or in situ cleaning to minimize particle contamination. A resistance heater surrounding the reactor tube is an example of a hot-wall reactor. A *cold-wall reactor* heats only the wafers and susceptor, which is where film is deposited. The walls of the reactor are cold and therefore do not have sufficient energy to allow a deposition reaction to occur. RF induction heating or infrared lamps mounted in the reactor are examples of this type of reactor. Localized heating at the wafer minimizes particle formation in the reactor.

CVD Reactor Configuration ■ The geometry of the reactor design depends closely on the pressure attained during the deposition process. Atmospheric-pressure reactors operate in the mass-transport limited region, so they must be designed so that an equal amount of reactant gases are delivered to each wafer. To achieve this condition, wafers are usually laid flat on a horizontal surface. A drawback to this approach is that wafers are susceptible to contamination from falling particles.

Low-pressure reactors (LPCVD) are reaction-rate limited. This means that they can be vertically stacked at close spacing with a large number of wafers per run, since adequate reactant gas is always supplied to the wafer surface. Quartz wafer holders (boats) hold up to 200 wafers. At the same time, LPCVD reactors must have precise temperature control because of their limitation from the reaction rate.

CVD Reactor Summary ■ The different types of CVD processes and their principal characteristics are shown in Table 11.2.[24]

TABLE 11.2 Types of CVD Reactors and Principal Characteristics

Process	Advantages	Disadvantages	Applications
APCVD (Atmospheric Pressure CVD)	Simple reactor, fast deposition, and low temperature.	Poor step coverage, particle contamination, and low throughput.	Low-temperature oxides (both doped and undoped).
LPCVD (Low Pressure CVD)	Excellent purity and uniformity, conformal step coverage, and large wafer capacity.	High temperature, low deposition rate, more maintenance intensive, and requires vacuum system.	High-temperature oxides (both doped and undoped), silicon nitride, polysilicon, W, and WSi_2.
Plasma-Assisted CVD: • Plasma-Enhanced CVD (PECVD) • High-Density Plasma CVD (HDPCVD)	Low temperature, fast deposition, good step coverage, and good gap fill.	Requires RF system, higher cost, stress is much higher with a tensile component, and chemical (e.g., H_2) and particle contamination.	High aspect ratio gap fill, low-temperature oxides over metals, ILD-1, ILD, copper seed layer for dual damascene, and passivation (nitride).

APCVD (Atmospheric Pressure CVD)

The first type of chemical vapor deposition used in the semiconductor industry is *atmospheric pressure CVD (APCVD)*.[25] As previously discussed, APCVD generally operates in the mass-transport limited regime. At any given time, there may not be sufficient gas molecules present at the wafer surface for a reaction to occur. Therefore, the reactor must be designed to have optimum reactant gas flow to every wafer in the system. The basic system design should never permit the reaction to slow down because insufficient reactant gas is available. Since the reactor is at atmospheric pressure, the reactor design can be simple and allows for high deposition rates.

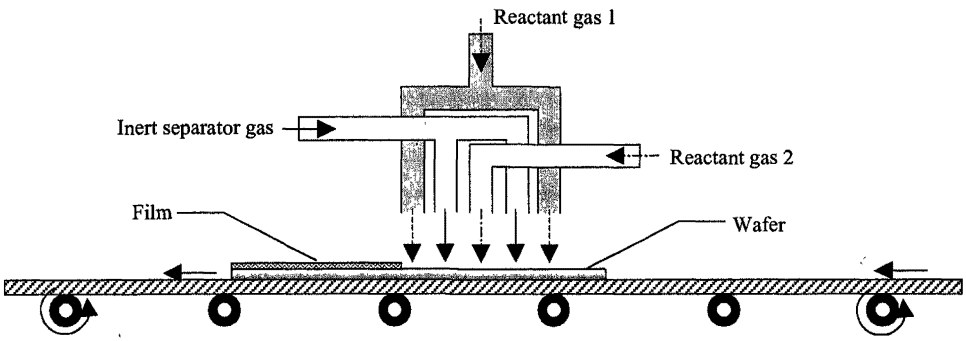
Two different types of continuous-processing APCVD systems are shown in Figure 11.12 on page 272. These equipment designs use a belt or conveyor to carry the wafer samples through the reactor gases, which flow through the center of the reactor.

Continuous-processing APCVD systems have high equipment throughput, good uniformity, and the capability to process large-diameter wafers. APCVD systems do have problems with high gas consumption and often need frequent reactor cleaning. Since film deposition also takes place on the conveyor as well as the wafer, the belt transport system must be cleaned (sometimes it is in cleaned in situ, or during use). APCVD deposited films often exhibit poor step coverage.

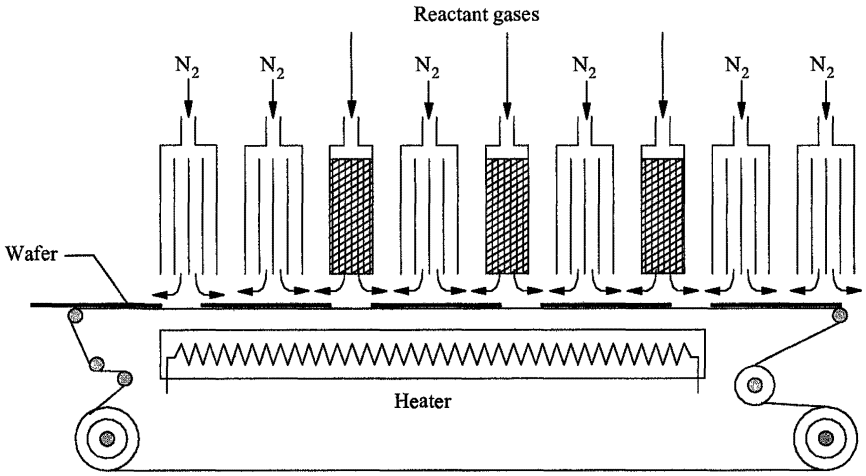
Silicon Dioxide ■ The most common use of APCVD is in the deposition of silicon dioxide (SiO_2) and doped oxides (e.g., PSG, BPSG, and FSG, as discussed earlier in this chapter). These films have been used traditionally as an interlayer dielectric (ILD), as a protective overcoat, or to planarize (smoothen) a nonuniform surface.

SiO_2 Deposition With Silane. SiO_2 is deposited by oxidizing *silane* (SiH_4) with oxygen. Pure silane is a highly pyrophoric, unstable gas that burns on exposure to air. For this reason, it is commonly supplied in a low dilution (typically 2 to 10% by volume) in argon or nitrogen in order to make it safer to handle. This reaction can be done at a low temperature of 450 to 500°C, which is an advantage for depositing SiO_2 on the aluminum metal lines as an ILD.[26] However, due to the small mean free path and poor surface migration through the boundary layer, this method has poor step coverage and poor gap-fill characteristics. This makes it unacceptable for critical ULSI applications.

SiO_2 Deposition With TEOS-Ozone. A common APCVD application is to deposit oxide by reacting *TEOS*, which is tetraethylorthosilicate ($Si(C_2H_5O)_4$) or tetraethoxysilane, with ozone (O_3). TEOS is an organic liquid precursor. It usually uses a carrier gas, typically nitrogen, bubbled through it for delivery of the TEOS gas mixture to the reactor. There are also flow controllers for delivering the TEOS source in liquid state to the reactor. Ozone is a triatomic oxygen molecule that is much more reactive than oxygen; therefore, the process can be done at low temperatures (e.g., 400°C) without plasma. Because the decomposition of TEOS is induced by O_3 without a plasma, the reaction takes place at atmospheric pressure (APCVD) of 760 torr or subatmospheric pressure (SACVD) at about 600 torr. There appears to be no significant difference between APCVD and



(a) Gas-injection type



(b) Plenum type

FIGURE 11.12 Continuous-Processing APCVD Reactors

SACVD.[27] The deposited oxide film has improved profiles over steps, is conformal, and has excellent electrical characteristics as a dielectric.[28] The reaction equation between TEOS and O_3 to obtain an oxide is:

$$\underset{\textit{TEOS}}{Si(C_2H_5O)_4} + \underset{\textit{ozone}}{8O_3} \rightarrow \underset{\textit{silicon dioxide}}{SiO_2} + \underset{\textit{water}}{10H_2O} + \underset{\textit{carbon dioxide}}{8\,CO_2}$$

An APCVD TEOS-O_3 as-deposited SiO_2 film is porous and often requires a reflow step to remove moisture as well as densify the film. This reflow adds a process step and reduces the thermal budget.[29] Doping uniformity is also a concern, as this can lead to problems for nonuniform removal rates during chemical mechanical planarization. The primary advantage of APCVD TEOS-0_3 is the improved step coverage for high-aspect ratio gaps for applications such as shallow trench isolation (see Figure 11.13). Another advantage to APCVD TEOS-O_3 is that it is just a basic thermal CVD process to deposit oxide, which could avoid wafer surface damage and wafer corner damage at steps. This type of damage is possible from high-density plasma CVD (discussed later in this chapter). An aspect of TEOS-O_3 films is that they are usually deposited in combination with other oxide film layers (e.g., PECVD oxide) for reasons such as reducing the TEOS-O_3 tensile stress on a thick film or to reduce sensitivity with the underlying material.[30]

Doped Oxides. APCVD oxides are usually doped with chemicals such as phosphorus and boron (discussed earlier in this chapter). If SiO_2 is undoped, it is referred to as *undoped silicate glass* (*USG*) or *undoped oxide* (*UDOX*). Phosphorus doped oxide is phosphosilicate glass (PSG). In the traditional deposition process, a high-temperature reflow (as high as 950°C for 15 to 30 minutes) heats and softens the doped oxide, causing it to planarize the surface (see Figure 11.14).

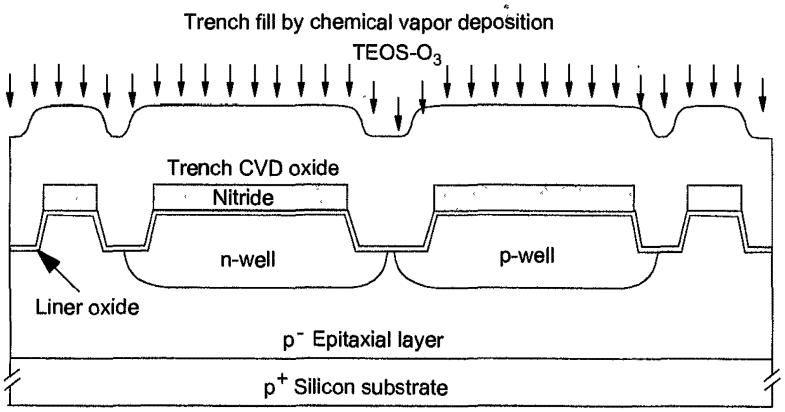


FIGURE 11.13 Improved Step Coverage of APCVD TEOS-O_3

Silane-based borophosphosilicate, or BPSG films, are deposited at low temperatures (400 to 450°C) and then immediately densified and stabilized at about 800°C for one hour. A smoother surface makes it easier to deposit and pattern the next film. Note that for the ULSI device technology, reflow has been replaced with chemical mechanical planarization (CMP) to planarize the surface. CMP is discussed in Chapter 18.

LPCVD (Low Pressure CVD)

Low pressure CVD (LPCVD) systems are more common now than APCVD because of their lower cost, higher production throughput, and superior film properties. LPCVD operates at a medium vacuum (about 0.1 to 5 torr), and employs temperatures between 300 and 900°C.[31] Conventional oxidation type furnaces (horizontal or vertical) and multichamber cluster tools can be used for LPCVD processing.

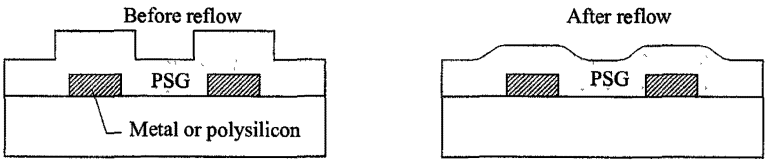


FIGURE 11.14 Planarized Surface after Reflow of PSG

LPCVD reactors typically operate in the reaction-rate limited regime. In this reduced-pressure regime, the diffusivity of the reactant gas molecules increases so that the mass-transfer of the gas to the wafer no longer limits the rate of the reaction. Because of this transfer state, the gas-flow conditions inside the reactor are not important, permitting the reactor design to be optimized for high wafer capacity (e.g., wafers can be closely spaced). Films are uniformly deposited on a large number of wafer surfaces as long as the temperature is tightly controlled.

The boundary layer in a LPCVD reactor is different from APCVD because it extends farther from the wafer surface due to the low-pressure condition (see Figure 11.15 on page 274). The boundary layer has a low molecular density, which permits incoming gas molecules to easily diffuse through this layer, exposing the wafer surface to more molecules than it can possibly consume. This condition explains why an LPCVD reactor operates in the reaction-rate limited regime. It is the reaction rate that is the limiting factor in the deposition rate, not the reactant supply. Furthermore, there are many collisions during LPCVD, so that material strikes the wafer in an undirected manner. This action is beneficial for conformal film coverage on high-aspect ratio steps and trenches. As a result, step coverage is usually good with LPCVD.[32]

The LPCVD reactor design favors the hot-wall reactor type so that uniform temperature control is achievable over a large operating length. A representative reaction chamber is shown in Figure 11.16 on page 274.

Since LPCVD reactors are often hot wall, particles deposit on the reactor wall. These deposits are minimized by reducing the partial pressure of the gas-phase reactants, which leads to fewer

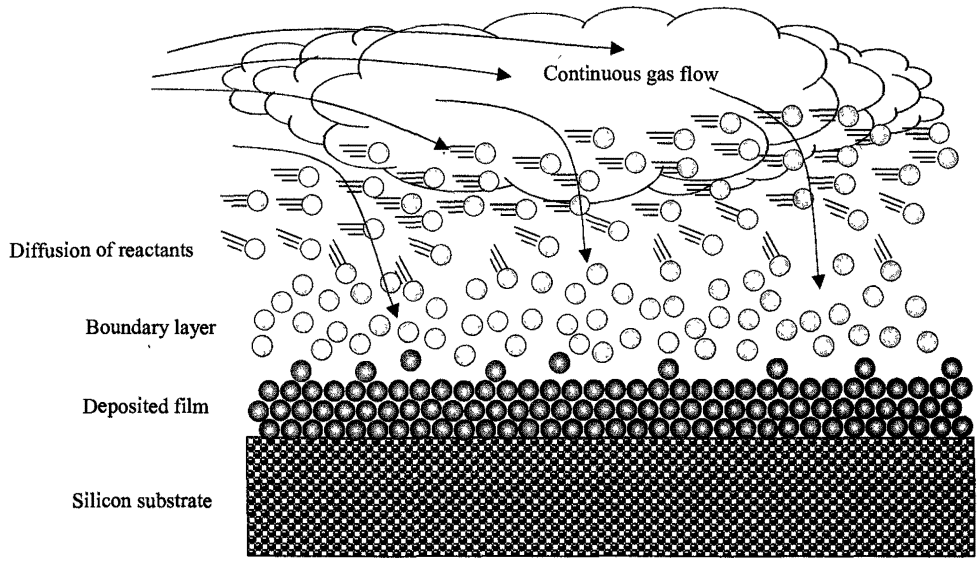


FIGURE 11.15 Boundary Layer at Wafer Surface

gas-phase reactions. For this reason, LPCVD reactors are usually operated at lower growth rates than APCVD systems.[33] Hot-wall reactors require periodic reaction chamber cleanup of particles by routine maintenance.

The traditional method of tube cleaning was to remove a dirty quartz tube, install a previously cleaned tube, and clean the dirty tube for future use. Cleaning involved manually rolling the tubes half-submerged in an acid bath such as aqueous HF. For production and safety reasons, an in situ clean is desirable. One in situ method is the use of plasma-generated fluorine gases that react with the solid residue in the reactor, forming volatile reaction products that are pumped from the system. Another in situ approach is to use chlorine trifluoride (ClF_3) with elevated temperature to activate a thermal clean. In situ tube cleaning for LPCVD reduces equipment downtime, lowers particle counts, and minimizes personnel exposure to chemicals.[34]

LPCVD reactors with a large number of wafers (e.g., 150 to 200 wafers) experience reactant depletion as gases are transported along the reaction chamber. This depletion leads to a reduction in growth rate. Adjusting the reactor temperature so that it increases slightly from inlet to outlet (typically by 25 to 50°C) compensates for this temperature nonuniformity.

Silicon Dioxide ■ There are numerous applications for LPCVD oxides (doped and undoped) in ULSI multilevel metallization. LPCVD oxides are used for interlayer dielectric (ILD), shallow trench isolation oxide fill, and sidewall spacers.

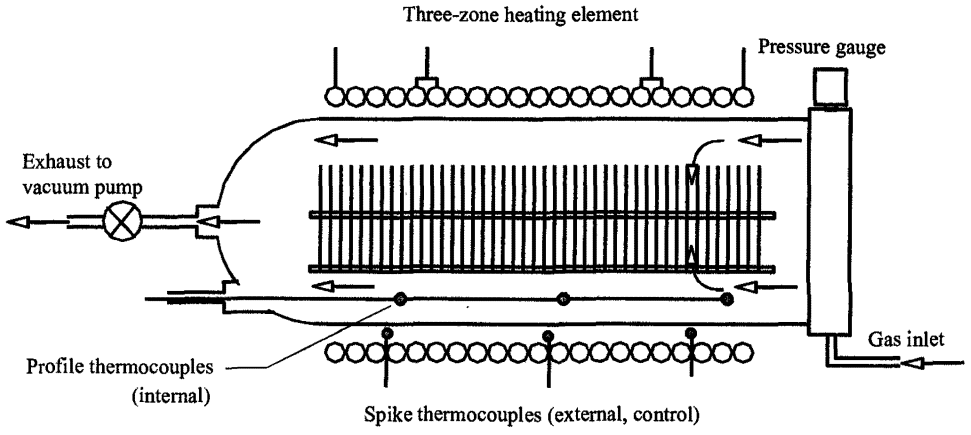


FIGURE 11.16 LPCVD Reaction Chamber

SiO_2 With TEOS. A common LPCVD method for depositing silicon dioxide is the pyrolosis (decomposition) of TEOS, with or without oxygen, at low pressure and at a temperature between 650 to 750°C. This method is sometimes referred to as LPTEOS, for low-pressure TEOS. LPTEOS yields very good oxide conformality due to rapid surface diffusion of gas molecules. A liquid TEOS source is used by bubbling a carrier gas through it (e.g., N_2, O_2, or He). The liquid source is heated by its own independent temperature source. The concentration of the liquid source vapor entering the reactor is controlled by the carrier gas-flow rate and liquid source temperature. A schematic of a typical TEOS deposition system is shown in Figure 11.17. The film growth rate for LPCVD oxides is significantly slower (100-150 Å/minute) than for APCVD films.[35]

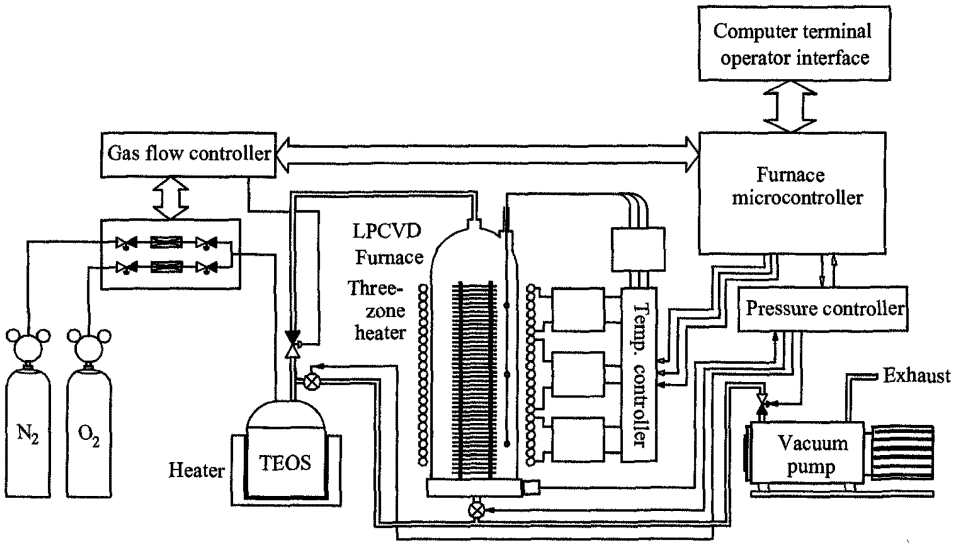


FIGURE 11.17 Oxide Deposition with TEOS LPCVD

SiO_2 With SiH_4. SiO_2 can also be deposited by LPCVD by oxidizing silane (SiH_4) at a relatively low temperature of about 450°C. Similar to APCVD oxide with SiH_4, this process has poor step coverage. A higher temperature deposition is done with dichlorosilane (SiH_2Cl_2 or DCS) and nitrous oxide (N_2O) at about 900°C to produce a good quality SiO_2 film. However, the high temperature is undesirable for the wafer thermal budget.

Silicon Nitride ■ *Silicon nitride (Si_3N_4)* is a film often used as the final wafer passivation layer because it provides good protection against the diffusion of impurities and moisture. It achieves excellent step coverage with high conformal coverage with LPCVD deposition. Silicon nitride is also used as a mask material (referred to as a hard mask), such as in the shallow trench isolation (STI) process. Silicon nitride has a high dielectric constant (i.e., a k value of 6.9, versus about 3.9 for CVD SiO_2), which makes it undesirable for the ILD dielectric due to the resultant high capacitance between conductor layers.

Reacting dichlorosilane ($SiCl_2H_2$) and ammonia (NH_3) in an LPCVD reactor produces silicon nitride at a reduced pressure and temperature between 700 and 800°C. The chemical reaction is:

$$\underset{\textit{dichlorosilane}}{3SiCl_2H_2\,(gas)} + \underset{\textit{ammonia}}{4NH_3(gas)} \rightarrow \underset{\textit{silicon nitride}}{Si_3N_4(solid)} + \underset{\textit{hydrogen chloride}}{6HCl\,(gas)} + \underset{\textit{hydrogen}}{6H_2(gas)}$$

The important variables that affect the properties of silicon nitride in an LPCVD process are total pressure, reactant concentrations, deposition temperature, and temperature gradients. For example, increasing the total pressure and partial pressure of dichlorosilane in the reaction chamber will increase the deposition rate.

Silicon nitride can also be produced in an APCVD process by reacting silane and ammonia, but better film uniformity and higher wafer throughput is achieved in the low-pressure process.[36] Deposited silicon nitride is an amorphous film that often contains large amounts of hydrogen. A high ammonia ratio and low deposition temperature will increase the hydrogen content.

Polysilicon ■ Polysilicon film (also referred to as *poly-si,* or *poly*) is commonly deposited using LPCVD. Recall that the term polysilicon means that the silicon is a polycrystal, with many small single-crystal regions separated by grain boundaries. The stress in thin polysilicon films is compressive.

Doped polysilicon serves as the gate electrode in MOS devices (see Figure 11.18). Key reasons for the use of doped polysilicon in the gate structure are:[37]

1. Ability to be doped to a specific resistivity.
2. Excellent interface characteristics with silicon dioxide.
3. Compatibility with subsequent high-temperature processing.
4. Higher reliability than possible metal electrodes (e.g., aluminum).
5. Ability to be deposited conformally over steep topography.
6. Allows for self-aligned gate process (see Chapter 12).

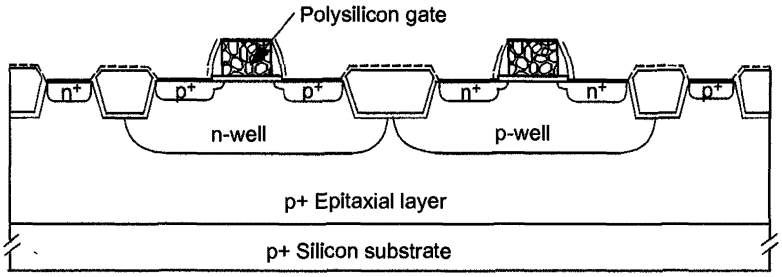


FIGURE 11.18 Doped Polysilicon as a Gate Electrode

Polysilicon is deposited in LPCVD by the pyrolysis (thermal decomposition) of silane at a temperature of 575 to 650°C. Low pressure reactors use either pure silane or 20 to 30% silane diluted with nitrogen, which is fed into the system at a pressure of 0.2 to 1.0 torr to fabricate polysilicon films. A practical deposition rate is about 100 to 200 Å/minute. The addition of diborane can enhance the polysilicon deposition rate because the diborane forms borane radicals, BH_3, which catalyze gas-phase reactions. The chemical reaction is:

SiH_4 (gas)	+	He (gas)	→	Si (solid)	+	$2H_2$ (gas)
silane		*helium (carrier gas)*		*silicon*		*hydrogen*

The polysilicon film is usually doped in situ during the deposition process by adding arsine (AH_3), phosphine (PH_3), or diborane (B_2H_6) to the gas mixture. An inert gas is often accompanied to improve film uniformity. Doping can be done after deposition by ion implantation (see Chapter 17). The resistivity of the polysilicon after doping is highly dependent on parameters such as deposition temperature, dopant concentration, and annealing temperature effects on grain size.

A refractory metal is deposited over the polysilicon gate to form a *polycide* or *silicide* structure (see Chapter 12). This action is taken to reduce electrical resistivity at the polysilicon interface with the other material (e.g., tungsten) that forms the electrical connection to other device components. Lightly doped polysilicon also serves as a resistor in memory cells, capacitors, and thin-film transistors.

Silicon Oxynitride ■ Silicon nitride films that contain oxygen are referred to as *silicon oxynitride* (SiO_xN_y) and combine the advantages of oxides and nitrides. Oxynitride films also have improved thermal stability, cracking resistance, and reduced film stress when compared to silicon nitride. In general, increased oxygen in oxynitride film also decreases the film's refractive index, which makes it less reflective and a useful antireflective layer in photolithography (see Chapter 14).[38] Another benefit of a silicon oxynitride film is the nitrogen in the film accumulates at the silicon interface to reduce the concentration of strained Si-O bonds, potentially reducing the creation of hot electrons (undesirable charge carriers) by as much as three orders of magnitude. An oxynitride layer at the Si/SiO_2 interface of thin gate oxides has been found to be beneficial for improving the device electrical performance.

Oxynitride films are formed by different techniques, including the oxidation of silicon nitride, the nitridation of SiO_2 with ammonia (NH_3), and the direct growth of SiO_xN_y using nitrous oxide as the oxidant species. One way of forming oxynitride films is by reacting silane (SiH_4) with nitrogen dioxide (N_2O) and NH_3. Nitride or oxynitride films are used frequently as an "etch stop" for selective etching of dielectric films.[39]

Plasma-Assisted CVD

The third major type of CVD equipment relies on plasma energy, in addition to thermal energy, to initiate and sustain the chemical reactions necessary for CVD. The advantages of using plasma during CVD are:

1. Lower processing temperature (250 to 450°C).
2. Excellent gap-fill for high aspect ratio gaps (with high-density plasma).
3. Good film adhesion to the wafer.
4. High deposition rates.
5. High film density due to low pinholes and voids.
6. Wide range of applications due to lower processing temperature.

Film Formation ■ The plasma-assisted CVD reaction necessary to form a film occurs when RF power is used to break up gas molecules in a vacuum. The frequency of the RF power depends on the application, with typical frequencies found at 40 kHz, 400 kHz, 13.56 MHz, and 2.45 GHz (microwave power). The molecular fragments (radicals) are chemically reactive species and readily bond to other atoms to form a film at the wafer surface (see Figure 11.19). Gaseous by-products are removed by the vacuum pumping system. The wafer is usually heated in order to assist the surface reactions and reduce the level of undesirable contaminants, such as hydrogen.

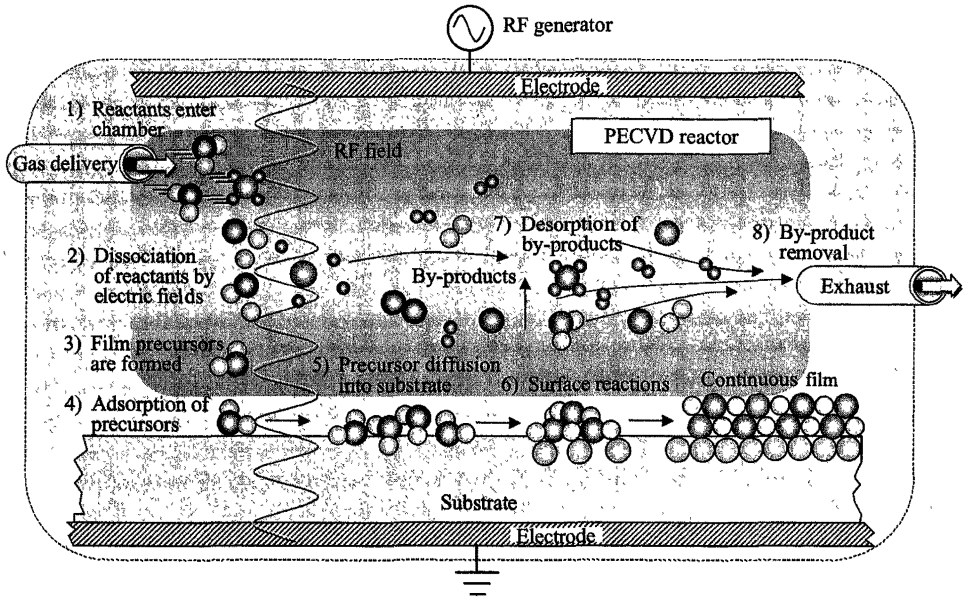


FIGURE 11.19 Film Formation During Plasma-Assisted CVD

The plasma-assisted CVD reactions at the wafer surface are very complex. The properties of plasma-deposited films depend on many variables, such as electrode configuration and separation, power level and frequency, gas composition, pressure and flow rate, and substrate temperature.[40] The essential aspect of these variables for production is embodied in the chamber design. The specific details of the intermediate reactions that occur in plasma glow discharge are not well understood. For these reasons, deposited films are often found to be not very stoichiometric. This means

that the deposited film will not necessarily have the same ratio of elements found in its nominal chemical formula.

There are two types of plasma processes used in CVD:

- Plasma-enhanced CVD (PECVD)
- High-density plasma CVD (HDPCVD)

Plasma-Enhanced CVD (PECVD) ■ The development of *plasma-enhanced CVD (PECVD)* uses plasma energy to create and sustain the CVD reaction. It is a natural follow-on to LPCVD since the system pressure for both types of CVD processes is comparable. The important difference is the much lower PECVD deposition temperature. For instance, silicon nitride (Si_3N_4) is deposited using LPCVD at 800 to 900°C, yet cannot be deposited over aluminum metallization because Al melts at 660°C. On the other hand, silicon nitride deposited with PECVD at a temperature of 350°C is suitable for this application.

PECVD is performed in a vacuum chamber between parallel conducting plates positioned several inches apart, typically with a variable gap for process optimization. A modern reactor is a multichamber cluster tool. The wafer (or wafers) is placed on the grounded bottom plate and RF power is applied to the top electrode. A plasma develops when the source gas is introduced through the gas manifold and the center of the electrodes. Exhaust gases are pumped out from the periphery of the bottom electrode. Sometimes the reactant gases are introduced at the periphery and pumped out at the center of the bottom electrode. A schematic of an overall PECVD system is shown in Figure 11.20.

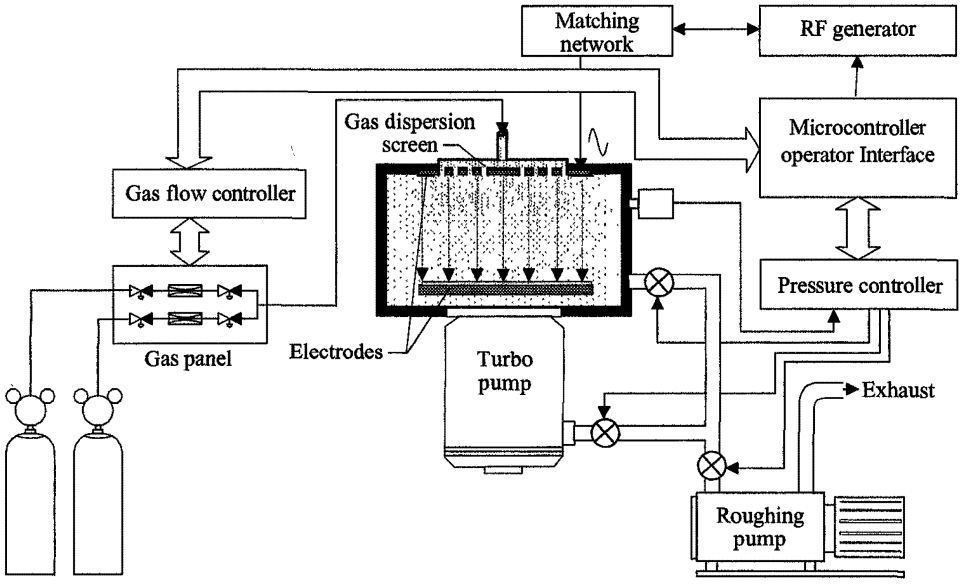


FIGURE 11.20 General Schematic of PECVD

A PECVD reactor is typically a cold-wall plasma reactor, with the wafer heated in its chuck while the remaining parts of the reactor are unheated. Deposition parameters must be controlled to ensure the temperature gradient does not affect the film thickness uniformity. Cold-wall reactors create fewer particles and require less downtime for cleaning. Deposition chambers often have in situ cleaning to reduce particles. With parallel-plate processing, there is the possibility that particles formed in the gas phase will fall on the wafer surface.

Silicon Dioxide PECVD. Oxide films with PECVD are usually formed by reacting silane (SiH_4) with either oxygen (O_2), nitrous oxide (N_2O), or carbon dioxide (CO_2) in a plasma. The processing temperature is about 350°C.[41] The oxide can be doped with boron or phosphorus to form BSG or PSG, respectively, or with both to deposit BPSG. PECVD PSG tends to be more crack resistant, conformal, and free of pinholes than APCVD PSG.

A plasma gas mixture of silane and oxygen is not normally used because oxygen reacts readily in the gas phase, which promotes particle generation and poor film quality (e.g., pinholes). The silane and nitrous oxide gas mixture produces more uniform films than through the use of an oxygen reactant. The reaction of silane with nitrous oxide to form silicon dioxide is:

$$\underset{\textit{silane}}{SiH_4\ (gas)} + \underset{\textit{nitrous oxide}}{2N_2O\ (gas)} \rightarrow \underset{\textit{silicon dioxide}}{SiO_2\ (solid)} + \underset{\textit{nitrogen}}{2N_2\ (gas)} + \underset{\textit{hydrogen}}{2H_2\ (gas)}$$

The silicon dioxide is nearly stoichiometric, although it does contain some hydrogen, with a small amount of nitrogen present in the film. Hydrogen can exist as a hydride (Si-H), silanol (Si-O-H), or as water (H-O-H). The presence of O-H groups is undesirable for MOS transistor electrical characteristics and must be minimized.[42]

Parallel-plate PECVD can also deposit SiO_2 using TEOS, and is referred to as *PETEOS*. However, PETEOS is not acceptable for filling narrow-spaced metal lines due to voids. PETEOS is combined with APCVD TEOS or high-density plasma CVD (HDPCVD) to attain conformal gap fill, followed by chemical mechanical planarization. PETEOS has a relatively high SiO_2 deposition rate, which makes it desirable for wafer throughput in integrated tools.

Silicon Nitride PECVD. *Silicon nitride PECVD* is a film used as a final passivation layer on the chip as scratch protection, and a moisture barrier and as a barrier to prevent sodium diffusion. It is also widely used as a mask material in shallow trench isolation (STI) and in self-aligned contact technology (see Chapter 12). The Si_3N_4 nitride film deposited by PECVD is not very stoichiometric, and is sometimes written as $Si_xN_yH_z$. This equation highlights its nonstoichiometric composition and identifies that the film contains hydrogen (usually 9 to 30%). Table 11.3 compares the properties of silicon nitride for LPCVD vs. PECVD.

TABLE 11.3 Properties of Silicon Nitride for LPCVD vs. PECVD

Property	LPCVD	PECVD
Deposition temperature (°C)	700 to 800	300 to 400
Composition	Si_3N_4	$Si_xN_yH_z$
Step coverage	Fair	Conformal
Stress at 23°C on silicon (dynes/cm^2)	$1.2 - 1.8 \times 10^{10}$ (tensile)	$1 - 8 \times 10^9$ (tensile or compressive)

PECVD nitride films are commonly produced from silane (SiH_4) and reacted with ammonia or nitrogen. The reaction equations are:

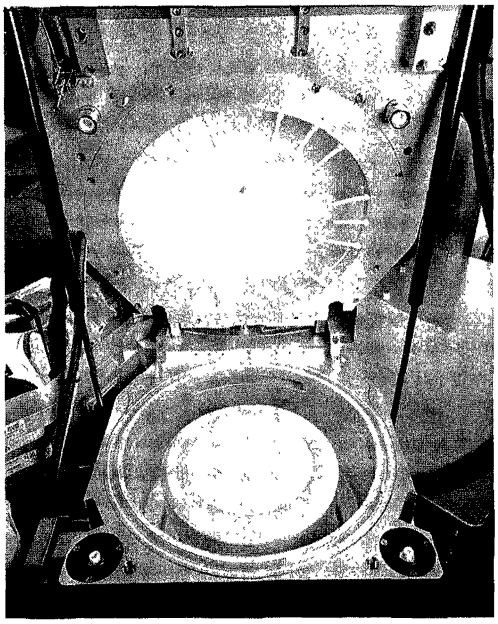
$$\underset{\textit{silane}}{SiH_4\ (gas)} + \underset{\textit{ammonia}}{NH_3\ (gas)} \rightarrow \underset{\textit{silicon nitride}}{Si_xN_yH_z\ (solid)} + \underset{\textit{hydrogen}}{3H_2\ (gas)}$$

$$\underset{\textit{silane}}{2SiH_4\ (gas)} + \underset{\textit{nitrogen}}{N_2\ (gas)} \rightarrow \underset{\textit{silicon nitride}}{2Si_xN_yH_z\ (solid)} + \underset{\textit{hydrogen}}{3H_2\ (gas)}$$

PECVD silicon nitride can have increased compressive film stress caused by ion bombardment during deposition that ruptures the Si-N or Si-H bonds. High compressive stress in a nitride layer can cause voids and cracks in the underlying aluminum metallization. There is usually substantial hydrogen content in the film. Hydrogen can reduce film stress, but it also degrades film properties.[43] When nitrogen is substituted for ammonia, the hydrogen content is reduced. However, ionization of nitrogen is difficult for plasma formation.

Silicon Oxynitride. Films of silicon oxynitride are deposited using PECVD by reacting Si_3N_4 with nitrous oxide (N_2O) at a temperature of 200 to 250°C.[44] Silicon oxynitride combines some properties from nitrides and oxides. It has good resistance to moisture and sodium penetration, with excellent mechanical, chemical, and electrical properties. These properties make silicon oxynitride suitable for topside passivation.

Gap Fill. A major limitation for PECVD in fine-geometry devices is gap fill. For gaps below 0.5 μm spacing (actual gap width), PECVD cannot fill the high aspect ratio gap without pinching off the top and leaving a void. A void in a dielectric material that is filling a gap is undesirable because of its effect on electrical performance and long-term reliability. For devices at the



High Density Plasma Deposition Chamber
(Photo courtesy of Applied Materials, Inc.)

0.25 μm technology node and below, HDPCVD has replaced PECVD because of its superior gap-fill properties.

It should be pointed out that other CVD methods can achieve fine-geometry gap-fill but have other limitations. For example, APCVD using TEOS oxide has good conformality and produces excellent gap fill. However, TEOS has different deposition characteristics depending on the underlying surface. One problem is that TEOS layers have dangling bonds that absorb moisture. To inhibit this condition, PECVD undercoat and overcoat layers are required. These layers add cost and also risk being removed during the subsequent chemical mechanical planarization (CMP).[45]

High-Density Plasma CVD (HDPCVD) ■ A recent development in plasma assisted CVD is *high-density plasma CVD (HDPCVD)*. This deposition method became widely accepted in advanced wafer fabs in the mid-1990s. As its name implies, the plasma in HDPCVD is a high-density mixture of gases at low pressure that is directed toward the wafer surface in the reaction chamber. Its main benefit is that it can deposit films to fill high aspect ratio gaps with a deposition temperature range of 300 to 400°C. HDPCVD was initially developed for interlayer dielectric (ILD) applications, but is also being used for deposition in ILD-1, shallow trench isolation, etch-stop layers, and deposition of low-k dielectrics.[46]

The HDPCVD reaction involves a chemical reaction between two or more gas precursors. For the deposition of oxide ILD, oxygen (or ozone) is often used with a silicon containing gas such as silane or TEOS, along with argon. To form the high-density plasma, a source excites the gas mixture with RF or microwave power (2.45 GHz) and directs the plasma ions into a dense region above the wafer surface. There are different high-density plasma sources, such as electron cyclotron resonance (ECR), inductively coupled plasma (ICP), and Helicon. These plasma sources are explained in detail in Chapter 16.

Wafer Bias and Heat Load. An RF bias (often between 1,500 to 3,000 kW of RF power) is applied to the wafer, pulling energetic ions out of the plasma and directing them toward the wafer surface. Biasing the wafer gives directionality to the energetic ions of the plasma. The high-density plasma ion density is about 10^{11} to 10^{12} ions/cm^3 at low pressures (between 2 to 10 mtorr). This high plasma density, in conjunction with directionality from wafer biasing, is an important reason for HDPCVD's ability to deposit films into narrow gaps with high-aspect ratio geometries of 3:1 to 4:1 and higher.

Much of the challenge for the use of high-density plasma is related not only to the performance of the plasma source but also to the details of the chamber design so that the technology works

in high-volume wafer fabrication.[47] A particular problem is that the high-density plasma will increase the thermal load to the wafer, since 2000W of RF bias can result in power density application of approximately 6 watts/cm^2 on the wafer surface. This condition leads to high wafer temperatures. However, the ILD must have process temperatures below 400°C to avoid harming the aluminum metallization; in addition, high heat loads cause thermal stress to the wafer.[48] This temperature limit requires cooling the wafer by applying a backside blanket of helium gas to the wafer through access ports in the electrostatic chuck (ESC). This action creates a thermal conductivity path between the wafer and the ESC, thus cooling the wafer and the chuck.

Simultaneous Deposition and Etching. An HDPCVD process uses a simultaneous deposition and etching action that is the basis for its ability to fill high-aspect ratio gaps with dielectric material without voids (see Figure 11.21). This is referred to as the *dep:etch (D:E) ratio,* which typically has a value of approximately 3:1 for HDPCVD.[49] Interpret this ratio as meaning the deposition rate (i.e., the rate material is deposited) is occurring three times faster than the etch rate. Increasing the ratio will increase deposition and therefore wafer throughput, but if the ratio is too high, then voids occur because the gap is not completely filled.

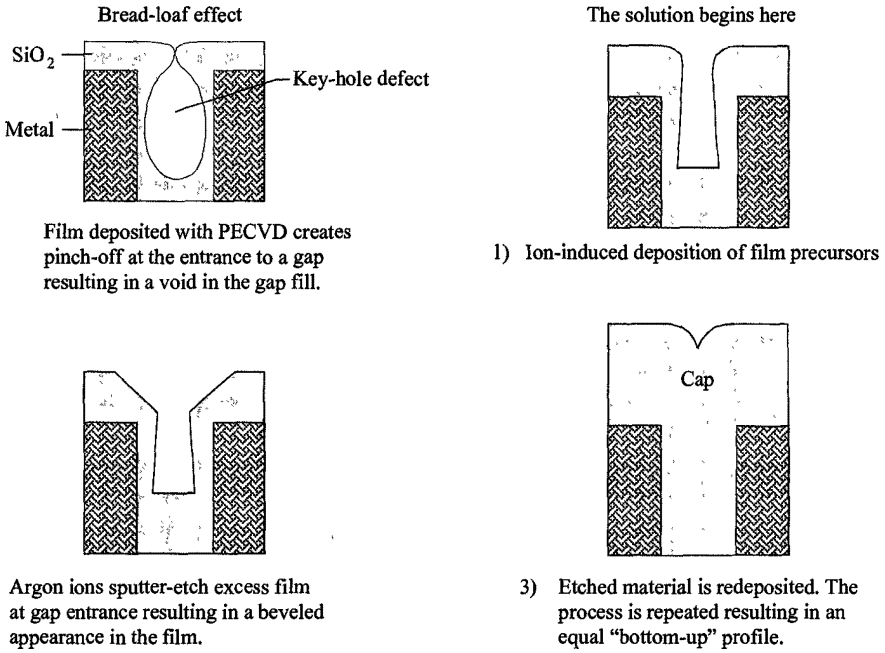


FIGURE 11.21 Dep-Etch-Dep Process

The surface reactions for the simultaneous deposition-etch include five steps, with the first three considered the dominant mechanisms:[50]

1. **Ion-induced deposition:** The ion is pulled from the plasma and deposited to produce the main gap-filling phenomena. It is driven by the kinetic energy of reactant ions breaking surface bonds to form reaction sites.
2. **Sputter etch:** Energetic argon and reactant ions attracted to the surface due to wafer bias will bombard the surface and etch (dislodge) atoms.
3. **Redeposition:** Atoms are dislodged from the bottom of the gap and usually redeposit on sidewalls. This is important for uniform thickness on sidewalls and the bottom of the gap.
4. **Hot-neutral CVD:** This is a minor contribution to the reaction where thermal energy drives some deposition.
5. **Reflection:** Another minor contribution of ions reflecting off sidewalls and then deposited.

Simultaneous deposition and etching is a beneficial by-product of the plasma directionality created in the oxygen and argon gas mixture in HDPCVD. For a silicon dioxide (SiO_2) deposition process, the oxygen reacts with silane (SiH_4) to form SiO_2, while argon acts to remove the deposited material away by sputtering (see Chapter 12 for an explanation of sputtering). Some factors that affect the dep:etch ratio are the ratio of the oxygen and argon gases, the chamber pressure, the ion energy, and the RF bias on the wafer.[51] Low pressure is important to reduce the mean free path, which reduces collisions and maintains good directionality of the plasma. For acceptable deposition rates and wafer throughput in HDPCVD, high gas-flow rates are required. As shown in Figure 11.22, the wafer often will sit directly on the throat of a high-speed turbo pump (e.g., 4,000 to 5,000 liters/sec pump speed for 300-mm diameter wafers).

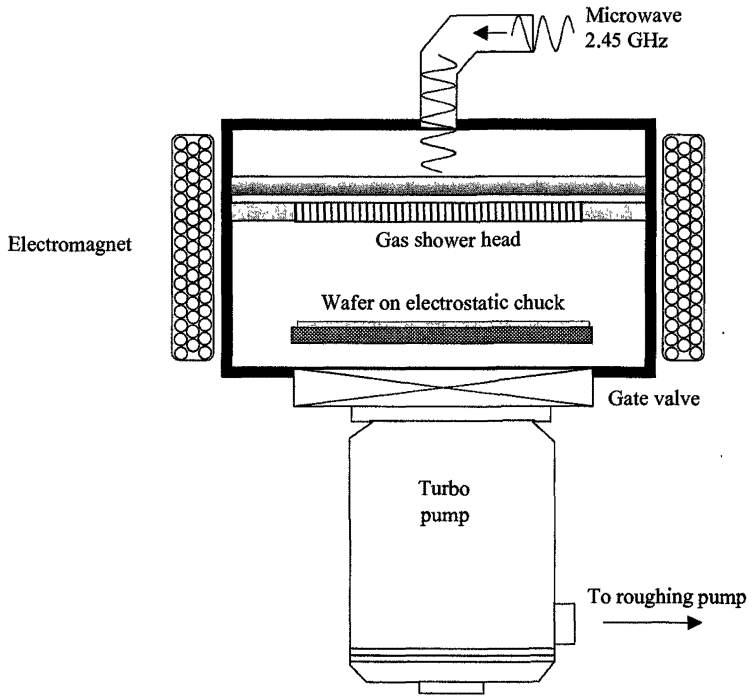


FIGURE 11.22 HDPCVD with Wafer at Throat of Turbo Pump

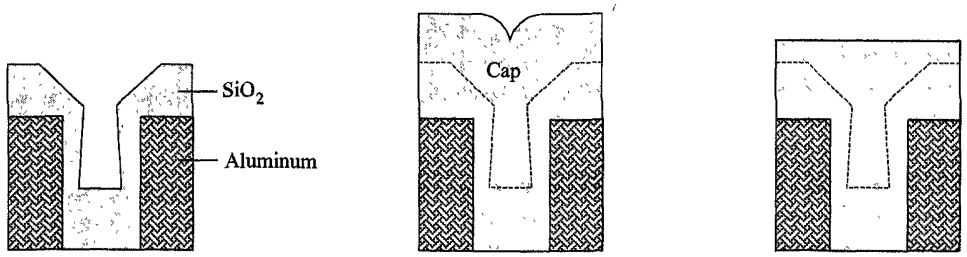
In practice, HDPCVD is sometimes used as the first step in a three-part process for dielectric gap fill.[52] High aspect ratio gaps (> 3:1) are filled with HDPCVD, followed by a regular density PECVD to deposit a cap film (see Figure 11.23). This capping film is then planarized by chemical mechanical planarization (see Chapter 18). Planarization maintains a smooth and constant oxide thickness on top of the metal conductor lines.

DIELECTRICS AND PERFORMANCE

The dielectric materials used in different film layers directly affect the performance of a microchip. Two important aspects of dielectrics are the dielectric constant and device isolation.

Dielectric Constant

The *dielectric constant* (k) of a nonconductive material represents its effectiveness at storing potential electrical energy under the influence of an electric field. In other words, this is the insulating material's ability to act like a capacitor. The lowest attainable k is 1.0 and represents air. A high k dielectric stores more electrical energy. The dielectric constant of thermal silicon dioxide (SiO_2) has a k value of about 3.9. The k value of a plasma-enhanced CVD (PECVD) oxide is about 4.1 to 4.3.



1) HDPCVD gap fill

2) PECVD cap

3) Chemical mechanical planarization

FIGURE 11.23 Three-Part Process for Dielectric Gap Fill

Doped SiO_2 has been traditionally the most common interlayer dielectric (ILD) material. Extensive research is underway to replace doped SiO_2 for the ILD with another dielectric material that has a lower dielectric constant (see Table 11.4).[53] Reducing the k value of the dielectric reduces capacitive losses between adjacent conductors because the dielectric stores less electric field and therefore takes less time to charge, allowing for an increase in speed performance of the metal conductors. A low-k dielectric material for the ILD becomes critical for smaller linewidths with less spacing between metal lines. As the metal linewidths decrease, the capacitive effects of the conductors and insulators increase, and using low-k materials compensates for this.

TABLE 11.4 Potential Low-k Materials for ILD of ULSI Interconnects

Potential Low-k Dielectric	Dielectric Constant (k)	Gap Fill (μm)	Cure Temp. (°C)	Remarks
FSG (silicon oxyfluoride, Si_xOF_y)	3.4 to 4.1	<0.35	No issue	FSG has almost the same k-value as SiO_2 and a reliability concern that fluorine will attack and corrode tantalum barrier metal.
HSQ (hydrogen silsesquioxane)	2.9	<0.10	350 to 450	Silicon-based resin polymer available in solution as FOX (Flowable Oxide) for spin-on coating application. May require surface passivation to reduce moisture absorption. Cure is done in nitrogen.
Nanoporous silica	1.3 to 2.5	<0.25	400	Inorganic material with tunable dielectric constant that relies on pore density. Increased porosity reduces mechanical integrity—porous material must withstand polishing, etching, and heat treatments without degradation.
Poly(arylene) ether (PAE)	2.6 to 2.8	<0.15	375 to 425	Spin-on aromatic polymer with excellent adhesion and ability to be polished with CMP.
a-CF (fluorinated amorphous carbon or FLAC)*	2.8	<0.18	250 to 350	Leading candidate for CVD deposition with high-density plasma CVD (HDPCVD) to produce film with good thermal stability and adhesion.
Parylene AF4 (aliphatic tetrafluorinated poly-p-xylylene)	2.5	<0.18	420 to 450	CVD film that meets adhesion and via resistance requirements with need to maintain gas delivery system at 200°C to control parylene precursor flow rate.

* P. Singer, "Technology News: Wafer Processing," *Semiconductor International* (October 1998): p. 44.

Chip Performance ■ An indicator of chip performance is the speed at which signals are transmitted. Continual die shrinks to achieve sub-0.25 μm technology nodes translate into reduced interconnect linewidths. This reduction leads to increased line resistance (R) for signals. Furthermore, reduced spacing between conductor lines creates more parasitic capacitance (C). The result is an increase in RC signal delay, which slows chip speed and lowers chip performance. This condition is a recent development due to sub-0.25 μm geometries and is often referred to as the *interconnect delay* (see Figure 11.24). In essence, reducing the interconnect dimensions results in more signal delay because of parasitic resistance and capacitance effects. This is the opposite of what occurs in transistors, where reducing gate length serves to reduce gate delay and increase transistor speed.[54]

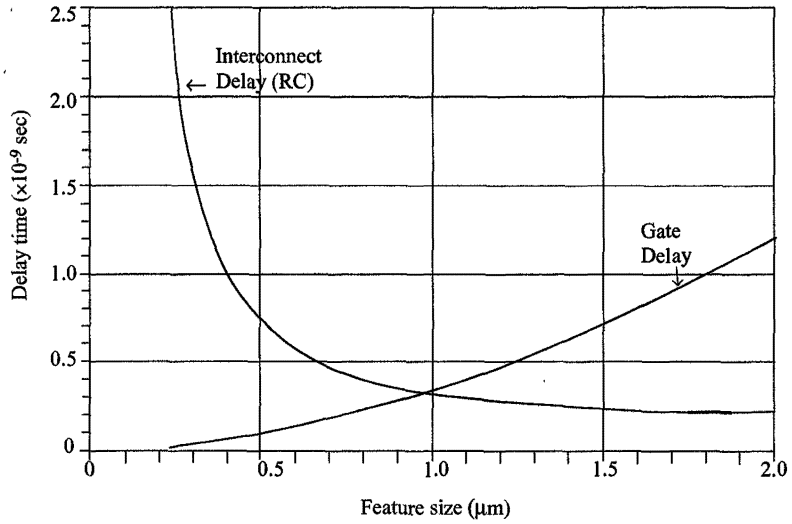


FIGURE 11.24 Interconnect Delay (RC) Versus Feature Size (μm)
(S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications, *Solid State Technology* (March 1996): p. 83.)

The line capacitance, C, is directly proportional to the *k*-value of the dielectric material. A low-*k* dielectric reduces the total interconnect capacitance of the chip (see Figure 11.25), reduces the RC signal delay, and improves chip performance. Lowering the total capacitance also decreases power consumption.[55] The use of a low-*k* dielectric material in conjunction with a low-resistance metal line provides the interconnect system with optimum performance for ULSI technology.

Low-*k* Dielectric Requirements ■ Table 11.5 outlines the typical requirements for a low-*k* dielectric constant film for introduction into wafer fabrication. It is anticipated in the semiconductor industry that the introduction of a low-*k* ILD material will follow the introduction of low-resistance metal lines made of copper (see Chapter 12 for a discussion of copper metallization). The investigation of alternative low-*k* dielectrics is being researched at this time.

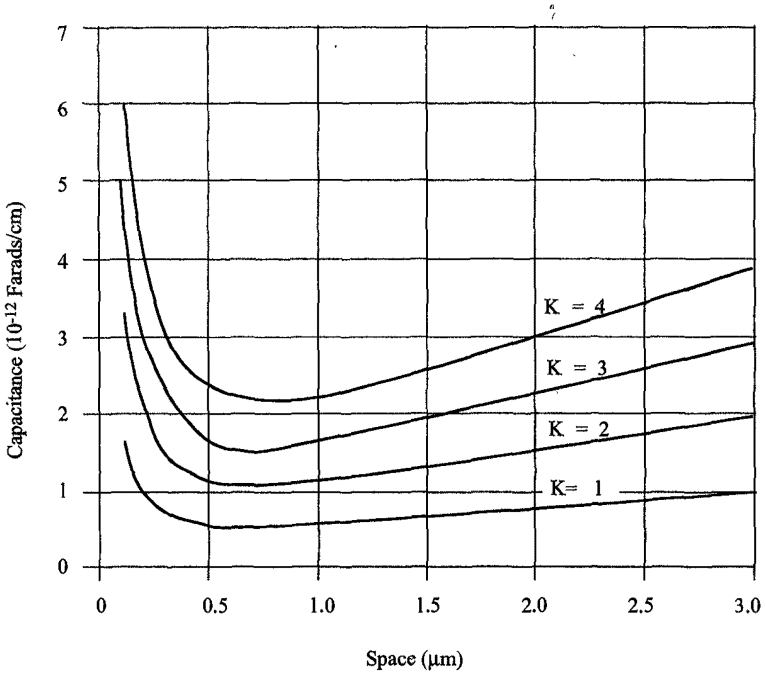


FIGURE 11.25 Total Interconnect Wiring Capacitance
(Used with permission from *Semiconductor International* (September 1998): p. 66.)

TABLE 11.5 Low-*k* Dielectric Film Requirements

Electrical	Mechanical	Thermal	Chemical	Processing	Metallization
Low dielectric constant	Good adhesion	Thermal stability	Resistant: acids and bases	Patternability	Low contact resistance
Low dielectric loss	Low shrinkage	Low coefficient of thermal expansion	Etch selectivity	Good gap fill	Low electromigration (corrosion)
Low leakage	Crack resistant	High conductivity	Low impurities	Planarization	Low stress voiding
High reliability	Low stress		No corrosion	Low pinhole	Hillock (smooth surface)
	Good hardness		Low moisture uptake	Low particulate	Compatible with barrier metals (Ta, TaN, TiN, and so on.)
			Acceptable storage life		

Adapted from P. Singer, "The Future of Dielectric CVD: High-Density Plasmas?" *Semiconductor International* (July 1997): p. 134.

High-*k* Dielectric Constant ■ There is ongoing investigation in the industry for new high-*k* dielectric materials, primarily for DRAM storage capacitors and as an eventual replacement for the extremely thin gate oxide (about 20 Å for 0.18 μm devices). DRAM technology has undergone a 4X increase in storage density every three years for the past 25 years due to processing and design improvements.[56] To attain the required charge storage with a SiO_2 and/or SiN_x dielectric, DRAM storage cell design has evolved into complex stacked capacitor structures (see Figure 11.26 on page 286). If the conventional SiO_2/SiN_x dielectric were replaced with a high-*k* dielectric, a simpler stack structure could be used for lower fabrication costs. A potential high-*k* material is tantalum pentoxide (Ta_2O_5), which has a value for *k* of 20 to 30 and can be easily integrated into the existing process. However, DRAM memory cells are sensitive to leakage and breakdown voltage of storage materials and interfaces. To compensate, the DRAM memory cell that requires a thicker Ta_2O_5

reduces its benefit. Another potential high-k material is barium strontium titanate ($(BaSr)TiO_3$, or BST), which gives a significant improvement in capacitance per unit area over conventional dielectrics.

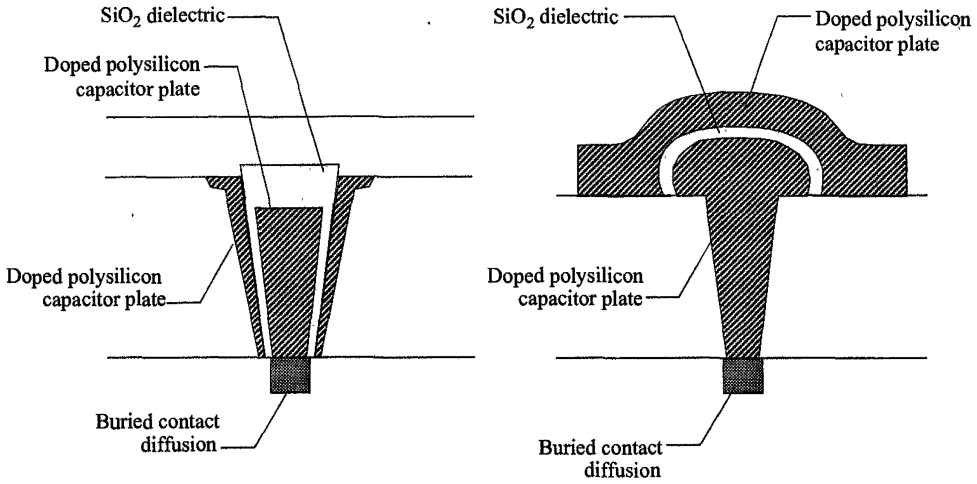


FIGURE 11.26 General Diagram of DRAM Stacked Capacitor

A new high-k gate dielectric is needed for the gate oxide because device scaling is forcing the existing SiO_2 to become extremely thin, with the future oxide thickness expected to be <10 Å for the 50 nm device generation. In a MOS transistor, the gate dielectric material must withstand a significant voltage difference between the gate electrode and the underlying silicon substrate. Thin gate oxides are affected by *tunneling currents,* which becomes a problem when the SiO_2 dielectric layer is below about 15 to 20 Å.[57] With the small gate dimensions in ULSI circuitry, electrons can tunnel through the gate material when the transistor is switched on and off. This leads to threshold voltage drift and eventual circuit failure because the device cannot switch states.

Device Isolation

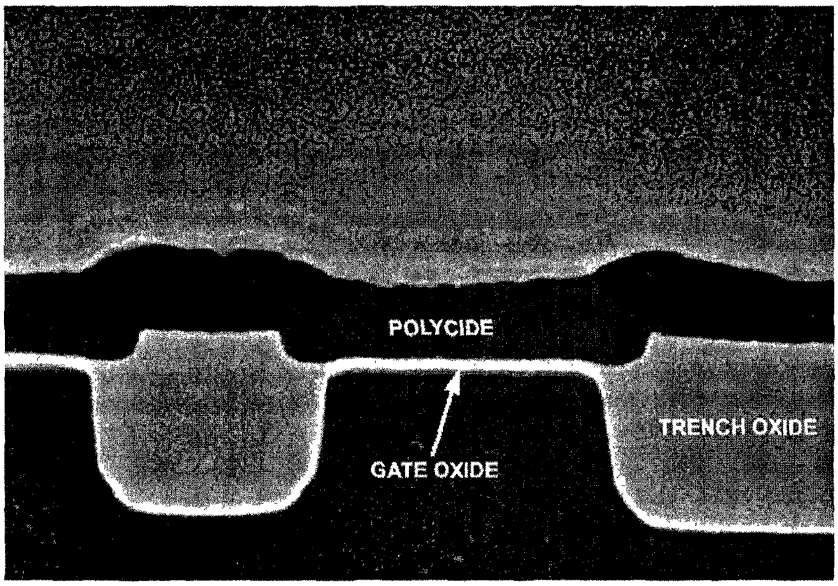
Device isolation in MOS wafer fabrication provides for electrical isolation between devices in the wafer. Isolation techniques serve to reduce or eliminate parasitic field transistors common in MOS planar fabrication (see Chapter 3). Isolation must accommodate scaling between the different device technologies (e.g., junction depth, gate oxide thickness, and so on.). This means that the space allocated for device isolation is being reduced for high-performance ICs. The two basic techniques for device isolation on MOS technologies are local field isolation through local oxidation of silicon (LOCOS), and shallow trench isolation (STI).[58]

Local Oxidation ■ The *local oxidation of silicon* (*LOCOS*) has been the traditional isolation technique for wafer fabrication of devices with critical dimensions of 0.35 μm and larger (see Chapter 10). This technique uses patterned islands of silicon nitride (Si_3N_4) to define the regions for oxidation growth. LOCOS isolation structures are too large for isolation in devices with deep submicron scaling.[59] The major factor limiting LOCOS for 0.25-μm geometries and below is the lateral growth of oxide during silicon oxidation, which imposes natural limits on the minimum area and topography achievable. LOCOS is not acceptable for high-density ULSI technology.

Shallow Trench Isolation ■ *Shallow trench isolation* (*STI*) is the preferred isolation process for wafers fabricated at the 0.25 μm and below technology nodes. The reasons why STI replaced the LOCOS process are:[60]

1. The need for more robust device isolation, especially in DRAM devices.
2. A significant reduction in the surface area for transistor isolation.
3. Superior latchup protection.
4. No channel encroachment.
5. Compatibility with chemical mechanical planarization (CMP).

In the basic STI process (see Chapter 10), trenches about 0.3 to 0.8 μm deep (up to several microns deep for DRAM trenches) are dry etched into the wafer substrate with nearly straight sidewalls and rounded corners (see Chapter 16 for an explanation of etching). The aspect ratio for STI trenches can vary from 2:1 to 5:1, with higher aspect ratios needed in DRAM devices due to their sensitivity to leakage. The trenches have a liner oxide grown on their surface, filled with CVD oxide and planarized using chemical mechanical planarization (CMP). STI technology is a more expensive process than LOCOS technology, primarily because it involves more complex process steps. Nevertheless, the benefits of STI outweigh its higher cost. The process steps for STI, including barrier layers and the liner oxide, are outlined in the CMOS process flow model in Chapter 9.



Shallow Trench Isolation
(Micrograph courtesy of Integrated Circuit Engineering)

SPIN-ON-DIELECTRICS

There are a wide range of low-*k* dielectrics that are designed to use wafer spin as an application method, referred to as spin-on-dielectric (SOD). The process for applying these materials is similar to a familiar process of spin-on-glass (SOG) used in low-cost IC fabrication, which we will review first.

Spin-On-Glass (SOG)

Spin-on-glass (*SOG*) was frequently used for gap fill and planarization of the ILD before the wide acceptance of chemical mechanical planarization (CMP) in the 1990s. SOG materials usually consist of two basic types: organic and inorganic.[61] The organic is based on siloxane, while the inorganic is based on a silicate. The organic siloxane SOG has significant water absorption after curing, is thermally unstable, and does not tolerate plasma exposure. After curing, the silicate SOG behaves like SiO_2, does not absorb excessive moisture, and is thermally stable. However, it shrinks significantly during curing, which leads to stress buildup and cracking if the layer is too thick. For these reasons, the siloxane SOG is usually used as a planarizing layer that undergoes an etchback process with a deposited oxide layer to render it smooth (see Chapter 18). The silicate SOG is used mainly for gap fill applications. There are many different commercial modifications of these two types.

Before applying the SOG, the wafer usually has a predeposited oxide layer, such as a PECVD oxide.[62] The SOG is a liquid and is applied by spinning the wafer at a predetermined speed in either a closed or open bowl. The spin speed and parameters such as ramp-up rate define the SOG thickness. The film is cured, typically at 400°C, and then etched back to planarize the surface. A capping oxide film is applied to seal and protect the SOG, which minimizes water absorption

(see Figure 11.27). Without the cap oxide, there is substantial risk of reliability problems from absorbed moisture, such as through poisoning where the via between the metal layers is corroded and has high resistance.

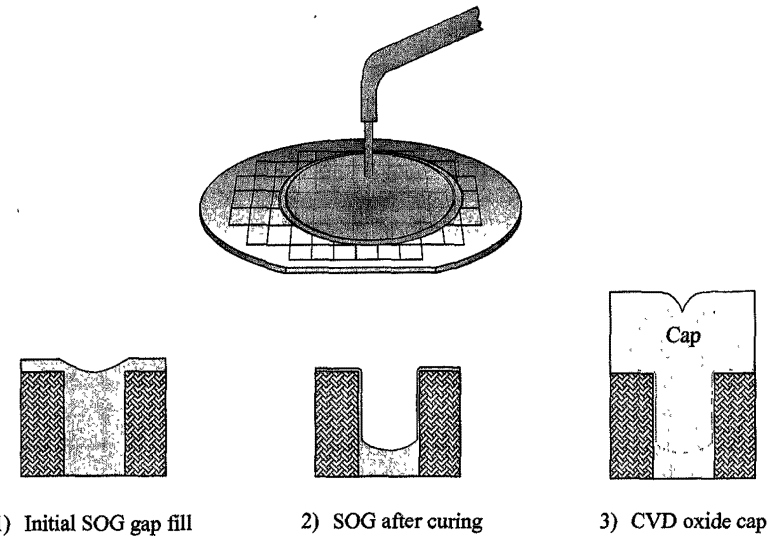


FIGURE 11.27 Gap-Fill with Spin-On-Glass (SOG)

Spin-On-Dielectric (SOD)

Low-*k* dielectric films are currently being investigated as *spin-on-dielectrics* (*SODs*). An example of an SOD low-*k* film is hydrogen silsesquioxane (HSQ), presented earlier in this chapter in Table 11.4 on page 283. The SOD technology is being considered as a cost-effective alternative to CVD processes for depositing low-*k* films. Most SOD applications have used standard spin coaters (used in photolithography which is discussed in Chapter 13) with batch curing in generic furnace tubes. There has been investigation into a single-wafer cluster tool approach to spin coat and cure the film, using a high-temperature hotplate at 350 to 475°C.[63] A proposed spin coat and cure process for HSQ is outlined in Table 11.6.

TABLE 11.6 Proposed HSQ Low-*k* Dielectric Processing Parameters

Major Operation	Process Step	Parameter
Spin Coating	Apply bowl speed	50 rpm
	Maximum bowl speed	800 to 1500 rpm
	Backside rinse	800 rpm, 5 sec
	Topside edge bead removal	1000 rpm, 10 sec
	Spin dry	1000 rpm, 5 sec
Cure	Initial soft-bake cure	200°C, 60 sec, N_2 purge
	In-line cure	475°C, 60 sec, N_2 ambient

In some cases, the nature of the SOD material is an extension of an existing SOG material. Other SOD materials are spin-on polymers that are substantially different than SOG materials. New low-*k* materials have less tendency to absorb moisture and have superior crack resistance. Because of these features, single layer deposition is usually acceptable with no need for a cap oxide layer. In some cases, however, an adhesion promoter is required.[64]

EPITAXY

Epitaxy is the deposition of a thin layer of single-crystal material upon the surface of a single-crystal substrate (see Figure 11.28). This epitaxial layer is often referred to as an epilayer. Epitaxy provides flexibility for the device designer to optimize performance by controlling epilayer doping thickness, concentration, and profile, independent of the wafer. This control is accomplished by the intentional addition of dopants during the epitaxial growth process. An epilayer also reduces the occurrence of latchup in CMOS devices (see Chapter 3). The most common epitaxial reactor in IC production is a high-temperature CVD system.[65]

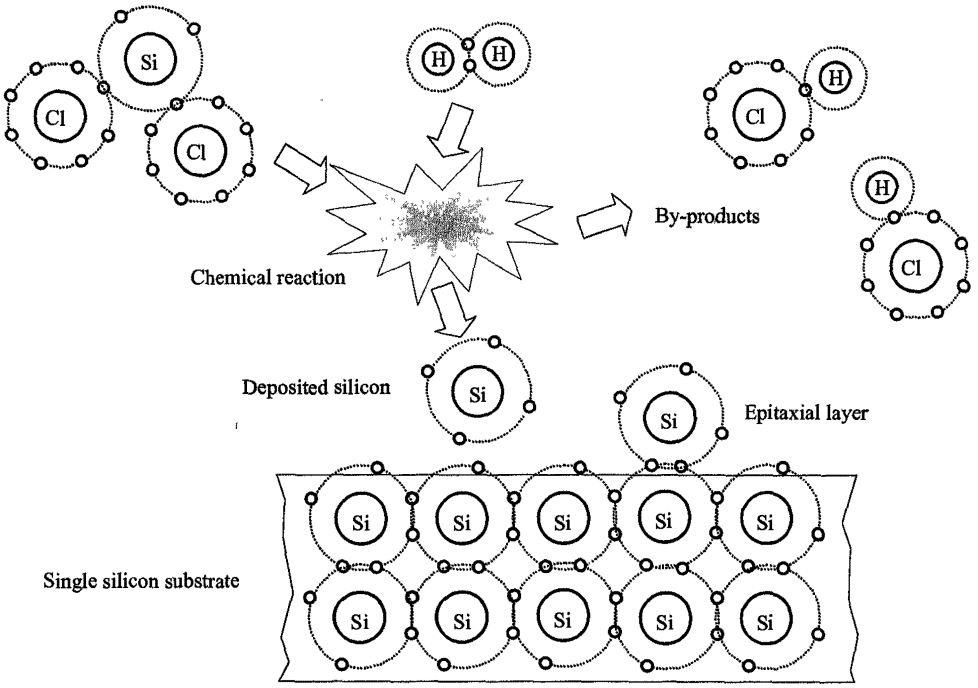


FIGURE 11.28 Silicon Epitaxial Growth on a Silicon Wafer

There are applications under investigation where epitaxy could contribute to advancing future IC performance. An example is *raised source/drain* (*S/D*) structures to achieve low contact resistance.[66] The raised S/Ds are formed by depositing epitaxial silicon on the source, drain, and gate areas of the device. This action effectively increases the surface areas of the S/Ds and therefore lowers the sheet resistance (similar to a larger diameter wire having lower resistance than the same wire with a smaller diameter). These deposits can decrease the contact resistance from 5 Ω/square on conventional devices to 1-2 Ω/square. The raised S/D structure is under investigation for possible implementation at the 0.15-μm technology node.

Undesirable conditions of nonuniform doping can occur during epitaxial deposition.[67] Since lightly doped epilayers are often grown on more heavily doped substrates, there is a condition of *autodoping* into the epitaxial layer. This condition occurs when dopant impurities evaporate from the wafer or are liberated due to chlorine etching of the wafer surface during deposition. These impurities move into the gas stream and cause unintentional doping of the epilayer. As the epilayer grows, less dopant is available from the wafer, and the impurities in the gas stream reach a constant level. Another form of irregular doping is the substrate acting as a source of dopant impurities that diffuse into the epilayer. This action is referred to as *out-diffusion.* Both autodoping and out-diffusion can affect the dopant transition between the substrate and epilayer, causing the transition to be less abrupt than desired.

If the film and substrate are of the same material (e.g., Si film on an Si substrate), then the film growth is *homoepitaxy.* A less common type of epitaxy is heteroepitaxy, which occurs when the film and substrate are of two different materials (e.g., Si on Al_2O_3).

Epitaxy Growth Methods

Epitaxial silicon is commonly deposited using CVD systems. Before epitaxy growth, the wafer must be cleaned of native oxide and any residual organic or metal impurities with a goal of achieving a perfect silicon surface (see Chapter 6). During the epitaxial deposition process, the atoms produced by the gas reaction strike the wafer surface and then move around until they find the correct location to bond to the surface atoms. This action forms the epitaxial layer with the same crystallographic arrangement as the substrate.

Possible gas sources for the epitaxy reaction are the hydrogen reduction of silicon tetrachloride ($SiCl_4$), silane (SiH_4), dichlorosilane (SiH_2Cl_2, or DCS) or trichlorosilane (TCS). Deposition temperatures are from 1050 to 1250°C. Nearly all silicon epitaxy for wafer fabrication is done by the reduction of chlorosilanes (SiH_xCl_{4-x}, where x = 0, 1, 2 or 3) that are diluted in hydrogen.[68] The temperature of the reaction can be lowered if there are fewer chlorine atoms in this precursor. Epitaxial silicon is not commonly grown with silane (SiH_4) because there is excessive particle formation when the silicon deposits on the warm surfaces of the reactor.

Different methods have been used to grow single-crystal layers on silicon wafers, including solid-phase, liquid-phase, vapor-phase, and molecular beam. Three methods used to grow epitaxial layers on silicon wafers for IC production are:

- Vapor-phase epitaxy (VPE)
- Metalorganic CVD (MOCVD)
- Molecular-beam epitaxy (MBE)

Vapor-Phase Epitaxy (VPE) ■ The most common method used for silicon epitaxial growth in wafer fabrication is *vapor phase epitaxy (VPE)*, which is a subset of CVD.[69] Silicon VPE is achieved by passing gas compounds of the desired chemicals over single-crystal silicon wafers that are heated from 800 to 1150°C. The heat from the high temperature provides the energy necessary to drive the chemical reactions, which take place on the surface of the wafer. This process is shown in Figure 11.29.

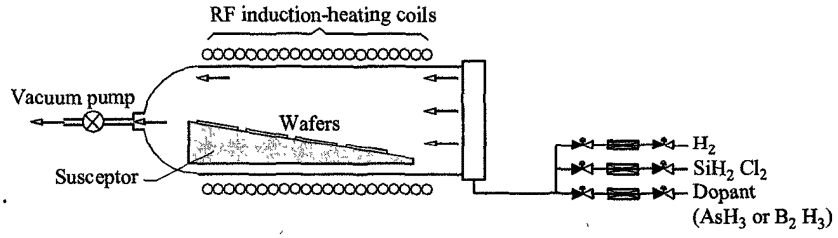


FIGURE 11.29 Illustration of Vapor-Phase Epitaxy

Before processing wafers in a VPE reactor, the system is purged with nitrogen or hydrogen, followed by a vapor HCl etching. The reactant gases, such as chlorosilane, along with dopant gases, are then introduced to the growth chamber where the wafer is heated to the desired temperature. Once the reactant and dopant gases are in the growth reactor, the chemical species undergo the necessary chemical and physical reactions to deposit the doped epilayer.

Typical equipment design for an epitaxial reactor consists of the gas distribution system, a reactor tube, a susceptor to hold and heat the wafer, a control system, and a gas exhaust system. Mass flow controllers and pneumatic valves are used in the gas distribution system to attain tight control over gas flows into the reactor chamber. The susceptor is typically made of graphite or a polysilicon that is coated with silicon carbide or silicon nitride. It must be strong and nonreactive to the reactants and their by-products.[70] Heating the susceptor is accomplished through inductive heating or by radiation from filament lamps. The horizontal and vertical reactors shown in Figure 11.30 are the most common. The susceptor in the horizontal reactor is tilted a few degrees to improve uniformity.

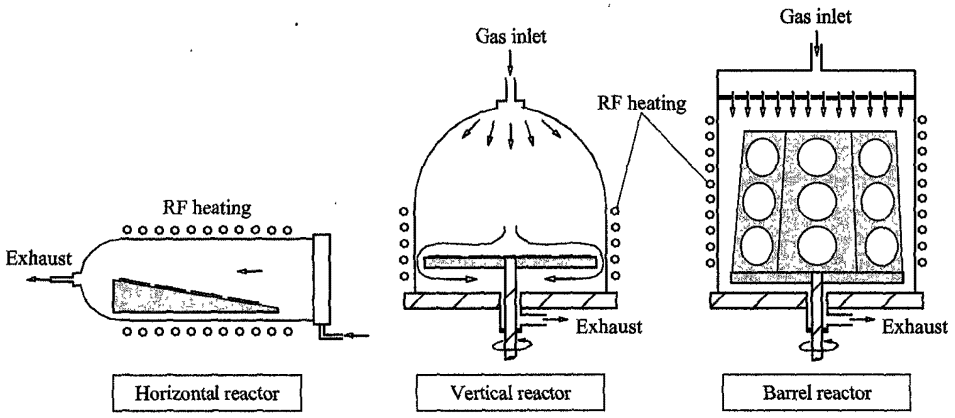


FIGURE 11.30 Silicon Vapor-Phase Epitaxy Reactors

Metalorganic CVD (MOCVD) ■ Another epilayer growth method is commonly referred to as *metalorganic CVD (MOCVD)*, but this is actually a broader term that is equally applicable to the deposition of polycrystalline and amorphous films of metals and oxides.[71] MOCVD is a form of VPE and is not typically used for silicon epitaxy because no suitable sources are readily available. It has been used to deposit compound semiconductor epitaxial layers, such as the III-V compound of gallium arsenide (GaAs) from organometallic sources at low temperature. As with the VPE process, compounds of the desired materials are transported to a heated substrate where the complex chemical reaction takes place on the wafer surface. MOCVD is important for the controlled deposition of ultra-thin, doped, or undoped semiconductor heterolayers (such as GaAs/Si), primarily used for lasers, light-emitting diodes (LED), and optoelectronic integrated circuits.[72] MOCVD is also used in development activity for depositing some organic low-k dielectrics for future IC fabrication.

Molecular-Beam Epitaxy (MBE) ■ *Molecular-beam epitaxy (MBE)* is one of the main methods used to deposit gallium arsenide (GaAs) heteroepitaxy with atomic resolution of thickness. It can also be used for depositing silicon onto a wafer substrate with very tight controls on epilayer thickness and doping uniformity. MBE takes place under conditions of high vacuum, usually at a base pressure of 10^{-10} to 10^{-11} torr or higher, typically with the use of a high vacuum cryopump. The reaction temperature range is between 500 to 900°C.

Most silicon MBE systems produce silicon atoms for the epitaxial reaction through the evaporation of silicon with an electromagnetically focused electron beam source. This process is similar to vacuum evaporation for deposition. The beam of silicon atoms leaving the silicon source travels through the evacuated chamber without collisions and deposits on the single-crystal wafer surface. Another more recent method is to deliver silicon atoms via a gas source at very low flow rates. The silicon growth rate of an MBE system is determined by measuring the number of atoms that leave the source and the fraction that actually strike the wafer surface and stick.

CVD QUALITY MEASURES

Important quality measures for CVD are listed in Table 11.7.

TABLE 11.7 Key Quality Measures for CVD

Quality Parameter	Types of Defects	Remarks
1. Voids during deposition of high-aspect ratio gap (>3:1) with PECVD SiO_2.	A. Keyhole voids are formed during deposition of high aspect ratio gaps. After chemical mechanical planarization (CMP) and removal of top surface, some voids become trenches (see Figure 11.31 on page 294). This trench can lead to an open circuit after metal deposition.	• Gap fill is critical as CDs are reduced and gaps have a high aspect ratio. • Voids are highly stressed regions and trap moisture or solvents that cause corrosion or outgas at high vacuum. • The root cause of this problem is trying to fill a high aspect ratio gap with a PECVD that is limited to deposition system. This application requires HDPCVD.
2. Film stress.	A. High film stress leads to cracking and delamination. B. Film stress can propagate silicon defects in the substrate. C. Stress may cause current leakage.	The presence of dopant in a glass film can reduce stress. Deposition parameters that affect film stress are: • RF power: adjust power to improve stress (e.g., reduce power by 10 watts to lower stress). • Pressure: adjust pressure to affect stress (e.g. increased pressure leads to higher film stress).
3. Film thickness.	A. Thickness of film exceeds requirements.	Deposition parameters that affect thickness are: • Time: reduce deposition time to reduce thickness (e.g., 1 sec reduction for each 100 Å of thickness). • Gas-flow rate: reduce gas flow to reduce thickness (e.g., reduce SiH_4 by 2 sccm for >5% thickness over nominal).
4. Refractive index (R/I).	A. R/I is a good monitor to assess quality of the film. B. R/I is highly dependent on the composition of the deposited film (stoichiometry).	R/I is an optical property of a material. Compared to thermal oxides, CVD oxides are inferior in quality and integrity. CVD oxides exhibit more particles and pinholes. The R/I of CVD SiO_2 can be compared to the R/I of SiO_2 (1.46) to assess relative quality: • High R/I means excessive Si in film. • Low R/I indicates a porous film, which tends to absorb moisture.

CVD TROUBLESHOOTING

Troubleshooting techniques for common CVD problems are described in Table 11.8. Plasma deposition systems have two areas of safety: chemical and electrical. Many plasmas use toxic gases (see Appendix A, Table A.3). There are high voltages and RF energy in the RF power supplies. Use caution when troubleshooting CVD systems.

TABLE 11.8 Common CVD Troubleshooting Problems

Problem	Probable Cause	Corrective Action
1. Particle contamination associated with film.	Source of particles is isolated by determining whether particles are found on top of film, in film, or below film: A. On film: particles formed after deposition. Check for particles on sidewalls of hot-wall reactors and belt-driven reactors. B. In film: Gas-phase nucleation particles from gas too rich in silane or silicon. Gas supply contamination causes particles. C. Under film: Particles from silicon carbide, quartz, or reactor walls fall on wafer before deposition.	• Particles on film: Particles on sidewalls indicate need for more frequent wet cleaning of quartzware and chamber surfaces for preventive maintenance. Inspect in situ dry cleaning process. Also verify the correct procedures for cleaning (manually or in situ). • Particles in film: Improper gas flows from MFC calibration or problem in process recipe of software program. Check for leaks in gas supply system or O-rings. Verify point-of-use filters are acceptable. • Particles under film: Check wafer cleaning process before deposition.
2. Film thickness.	Thickness is related to both equipment and process problems. Variables affecting thickness are: A. Incorrect temperature control. B. System pressure is too high or low. C. System power needs adjustment. D. Improper gas flow.	• Temperature controller may require calibration. A common problem is a defective thermocouple. • System pressure is controlled by process recipe. Check vacuum system for leaks. • Adjust RF power to optimize film thickness. • Check calibration of MFC to ensure proper gas flow. • If checking thickness with test wafer, verify the test wafer is clean and does not have a thickness variation before the test.
3. Cracking of dielectric material on top of electrostatic chuck (ESC).	High thermal load at the ESC causes the dielectric material to erode or crack. This condition leads to plasma arcing or wafer chucking/dechucking problems.	• Investigate the ESC backside cooling system to ensure it is functioning properly. • Inspect ESC materials to verify there is no breakdown due to high power exposure, temperature, or plasma clean conditions.

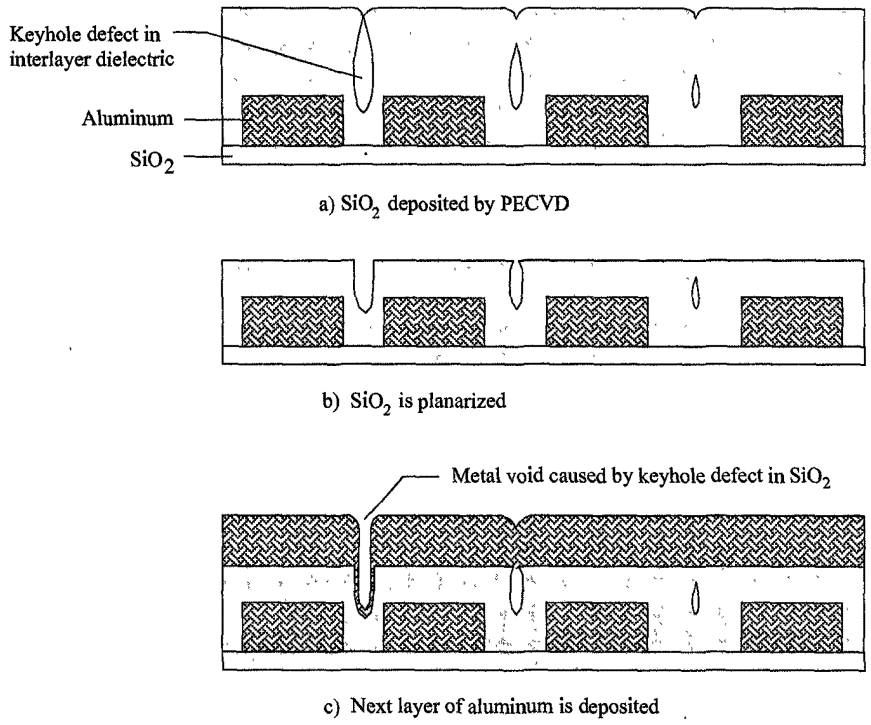


FIGURE 11.31 Effects of Keyholes in ILD on Metal Step Coverage

SUMMARY

Multilevel metallization with dielectric and metal layers is necessary for advanced ICs. Thin films have special characteristics to meet the requirements for wafer fabrication, including high aspect ratio gap fill. A film grows in stages from a cluster of nuclei to a continuous film. There is a wide range of tools for depositing film layers, with the focus on dielectric chemical vapor deposition (CVD). CVD deposits a thin film on the wafer surface through a chemical reaction that is assisted by thermal or plasma energy. CVD surface reactions include pyrolosis, oxidation, and reduction. The CVD chemical reaction follows defined steps and can be limited by the supply of chemicals or by the nature of the surface reaction. The thin film is sometimes doped to obtain improved performance. Reactors for CVD are categorized by atmospheric pressure CVD (APCVD), low pressure CVD (LPCVD), or plasma-assisted CVD. APCVD is simple in design and is primarily used to deposit conformal oxide with a TEOS-O_3 process. LPCVD deposits various films, including oxide, silicon nitride, and polysilicon. The LPCVD process is reaction-rate limited (gas flow is not critical) and often occurs at lower temperatures. Plasma-assisted CVD is divided into plasma-enhanced CVD (PECVD) and high-density plasma CVD (HDPCVD). PECVD was the first plasma-assisted CVD but had limited use for filling gaps less than 0.5 μm. HDPCVD is used for advanced ICs because of its excellent gap-fill capability for high-aspect ratios at relatively low temperatures. The deposited dielectric material increases chip performance by reducing the RC interconnect delay with a low-*k* dielectric. Dielectric material is also important for isolation technologies. Dielectric materials have been applied in liquid form by spinning, which may be an application method for future low-*k* dielectrics. Epitaxy is grown primarily by CVD methods, including vapor-phase epitaxy (VPE), metalorganic CVD (MOCVD), and molecular-beam epitaxy.

KEY TERMS

multilevel metallization
aluminum metallization
critical layers
noncritical layers
first interlayer dielectric (ILD-1)
premetal dielectric (PMD)
interlayer dielectric (ILD)
thin film
stoichiometry
aspect ratio
nucleation
nuclei coalescence
continuous film
chemical vapor deposition (CVD)
reactor
heterogeneous reaction (also surface catalyzed)
homogeneous reaction
adsorption
desorption
species
precursors
mass-transport limited
reaction-rate limited
kinetically controlled
boundary layer
stagnant layer
phosphosilicate glass (PSG)
borosilicate glass (BSG)
borophosphosilicate glass (BPSG)
fluorosilicate glass (FSG)
cluster tools
hot-wall reactor
cold-wall reactor
atmospheric pressure CVD (APCVD)
silane
TEOS oxide
low pressure CVD (LPCVD)
silicon nitride (Si_3N_4)
silicon oxynitride (SiO_xN_y)
plasma-enhanced CVD (PECVD)
PETEOS
silicon nitride PECVD
high-density plasma CVD (HDPCVD)
dep:etch ratio
dielectric constant (k)
device isolation
local oxidation of silicon (LOCOS)
shallow trench isolation (STI)
spin-on-glass (SOG)
spin-on-dielectric (SOD)
epitaxial growth
raised source/drain structures
autodoping
out-diffusion
homoepitaxy
heteroepitaxy
vapor-phase epitaxy (VPE)
metalorganic CVD (MOCVD)
molecular-beam epitaxy

REVIEW QUESTIONS

1. What is multilevel metallization? Why is it necessary for chip fabrication?
2. What is aluminum metallization? Describe a chip critical layer and a noncritical layer.
3. Explain the purpose of an ILD layer. Where is the ILD-1 layer located on a chip?
4. What is a thin film? List and describe eight characteristics of an acceptable thin film.
5. Define the aspect ratio. Why is a high-aspect ratio important for ULSI devices?
6. List and describe the three stages of thin film growth.
7. List the five major techniques for deposition.
8. Define chemical vapor deposition (CVD).
9. What are the five basic CVD chemical reactions? Give a short description of each reaction.
10. In a CVD reaction, is a heterogeneous or homogeneous reaction preferred? Why?
11. Identify and describe the eight steps in a CVD reaction.
12. Explain the difference between a mass-transport limited and a reaction-rate limited CVD process. Which process is temperature dependent?
13. What is a boundary layer in CVD? Under what conditions does it become a stagnant layer?
14. What is the beneficial effect of low pressure in a CVD reaction?
15. What do the abbreviations PSG, BPSG, and FSG stand for? What are three benefits of adding dopants to oxide?

16. Describe the different types of CVD reactors and their principal advantages.
17. Explain the difference between a hot-wall reactor and cold-wall reactor for CVD.
18. Explain APCVD. What is the principal disadvantage with APCVD SiO_2, using silane as a source?
19. What is TEOS? Discuss the primary advantages to using APCVD TEOS-O_3.
20. Why is LPCVD more common than APCVD? Describe LPCVD processing.
21. Why does LPCVD operate in the reaction-rate regime?
22. State what film is deposited using LPCVD TEOS, and explain the advantage of this film.
23. Why is silicon nitride often used as a passivation layer?
24. What CVD tool is used to deposit the polysilicon gate material? List six reasons why polysilicon is used as a gate electrode.
25. What are the benefits to using a silicon oxynitride film?
26. State six advantages to using plasma during CVD.
27. What is PECVD? What is the main difference between PECVD and LPCVD?
28. Describe the main differences between silicon nitride PECVD and LPCVD.
29. Discuss a major limitation for using PECVD in ULSI microchips.
30. Explain what HDPCVD is. What are its main benefits in advanced ICs?
31. Describe the effect wafer biasing has on HDPCVD directionality.
32. Explain simultaneous deposition and etching for HDPCVD. What is the value for the typical dep-etch ratio?
33. List and describe the five steps used for simultaneous deposition-etching in CVD.
34. Discuss the importance of the dielectric constant of the ILD.
35. Explain the interconnect delay and why it is beneficial to lower the ILD constant dielectric.
36. What is an application for a high-k dielectric material and why is it needed?
37. What are LOCOS and STI? Why has STI replaced LOCOS for advanced ICs?
38. List the processing steps for STI.
39. What is spin-on-glass? Explain spin-on-dielectric and what it could be used for in future applications.
40. Describe what epitaxy is. Define the terms *autodoping* and *out-diffusion*.
41. List and discuss three methods for epitaxial growth.

DEPOSITION EQUIPMENT SUPPLIERS' WEBSITES

Amtech Systems Inc.	http://www.amtechsystems.com/
Applied Materials	http://www.appliedmaterials.com/products/
ASM	http://www.asm.com/
CVC Incorporated	http://www.cvc.com/
CVD Equipment Corporation	http://www.cvdequipment.com/
Genus Incorporated	http://www.genus.com/
Kokusai Semiconductor Equipment	http://www.ksec.com/
Novellus Systems Inc.	http://www.novellus.com/index.htm
TEL, Tokyo Electron Ltd.	http://www.teainet.com
Tystar Corporation	http://www.tystar.com/

REFERENCES

1. R. Liu, "Metallization," *USLI Technology,* ed. C. Change and S. Sze (New York: McGraw-Hill, 1996), p. 379.
2. R. Jackson et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998): p. 49.
3. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics for Future ICs," *Solid State Technology* (September 1998): p. 69.
4. P. Singer, "The Future of Dielectric CVD: High-Density Plasmas?" *Semiconductor International* (July 1997): p. 134.
5. S. Sivaram, *Chemical Vapor Deposition: Thermal and Plasma Deposition of Electronic Materials* (New York: Van Nostrand Reinhold, 1995), p. 2.
6. Adapted from S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era*, vol. 1 *Process Technology,* 2nd ed., (Sunset Beach, CA: Lattice Press, 2000) pp. 108–114.
7. J. Baliga, "Options for CVD of Dielectrics Include Low-k Materials," *Semiconductor International* (June 1998): p. 139.
8. C. Rye et al., "Barriers for Copper Interconnections," *Solid State Technology* (April 1999): p. 53.
9. P. Singer, "Film Stress and How to Measure It," *Semiconductor International* (October 1992): p. 54.
10. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., (New York: John Wiley & Sons, 1994), p. 522.
11. Ibid.

12. G. Anner, *Planar Processing Primer* (New York: Van Nostrand Reinhold, 1990), p. 361.
13. G. Anner, *Planar Processing Primer,* p. 361.
14. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology* (Sunset Beach, CA: Lattice Press, 1986), p. 151.
15. A. Jones and P. O'Brien, *CVD of Compound Semiconductors: Precursor Synthesis, Development and Applications* (Weinheim, Germany: VCH, 1997), p. 31.
16. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology,* p. 151.
17. Ibid., p. 154.
18. Ibid., p. 154.
19. G. Anner, *Planar Processing Primer,* p. 365.
20. SEMATECH, "Desposition Processes," *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 6.
21. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 532.
22. S. Nag et al., "Low-Temperature Pre-Metal Dielectrics," p. 69.
23. J. Baliga, "Options for CVD of Dielectrics," p. 140.
24. H. Cheng, "Dielectric and Polysilicon Film Desposition," *ULSI Technology,* ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 211.
25. M. Hammond, "Introduction to Chemical Vapor Deposition," *Solid State Technology* (December 1979): p. 61.
26. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 529.
27. G. Schwartz and K. Srikrishnan, "Interlevel Dielectrics," *Handbook of Semiconductor Interconnection Technology,* ed. G. Schwartz, K. Srikrishnan, and A. Bross (New York: Marcel Dekker, 1998), p. 231.
28. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies* (Boston: Kluwer Academic Publishers, 1995), p. 118.
29. J. Baliga, "Options for CVD of Dielectrics," p. 140.
30. G. Schwartz and K. Srikrishnan, "Interlevel Dielectrics," p. 233.
31. H. Cheng, "Dielectric and Polysilicon Film Deposition," *ULSI Technology,* p. 211.
32. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 524.
33. Ibid., p. 520.
34. C. Gugliemini and A. Johnson, "Properties and Reactivity of Chlorine Trifluoride," *Semiconductor International* (June 1999): p. 162.
35. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 529.
36. H. Cheng, "Dielectric and Polysilicon," p. 225.
37. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* Vol. 1, *Process Technology,* 2nd ed., p. 175.
38. Y. Trouiller et al., "Inorganic Bottom ARC SiO_xN_y for Interconnection Levels on 0.18 μm Technology," *SPIE* 3508, (1998), p. 122.
39. L. Peters, "Pursuing the Perfect Low-*k* Dielectric," *Semiconductor International* (September 1998), p .122.
40. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* p. 125.
41. Ibid.
42. SEMATECH, "Desposition Processes," *Furnace Processes and Related Topics,* p. 35.
43. C. Apblett et al., "Silicon Nitride Growth in a High-Density Plasma System," *Solid State Technology* (November 1995): p. 73.
44. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 535.
45. E. Korczynski, "HDP-CVD: Trying to Lasso Lightning," *Solid State Technology* (April 1996): p. 63.
46. P. Singer, "Future of Dielectric CVD," p. 128.
47. P. Burggraaf, "Advanced Plasma Sources: What's Working?" *Semiconductor International* (May 1994): p. 57.
48. J. Bondur et al., "Impact of Electrostatic Chuck Performance on HDP CVD SiO_2 Films," 52nd Symposium on Semiconductors and Integrated Circuits Technology, Osaka, Japan.
49. P. Singer, "Future of Dielectric CVD," p. 127.
50. E. Korczynski, "HDP-CVD," p. 64.
51. P. Singer, "Future of Dielectric CVD," p. 127.
52. Ibid.
53. L. Peters, "Pursuing the Perfect Low-*k* Dielectric," p. 68.
54. M. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *Solid State Technology* (September 1996): p. 105.
55. S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications," *Solid State Technology* (March 1996): p. 83.
56. D. Kotecki, "High-*k* Dielectric Materials for DRAM Capacitors," *Semiconductor International* (January 1998): p. 38.
57. P. Singer, "Wafer Processing: New Gate Dielectric Material Needed," *Semiconductor International* (January 1998): p. 38.
58. P. Van Cleemput et al., "HDPCVD Films Enabling Shallow Trench Isolation," *Semiconductor International* (July 1997): p. 180.
59. Ibid., p. 182.
60. L. Peters, "Choices and Challenges for Shallow Trench Isolation," *Semiconductor International* (April 1999): p. 69.
61. R. Liu, "Metallization," p. 419.
62. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* p. 569.
63. T. Batchelder et al., "In-Line Cure of SOD Low-*k* Films," *Solid State Technology* (March 1999): p. 29.
64. L. Peters, "Pursuing the Perfect Low-*k* Dielectric," p. 66.
65. S. Campbell, *The Science and Engineering of Microelectronic Fabrication* (New York: Oxford University Press, 1996), p. 340.
66. L. Peters, "Is the 0.18 μm Node Just a Roadside Attraction?" *Semiconductor International* (January 1999): p. 52.
67. R. Jaeger, *Introduction to Microelectronic Fabrication* vol. V, (Reading, MA: Addison-Wesley, 1988), p. 125.

68. S. Campbell, *Science and Engineering,* p. 347.

69. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., p. 283.

70. P. J. Wang, "Epitaxy," *ULSI Technology,* ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 124.

71. A. Thompson, R. Stall, and B. Kroll, "Advances in Epitaxial Deposition Technology," *Semiconductor International* (July 1994): p. 173.

72. B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* p. 102.

CHAPTER 12

METALLIZATION

Metallization in wafer fabrication is the process of depositing metal film over a dielectric film and later patterning it to form the interconnecting metal lines and plugs of integrated circuits. This is similar to using insulated copper wire in an automobile to interconnect all the electrical components to make a fully-functional electrical system. The metal lines are sandwiched between dielectric layers for electrical integrity. High-performance microprocessors use metal lines to interconnect tens of millions of devices on one chip. Transistor densities are projected to reach 1 billion transistors per chip by the year 2010, with a corresponding increase in interconnect complexity.

Microchip interconnect technology has become a critical challenge for future IC performance due to the need to decrease signal propagation delay. Because the density of ULSI circuit elements is increasing, interconnect resistance and parasitic capacitance increase, thereby slowing the signal propagation.

A significant change currently underway in wafer fabrication technology is the reduction in interconnect metal resistivity, ρ. This reduction is achieved by the replacement of aluminum alloy with copper as the primary conducting metal. With deep submicron linewidths, there is also a need for a low-k interlayer dielectric (ILD). Lowering the dielectric constant will reduce parasitic capacitance, which contributes to signal delay.

Traditionally, the methods used to deposit metals on a wafer have been physical processes. On the other hand, processes used to deposit insulating and semiconducting layers usually involved the CVD chemical reactions studied in Chapter 11. This separation into physical and chemical processes is less defined with the introduction of new IC metallization technology.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the terminology for metallization.
2. List and describe the six categories of metals used in wafer fabrication. Discuss the performance requirements and give applications for each metal category.
3. Explain the benefits for using copper metallization in wafer fabrication. Describe the challenges for implementing copper.
4. State the advantages and disadvantages to sputtering.
5. Describe the physics of sputtering and discuss different sputtering tools and applications.
6. Describe the benefits and applications for metal CVD.
7. Explain the fundamentals of copper electroplating.
8. Describe a process flow for dual-damascene processing.

INTRODUCTION

Wafer metallization is the deposition of a thin film of conductive metal onto a wafer by use of a chemical or physical process. This process is closely linked to dielectric deposition. Metal lines conduct the signal through the IC circuit while the dielectric layers ensure signals are not influenced by adjacent lines. Both metal and dielectrics are thin film processes. In some cases, metal and dielectrics are deposited by the same type of equipment.

Metallization has unique terminology for the different metal connections. The term *interconnect* describes the conductor materials, such as aluminum, polysilicon, or copper, that create the metal wiring that carries electrical signals to different parts of the chip. Interconnect is also used as a general term for the wiring between devices on a die and the overall package. A *contact* is the electrical connection at the silicon surface between the devices in the silicon wafer and the first metal layer. *Vias* are openings that pass through the various dielectric layers to form an electrical pathway from one metal layer to the adjacent metal layer. A metal *plug* fills the vias to form the electrical connection (interconnect) between the two metal layers. These connections are illustrated in Figure 12.1.

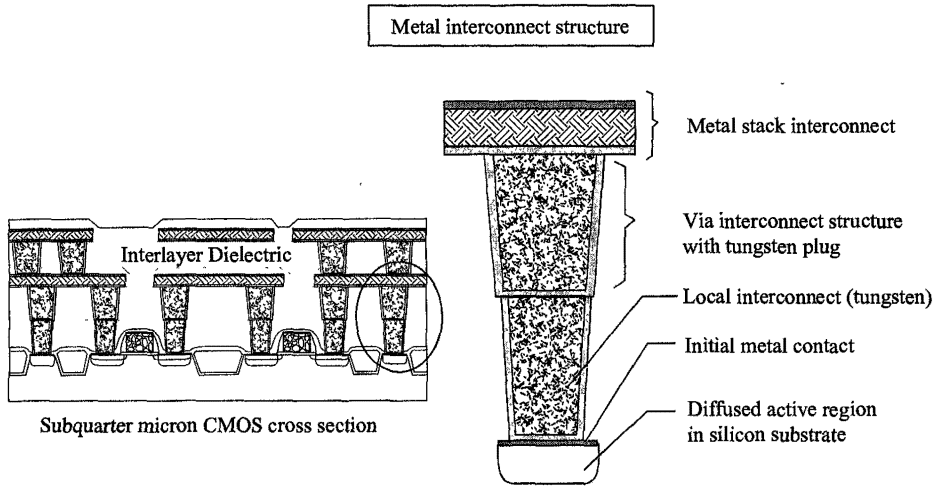


FIGURE 12.1 Multilevel Metallization

The interlayer dielectric (ILD) is an insulating material that electrically separates the metal levels. Once deposited, the ILD is patterned and etched to form via pathways for the various metal layers and the silicon. The vias are filled with a metal, traditionally tungsten (W), to form the via plug. Appreciate that there are many vias on a wafer, up to 10^{11} vias on each individual layer of a 300-mm product wafer.[1] This process of creating vias in the ILD is repeated for every ILD layer on the die. In traditional metallization, a blanket layer of aluminum alloy metal is deposited on the dielectric layer and then patterned and etched to form metal lines. Metal etch is an important technology for traditional metallization.

Metallization is in a transition period and is undergoing rapid change with the introduction of copper metallurgy to replace aluminum alloy. This change is due to the difficulty of etching copper. To overcome this problem, copper metallurgy uses the dual damascene process to form vias and copper interconnects (explained later in this chapter). This metallization process is literally the opposite approach of traditional metallization (see Figure 12.2). Damascene starts with deposition of a blanket dielectric, planarizes the dielectric, patterns and etches holes or trenches in the dielectric for vias and connecting metal lines, deposits blanket metal into the trenches, and then planarizes the metal down to the dielectric to define the metal interconnects.

Metallization technology is critical for increasing performance in advanced ICs, as explained in Chapter 11. Reducing chip performance signal delay caused by interconnect lines was not a significant concern for older IC technologies. The dominant signal delay has traditionally been caused by the device. However, for newer ULSI products manufactured with denser wiring, signal delay

due to interconnect has become a larger portion of the clock cycle time and has more effect on limiting the IC performance.[2]

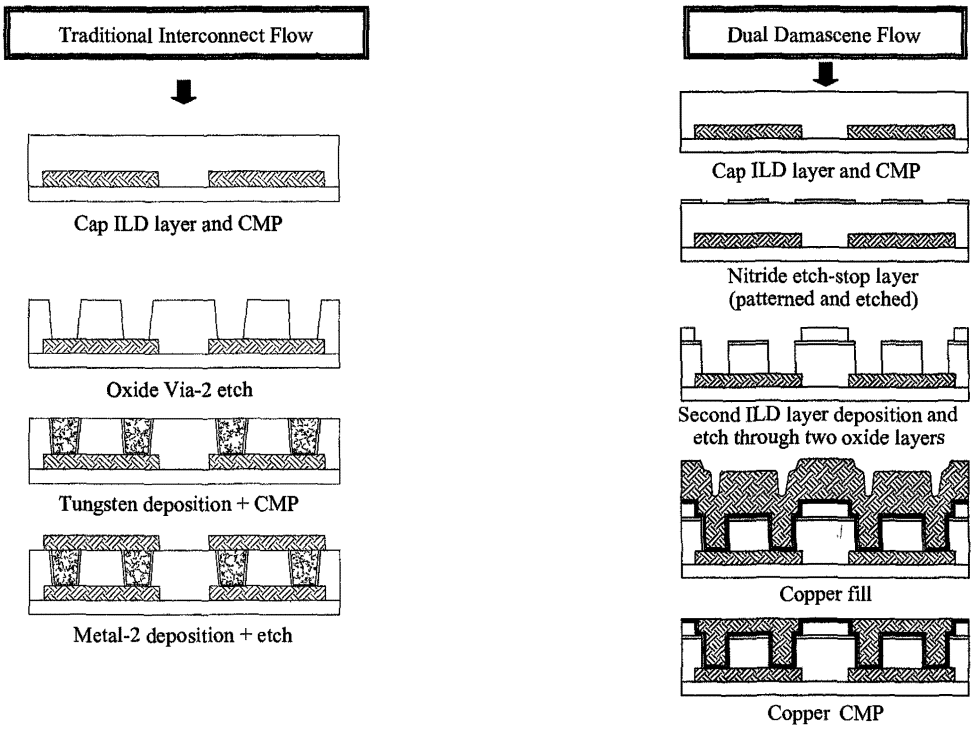
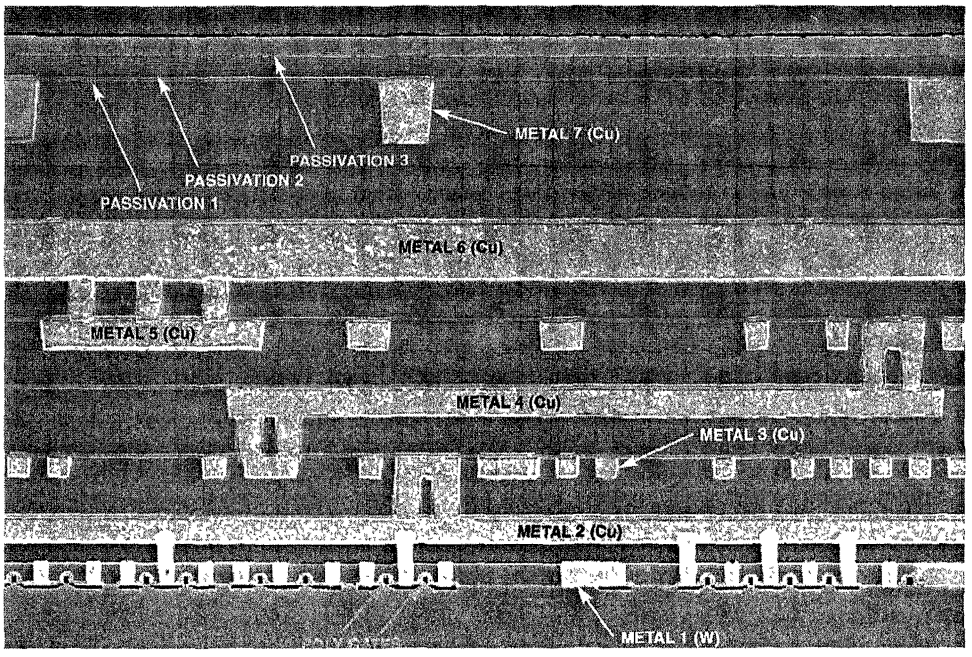


FIGURE 12.2 Traditional Versus Damascene Metallization



Copper Metallization
(Micrograph courtesy of Integrated Circuit Engineering)

TYPES OF METALS

The types of metals and metal alloys used for chip interconnects are evolving based on the performance requirements. Requirements for a successful metal material are:[3, 4]

1. **Conductivity:** Must be highly conductive and capable of handling high current densities while maintaining electrical integrity.
2. **Adhesion:** Ability to adhere to the underlying substrate and easily connect to external connections. Low contact resistance to semiconductor and metal interfaces.
3. **Deposition:** Readily deposited with a uniform structure and composition (for alloys) by a relatively low-temperature process. Deposition into high-aspect ratio gaps for damascene metallization technique.
4. **Patterning/Planarization:** High-resolution patterning for traditional Al-based metallization without etching the underlying dielectric. Ease of planarization for damascene metallization.
5. **Reliability:** The metal is relatively soft and ductile in order to withstand cyclic temperature variations during processing and service.
6. **Corrosion:** High resistance to corrosion with minimal chemical interactions with the adjacent layers and underlying device regions.
7. **Stress:** Resistance to mechanical stress to reduce wafer distortion and material failures for reasons such as cracking, void formation, and stressed-induced corrosion.

The melting temperature and resistivities of different metals found in wafer fabrication and silicon are shown in Table 12.1.[5]

TABLE 12.1 Silicon and Select Wafer Fab Metals (at 20°C)

Material	Melting Temperature (°C)	Resistivity (μΩ-cm)
Silicon (Si)	1412	$\approx 10^9$
Doped Polysilicon (Doped Poly)	1412	≈ 500 to 525
Aluminum (Al)	660	2.65
Copper (Cu)	1083	1.678
Tungsten (W)	3417	8
Titanium (Ti)	1670	60
Tantalum (Ta)	2996	13 to 16
Molybdenum (Mo)	2620	5
Platinum (Pt)	1772	10

The various metals and metal alloys used in wafer fabrication can be grouped into the following categories:

- Aluminum
- Aluminum-copper alloys
- Copper
- Barrier metals
- Silicides
- Metal plugs

Aluminum

The earliest interconnect metal in semiconductor fabrication was aluminum, and it is still the most common interconnect metal in wafer fabrication. Aluminum is projected to continue as the dominant fab interconnect metal for several more years.[6] Copper interconnect metal is expected to replace aluminum for high-performance IC fabrication during the early 2000s. Nevertheless, it is beneficial to review the background into the selection of aluminum for metallization because many of the same basic technical challenges exist today.

Selection of Aluminum ■ Aluminum has been one of the major materials used in wafer fabrication, along with silicon and silicon dioxide. It is used during wafer fabrication in thin film form to interconnect the different devices in the silicon wafer (see Figure 12.3). At the same time, aluminum is one of the thickest films deposited on the wafer, with the first metal layer deposition about 5,000 Å thick. The upper noncritical layers on a wafer (e.g., metal layer with bonding pads) can range up to 20,000 Å thick.

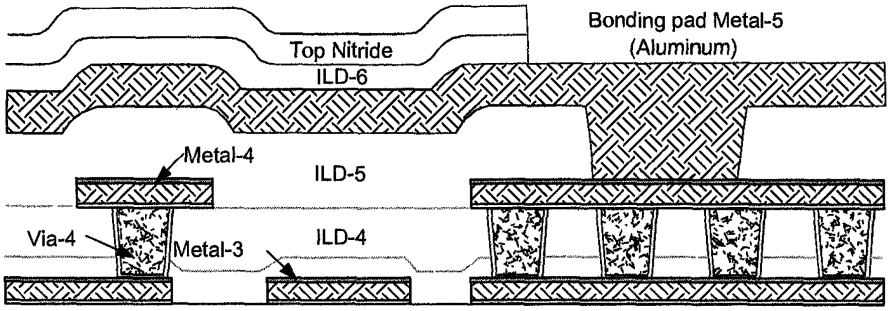


FIGURE 12.3 Aluminum Interconnect

Aluminum's low resistivity of 2.65 μΩ-cm at 20°C has similar but slightly higher resistivity than copper, gold, and silver. However, silver and copper are both prone to corrosion and have high diffusivity in Si and SiO_2, a fact that keeps them from being used for semiconductor fabrication. Gold and silver are much more expensive than aluminum and do not adhere well to oxide films. Gold was sparingly used in the beginning of wafer fabrication, but its high contact resistance with silicon required a platinum barrier metal. On the other hand, aluminum reacts readily with silicon dioxide (SiO_2) when heated to form a thin layer of aluminum oxide (Al_2O_3), which promotes adhesion between the silicon dioxide and aluminum. Aluminum is easily deposited on the wafer and etches in solutions that do not attack underlying films. For these reasons, aluminum was selected as the preferred metal for metallization.

Overall, aluminum is compatible with the major processes in silicon IC fabrication and is relatively inexpensive, making it the choice for metallization since the early days of IC fabrication. However, because of increased circuit density, increased number of metal layers on a wafer, and the reduction in linewidth due to scaling, the metallization technology has evolved from the simple one-level metal layer to a scheme of multiple metal layers. With its lower resistivity, copper is desirable to replace aluminum as the primary interconnect material.

Ohmic Contact ■ Before VLSI circuitry, pure aluminum was used for metallization. Silicon melts at a temperature of 1412°C and pure aluminum melts at 660°C. However, aluminum and silicon alloyed together actually have a lower melting temperature depending on their composition. For an alloy of 88.7% aluminum and 11.3% silicon, the alloy melts at 577°C, which is referred to as the *eutectic temperature.*[7] The eutectic temperature is the lowest temperature that an alloy melts at a particular eutectic composition.

To form a contact between aluminum and silicon, it is necessary to heat the interface, usually in an inert or reducing H_2 atmosphere at a temperature of 450 to 500°C. This thermal bake is also referred to as a low temperature *anneal* or *sinter.* Baking the aluminum on silicon forms a desirable electrical interface referred to as an *ohmic contact.* An ohmic contact has very low resistance (the voltage-current characteristics of the contact interface behave according to Ohm's law). However, there is a small resistance associated with the ohmic contact and it is inversely proportional to the area of the contact. That is, smaller contacts have higher resistance. In modern chip design, ohmic contacts use a special refractory metal (e.g., titanium in the form of a silicide) as a contact at the silicon interface to reduce electrical resistance and improve adhesion (see Figure 12.4 on page 304). Ohmic contacts are fabricated with a salicide (self-aligned silicide) process to achieve optimal positioning above the source/drain and in close proximity to the gate structure. The salicide process is discussed later in this chapter.

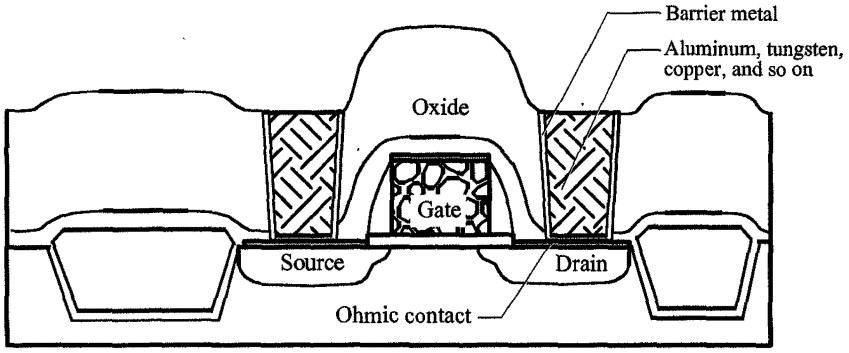


FIGURE 12.4 Ohmic Contact Structure

Given the hundreds of millions of contacts that can exist on a particular die, a reliable contact interface with low resistance and excellent adhesion is extremely important for optimum electrical performance. A failure in any individual contact can cause the entire chip to fail during test or use.

A problem encountered during the earliest work to develop ohmic contacts was undesirable interaction between the aluminum and silicon during the thermal step. This condition led to the microalloying of the contact metal to the silicon, and was referred to as *junction spiking.* Spiking occurs when the interface of pure aluminum and silicon is heated (see Figure 12.5), causing the diffusion of silicon into the aluminum. The amount of silicon dissolved into the aluminum is not uniform and will depend on the time and temperature involved in the heating process. If pure aluminum is heated to 450°C and a source of silicon is present, silicon will begin to dissolve in the aluminum until it reaches a concentration of about 0.5%.[8] The problem is that the source of this silicon is the wafer. As the silicon is dissolved in the aluminum, it leaves behind voids in the wafer, permitting spikes of aluminum to form and penetrate into the silicon contact region. If the aluminum forms an ohmic contact to a shallow junction, the spike may cause a junction short.

The problem of junction spiking was addressed by two methods: the addition of silicon to the aluminum and barrier metallization. The first approach was to use an alloy of aluminum and silicon instead of pure aluminum. If there is already silicon in the aluminum, this will slow down the dissolution of additional silicon from the substrate. However, silicon alloying in aluminum is limited and can lead to silicon nodule formation (small regions of high silicon concentration) in the aluminum due to silicon condensation. Nodule formation can significantly increase the contact resistance and pose a serious reliability concern from local heating at the nodules. The main method to resolve junction spiking was the introduction of barrier metallization to inhibit diffusion (see the following section).

It is important that the oxide at the contact interface be as thin as possible. Wafers with exposed contact regions are often dipped in a dilute hydrofluoric (HF) acid solution immediately before placing the wafers in the deposition chamber. This serves to remove the native oxide layer prior to contact formation.

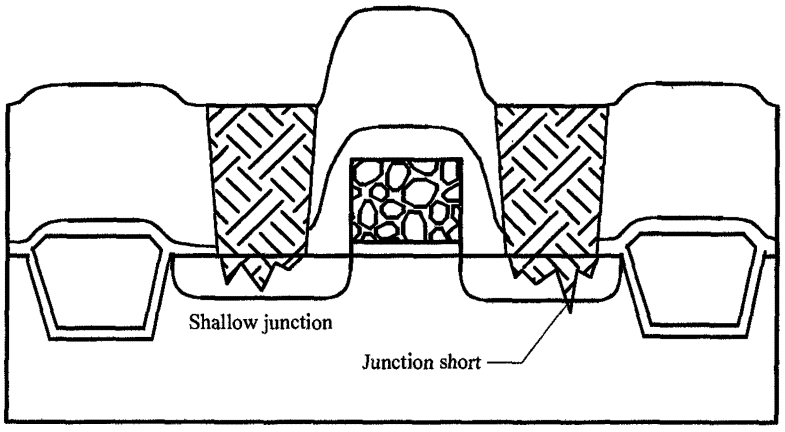


FIGURE 12.5 Junction Spiking

Aluminum-Copper Alloys

Aluminum metal was selected as the primary interconnect material for ICs because of its low resistivity and its compatibility with silicon and the wafer fabrication process. However, aluminum suffers from a reliability problem known as *electromigration.* This is the movement of aluminum atoms in the conductor due to momentum transfer from the electrons carrying the current.[9] Under high current-density conditions, electrons collide with aluminum atoms, causing the atoms to gradually move. This movement of atoms leads to a depletion of atoms at the negative end of the conductor. In conductor regions where depletion occurs, this action leads to voids, thinning of lines, and a potential open circuit. In other regions of the conductor that have an accumulation of metal atoms, metal atoms pile up and form hillocks (see Figure 12.6).[10] *Hillocks* are protrusions on the surface of metal films due to electromigration. If excessive or large hillocks form, adjacent lines or lines on two levels can short together. With advanced circuitry designs in ULSI technology, there is increased current density and chip temperature, both of which make the aluminum chip metallization more prone to electromigration.

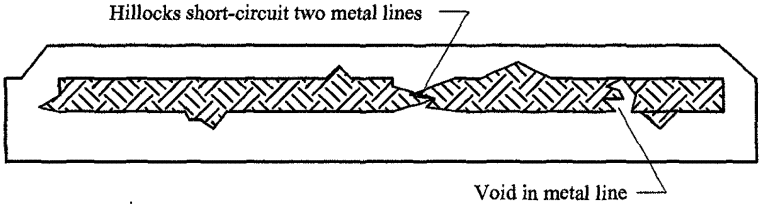


FIGURE 12.6 Hillock on a Metal Line

Electromigration in interconnects is controlled by alloying the aluminum with between 0.5 to 4% copper. This alloying essentially increases the conductor's current-carrying capacity by reducing the grain boundary diffusion effects in the aluminum. At the same time, it has been shown that values of copper in aluminum greater than 8% will actually increase electromigration.[11] The exact reason why the addition of copper reduces the chance of electromigration failure in Al-Cu alloys is not well understood.[12] If copper is alloyed with aluminum, more care must be given to plasma etching this alloyed aluminum (etching is discussed in Chapter 16). Copper is difficult to etch, and any residual copper remaining after etching the aluminum alloy interconnect can promote corrosion.[13] It is also possible to have electromigration in contacts and vias. Electromigration failure in contacts is resolved by the use of barrier metallization.

Electromigration is probably one of the most widely studied failure mechanisms of integrated circuits. Studies of this failure began in the 1950s with the development of the semiconductor and intensified in the late 1960s with the discovery of crack formation in the aluminum conductor leading to failures.[14] The significance of electromigration is that it usually occurs after the chip has been in service for an amount of time, which means that failure occurs catastrophically during customer usage.

Copper

A major transformation is underway in IC design and fabrication by means of the introduction of *copper interconnect* technology. Copper interconnects will replace aluminum metallization with copper to achieve significant benefits in chip performance. A legitimate question that someone may pose is why introduce copper if aluminum has performed so admirably over the years.

Need for Copper ■ The benefits for introducing copper for IC interconnect metallization are:[15]

1. **Reduction in resistivity.** The interconnect wiring resistivity reduces from 2.65 μΩ-cm for aluminum to 1.678 μΩ-cm at 20°C for copper, reducing RC signal delay and increasing chip speed.
2. **Reduction in power consumption.** Narrower lines consume less power.

4. **Tighter packing density.** Narrower lines permit tighter circuitry packing, which means that fewer levels of metal are needed.
5. **Superior resistance to electromigration.** Copper does not have a concern for electromigration.
6. **Fewer process steps.** Potential for 20 to 30% fewer processing steps with damascene processing of copper.

As wafer fabrication design rules reduce to 0.15-μm linewidths and below, the increased packing density of devices on the chip permits more electrical signal speed from device to device (the transistors are closer; therefore, the signal has less distance to travel). This density leads to improved chip performance. However, this improved chip performance is only possible if the interconnect system between the devices is optimized. Narrower linewidths lead to increased line resistance. Tightly spaced conductor lines with a dielectric material between them act as capacitors, leading to a degradation in performance from an increase in resistance (R) and capacitance (C). This condition is the signal delay or interconnect delay discussed in Chapter 11. If either or both of these two parameters are reduced, then the signal delay reduces, leading to increased chip performance.

One possible method reducing interconnect resistance is to increase the conductor cross section. However, this contradicts the goal of increased packing density since wider conductors will take substantially more space. Smaller IC feature sizes are not achievable with larger linewidths. This is why the semiconductor industry is placing emphasis on copper—to lower the interconnect resistance, which lowers R and therefore the overall signal delay. The optimum improvement to RC signal delay is gained when resistance (R) is reduced and the capacitance (C) is lowered by using a low-*k* dielectric along with thinner barrier metals (see Table 12.2).[16]

TABLE 12.2 Change in Interconnect Delay Compared to 0.25-μm Device Generation

Technology	0.25 μm	0.18 μm	0.13 μm
Conventional Interconnect Technology: • Al/Cu interconnect alloy and TiN barrier metal	0	+21%	+93%
New Technology Introduced by Generation: • Reduced barrier thickness • Low-*k* (3.0) dielectric • Dual damascene Cu interconnect and plugs	−10%	−27%	−16%

Another benefit from the implementation of copper is that smaller linewidths can carry the same amount of current, permitting a tighter packing density on each metal level. This condition produces fewer overall levels of metal on a chip, leading to significantly reduced manufacturing costs.[17]

Copper is a relatively soft metal. It has superior resistance to electromigration, a common reliability problem with aluminum.[18] This means that chips fabricated with copper can handle higher electrical power densities, which permits the development of new product applications. Table 12.3 compares Cu and Al for different properties and processes common in wafer fabrication.[19]

TABLE 12.3 Comparison of Properties/Processes Between Al and Cu

Property/Process	Al	Cu
Resistivity (μΩ-cm)	2.65 (3.2 for Al—0.5% Cu)	1.678
Electromigration resistance	Low	High
Corrosion resistance (in air)	High	Low
Etch processing	Yes	No
CMP (chemical mechanical planarization) processing	Yes	Yes

Copper Challenges ■ There are three major challenges involved with using copper as semiconductor interconnects that distinguish it from traditional aluminum interconnects. These three challenges differ significantly from aluminum technology and must be resolved before the implementation of copper for IC interconnects:[20]

1. Copper diffuses quickly into oxides and silicon. This is a concern if copper diffuses into the active region of the silicon (i.e., source/drain/gate region of the transistor) because it will damage the device by creating junction or oxide leakage.
2. Copper cannot be easily patterned using regular plasma etching techniques (see Chapter 16 for a discussion of etching). Copper dry etching does not produce a volatile by-product during the chemical reaction that is necessary for economical dry etching.
3. Copper oxidizes quickly in air at low temperatures (<200°C) and does not form a protective layer to stop further oxidation.

These challenges are addressed by converting to dual damascene processing with special barrier metals optimized for copper. Damascene processing eliminates the need to etch copper. Furthermore, tungsten plugs are expected to be used as the first level of metal to contact the source, drain, and gate regions.[21] This use of tungsten overcomes the concern of copper contaminating the silicon (referred to as copper poisoning). The tungsten may even be patterned for wiring of the local interconnect (LI). All other metal lines and vias for the multiple metal layers are expected to be copper.

Another major hindrance to converting to copper in semiconductor manufacturing is the natural reluctance to introducing new materials into production. New materials means there are new sources of contamination, new equipment and procedures, unpredictable results, and so on. Change always brings unforeseen problems that pose increased risks to the semiconductor manufacturer. However, these risks are balanced by the benefits from changing the interconnect material in order to improve IC performance.

Barrier Metals

A more effective way to make reliable ohmic contacts to shallow junctions without material diffusion or problems such as junction spiking is to use barrier metallization. A *barrier metal* is a thin layer of deposited metal or metals that is designed to prevent intermixing of the materials above and below the barrier (see Figure 12.7). The thickness of barrier metals has typically been around 100 nm for features sizes in the 0.25-μm generation, which is substantially thinner than the 400 to 600 nm thickness required for 0.35-μm generation of devices. Thickness of barrier metals is projected to decrease to 23 nm or less for the 0.18-μm technology node and beyond.[22]

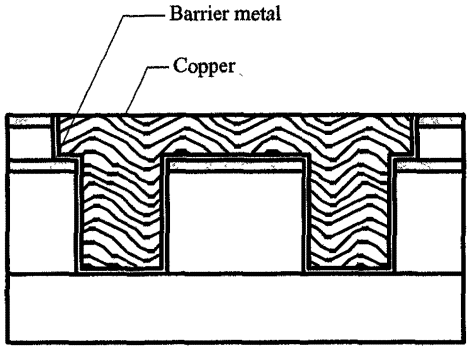


FIGURE 12.7 Barrier Layer for Copper Interconnect Structure

Barrier metals are widely used in semiconductor manufacturing. For contacts between the tungsten (W) plug that connects the aluminum interconnect wiring and the silicon source/drains, the barrier metal prevents the silicon and W from coming into contact with each other. This barrier prevents diffusion between the W and silicon and any junction spiking.

The essential properties of an acceptable barrier layer metal are:

1. Good diffusion barrier properties so that the diffusivity of the two interface materials (e.g., tungsten and silicon) is low at the sintering temperature (*sintering* refers to joining the materials by thermal means).
2. High electrical conductivity with low ohmic contact resistance.
3. Good adhesion between the semiconductor and the metal.

4. Resistance to electromigration.
5. Stability when thin and at high temperature.
6. Resistance to corrosion and oxidation.

Metals commonly used for barrier metals are a class of high melting point metals known as *refractory metals.* Common refractory metals used in multilevel metallization for wafer fabrication are titanium (Ti), tungsten (W), tantalum (Ta), molybdenum (Mo), cobalt (Co), and platinum (Pt). Refractory metals have been used in wafer fabrication since the 1960s, such as in the formation of Schottky barrier diodes in bipolar processes. The benefits of using titanium as a barrier metal for aluminum alloy wiring are improved adhesion, reduced contact resistance, reduced stress, and controlled electromigration. To achieve good barrier properties, wafers undergo a clean step (referred to as *sputter etch* and explained in a later section) before deposition in the processing chamber to remove native oxides and oxide residues on the wafer.

Titanium tungsten (TiW) and titanium nitride (TiN) are two common barrier metal layers that inhibit diffusion between the silicon substrate and aluminum. TiN is widely used in ULSI manufacturing for its superior barrier performance in aluminum alloy interconnect processing. TiN functions as a barrier metal for both tungsten and aluminum. TiN is also widely used as an antireflective coating on aluminum to improve the photolithography patterning process (see Chapter 14). However, TiN does not produce good contact resistance to silicon. To resolve this problem, a thin layer of titanium (e.g., several hundred angstroms or less) is typically deposited before the TiN so that it can react with the underlying material as a silicide (see the following section for an explanation of silicide) and lower the contact resistance. The Ti and TiN are typically deposited in a cluster tool to keep oxides from forming between the two layers.

Copper Barrier Metals ■ Barrier metals are critical for copper metallurgy. Cu has high diffusivity in silicon and silicon dioxide, which will destroy device performance. Traditional barrier metals are not sufficient barriers for copper. Cu requires complete encapsulation by a thin-film barrier layer that functions as an adhesion promoter and effective diffusion barrier.[23] There is a trade-off between these two requirements since good adhesion requires some reaction with Cu, whereas a good barrier metal should not react with Cu. The special barrier metal requirements for copper are:[24]

1. Prevent copper diffusion.
2. Low film resistivity.
3. Good adhesion to both dielectric material and copper.
4. Compatible with chemical mechanical planarization (CMP).
5. Metal layer is continuous and conformal with good step coverage and deposition in high aspect ratio gaps.
6. Minimal thickness to allow the copper to occupy the maximum cross-sectional area.

For Cu interconnect metallurgy, tantalum (Ta), tantalum nitride (TaN), and tantalum silicon nitride (TaSiN) are candidate materials for barrier metals (see Figure 12.8). The diffusion barrier must remain thin (about 75 Å) so that it does not affect the resistivity of the high-aspect ratio plug while still acting as a barrier metal. This condition is difficult to maintain as geometries continue to shrink and metals are deposited into high-aspect ratio vias. Research has shown Ta has good barrier and adhesion properties for copper, whereas TiN, a traditional barrier metal for Al/SiO_2 interconnects, has good barrier properties but poorer adhesion.[25] If TaN is used, it is obtained by small amounts of nitrogen doping or by deposition with a tantalum nitride compound. Investigations have also shown tungsten nitride (WN) functions as an effective barrier for Cu metal.[26] There is ongoing research in the development of barrier metals for Cu metallurgy.

Copper barrier layers can be deposited by using high-density plasma CVD (HDPCVD) or ionized metal plasma physical vapor deposition. Ionized metal plasma PVD for tantalum achieves good step coverage. For deposition with high-aspect ratio gaps, then HDPCVD barrier deposition is often the choice.[27]

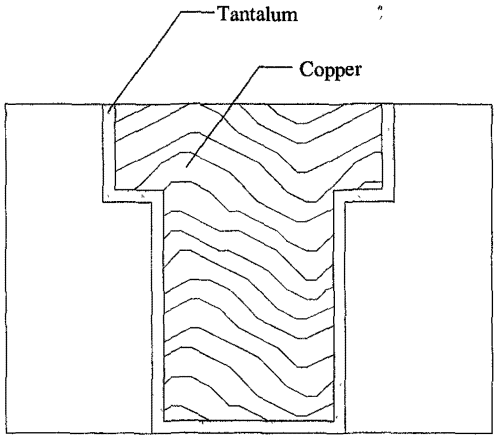


FIGURE 12.8 Ta for Copper Barrier Metal

Silicides

Refractory metals react with silicon when alloyed together to form a silicide. A *silicide* is a metal compound that is thermally stable and provides for low electrical resistivity at the silicon/refractory metal interface (see Figure 12.9). Refractory metal silicides are important in wafer fabrication because of the need to reduce the electrical resistance of the many silicon contacts in the source/drain and gate regions for chip performance. Titanium and cobalt are common refractory metal used for contacts in aluminum interconnect technology.

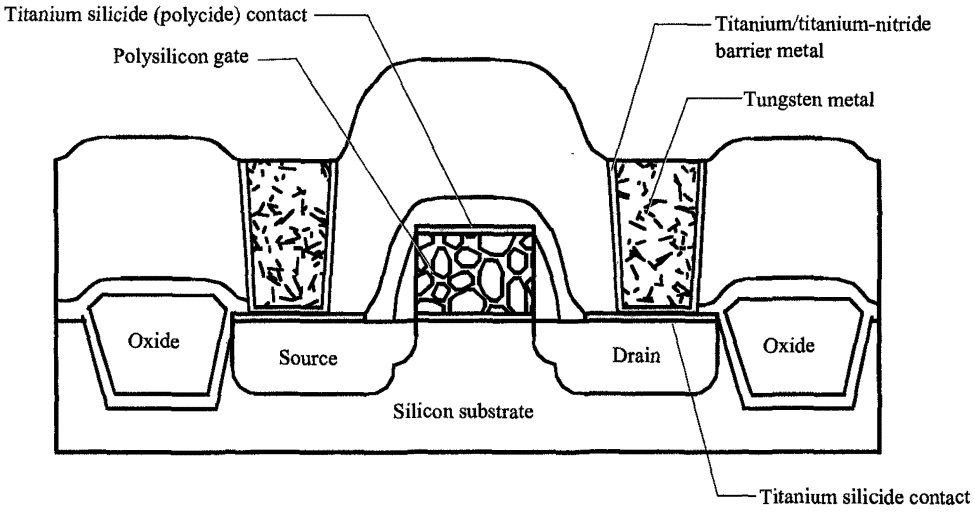


FIGURE 12.9 Refractory Metal Silicide at a Silicon Contact

If the refractory metal is reacted with polysilicon, it is called a *polycide* (see Figure 12.10 on page 310). Doped polysilicon is used as the gate electrode and has a relatively high resistivity (about 500 μΩ-cm), which leads to an undesirable RC signal delay. A polycide is beneficial for reducing the series resistance of an interconnection to polysilicon. At the same time, it maintains polysilicon's good interfacial characteristics to oxides.

Silicides improve contact resistance by reducing residual oxides on the silicon surface during silicide formation. Silicides form extremely good metallurgic contact to the silicon, serving as a critical adhesion layer between the contact metal and the silicon junction regions. Many silicides are stable at temperatures exceeding 1000°C with a relatively high eutectic temperature. Table 12.4 on page 310 lists some properties of common silicides used in wafer fabrication.[28] Recall that the eutectic temperature is the lowest temperature where the alloy melts. It is undesirable for the silicide to melt because the liquid alloy can extend into the substrate material and cause junction spiking.

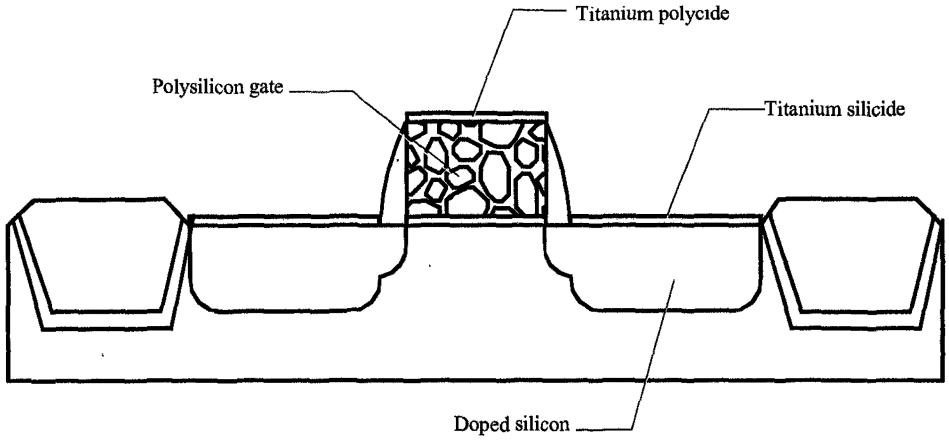


FIGURE 12.10 Polycide on Polysilicon

TABLE 12.4 Some Properties of Select Silicides

Silicide	Lowest Eutectic Temperature (°C)	Typical Forming Temperature* (°C)	Resistivity (μΩ-cm)
Cobalt/silicon ($CoSi_2$)	900	550–700	13–19
Molybdenum/silicon ($MoSi_2$)	1410	900–1100	40–70
Platinum/silicon (PtSi)	830	700–800	28–35
Tantalum/silicon ($TaSi_2$)	1385	900–1100	35–55
Titanium/silicon ($TiSi_2$)	1330	600–800	13–17
Tungsten/silicon (WSi_2)	1440	900–1100	31

* B. El-Kareh, *Fundamentals of Semiconductor Processing Technologies,* (Boston: Kluwer Academic Publishing 1995), p. 537.

Silicide formation typically requires the refractory metal be deposited on the silicon wafer, followed by a high temperature thermal anneal process to produce the silicide material. In areas where there is silicon, the metal reacts to form a silicide. In other areas on the wafer surface, such as where there is SiO_2, there is little or no silicide formation. Often this thermal anneal step uses a rapid thermal anneal (RTA) process in a multichamber cluster tool.

$TiSi_2$ has traditionally been the most common contact silicide for wafer fabrication, serving as the contact between the transistor active regions of silicon and the tungsten plug. It is often referred to as the glue that holds the tungsten to the silicon. Its beneficial properties are high temperature stability, compatibility with self-aligned contact processing (discussed in the following section), low resistivity when compared to other silicides, and its compatibility with TiN barrier metals.[29] $TiSi_2$ forms two different grain phases (a phase is a physically homogeneous material state) as it is annealed, a low-temperature C49 phase and a high-temperature C54 phase (see Figure 12.11). Both of these phases are $TiSi_2$.

The C49 phase of $TiSi_2$ forms at an anneal temperature of 625 to 675°C with a resistivity of 60 to 65 μΩ-cm. The C54 phase is formed in a second anneal following the C49 phase formation and requires a temperature of 800°C or more. It has a much lower resistivity of only 10 to 15 μΩ-cm, which is desirable to lower the overall contact resistance.[30]

However, at the time of this writing, $TiSi_2$ appears limited in its future use as a contact silicide in sub-0.25 μm technologies. Contacts for ultrashallow source/drain junctions are becoming thin. $TiSi_2$ is undesirable for thin contacts because the silicide resistivity increases.[31] It is also difficult to form this contact during the second anneal because of critical time at temperature requirements.

The silicide that appears promising for the 0.18 μm and below technology nodes is cobalt silicide ($CoSi_2$). This silicide maintains a reduced contact resistance of 13 to 19 μΩ-cm after its

	Sintering Temperature	Resistivity
$TiSi_2$ – C49	625 – 675°C	60 – 65 μΩ-cm
$TiSi_2$ – C54	800°C	10 – 15 μΩ-cm

FIGURE 12.11 Anneal Phases of $TiSi_2$

anneal operation, even with deep submicron geometries of 0.18 μm and less. This reduced resistivity occurs because the grain size of $CoSi_2$ is about ten times smaller than the grain size of $TiSi_2$. Therefore, the low resistance phase is completely nucleated and grown during the thermal anneal operation. A $CoSi_2$ contact is also easier to form because of the smaller grain size.

Note that a silicide is not a barrier metal. In some silicides it is found that silicon diffuses rapidly through them. This diffusion occurs during heat treatment of the metal-silicide-silicon system where silicon diffuses through the silicide to the metal, which then reduces the integrity of the system. The solution to this problem is to place a metal barrier layer between the silicide and the metallization layer (discussed in the preceding section). A common silicide barrier film is titanium nitride (TiN), which is effective for both tungsten and aluminum. A tantalum-based barrier layer is anticipated for copper metallurgy.

Salicide ■ Because of the need to scale device sizes in VLSI and ULSI to optimize performance, the cross-sectional area of the source/drain (S/D) contact between the silicon and the first metal layer is small. This small size leads to increased resistance. A technique to decrease contact resistance at the source and drain areas by providing stable contact structures is referred to as a salicide. A *salicide* approach (taken from the expression *self-aligned silicide*) is used to create silicides that are properly aligned with the exposed silicon of the source, drain, and polysilicon gate. There are numerous chip performance issues that depend on salicide formation (see Figure 12.12).

The basic salicide steps are shown in Figure 12.13 on page 312. This process flow corresponds to Step 6 of the CMOS process flow described in Chapter 9. To form a salicide, oxide has been previously deposited and etched back with a dry plasma etch to leave oxide sidewall spacers

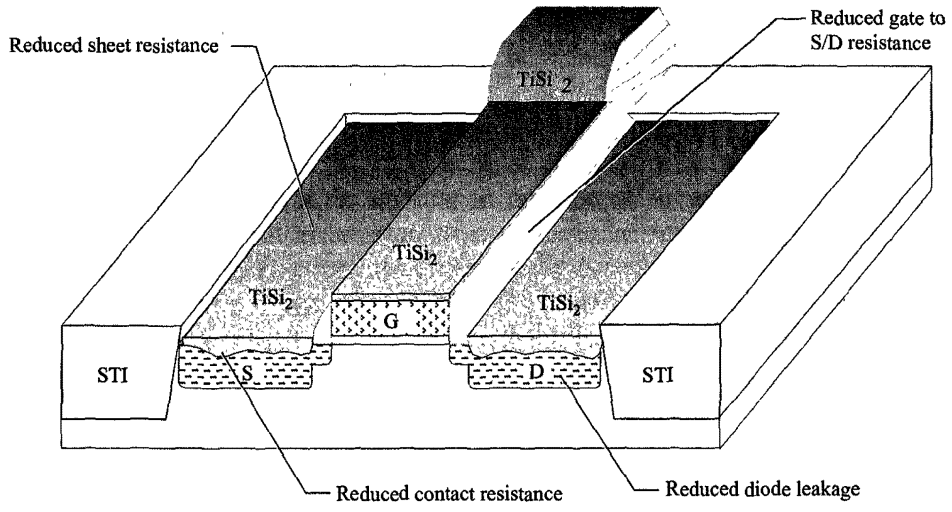


FIGURE 12.12 Chip Performance Issues Related to a Salicide Structure

on the side of the polysilicon gate (Step 5 of the CMOS process flow described in Chapter 9). With the sidewall spacers, only the top poly surface of the gate is exposed. After an HF-dip clean step to remove native oxides, a 250 to 350 Å thin layer of titanium metal is deposited on the wafer. The refractory metal undergoes an RTP rapid anneal at 600 to 800°C to create the high-resistivity C49 titanium silicide phase wherever the refractory metal contacts silicon. The sidewall spacer keeps the S/D $TiSi_2$ from shorting out to the sides of the gate poly, while the oxide of the shallow trench isolation separates the devices. After the first RTP anneal, all unreacted titanium is removed by a wet chemical etch in ammonium hydroxide (NH_4OH) and hydrogen peroxide (H_2O_2). $TiSi_2$ remains and covers the S/D areas and top surface of the poly gate. A second RTP silicide anneal at 800 to 900°C creates the low-resistivity C54 metallic silicide. A major benefit of the salicide process is the avoidance of alignment tolerances that would occur if patterning were required to align the refractory metal on the silicon contact. Care must be taken during the anneal steps to avoid oxygen contamination in the furnace tube. Titanium reacts easily with oxygen to form undesirable titanium oxides. This oxide contamination can promote silicide formation on top of the oxide spacer regions, leading to a short between the polysilicon gate and source or drain.

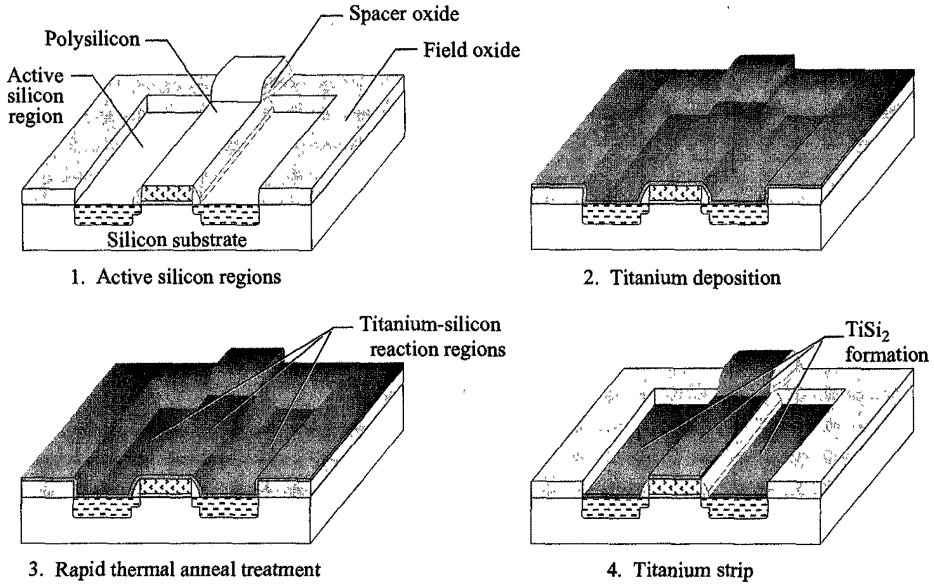


FIGURE 12.13 Formation of Self-Aligned Metal Silicide (Salicide)

Metal Plugs

Multilevel metallization creates the need for billions of vias filled with metal plugs to form electrical pathways between two metal layers. A contact plug is also used to connect the silicon devices in the wafer to the first level of metallization. The most common metal currently used for plugs is tungsten (W), which is why plugs are often referred to as *tungsten plugs* (see Figure 12.14). Tungsten has been traditionally used as a plug material because of its ability to uniformly fill the high-aspect ratio vias when deposited by chemical vapor deposition (CVD) methods. Tungsten is resistant to electromigration failure. It also serves as a barrier to inhibit diffusion and reaction between Si and the first metal layer. Tungsten is a refractory material with a melting point of 3,417°C and a bulk resistivity of 52.8 μΩ-cm at 20°C.

Aluminum would be desirable as a plug material because of its lower resistivity (2.65 μΩ-cm at 20°C), but sputtered aluminum cannot fill high-aspect ratio vias (see the next section for a description of sputtering). For this reason aluminum is used as the interconnect material and tungsten is limited to the plug. There has been recent interest in reflow aluminum plugs, where aluminum is sputter deposited into the via and then reflowed by a high-temperature anneal with a rapid thermal processor (RTP).[32]

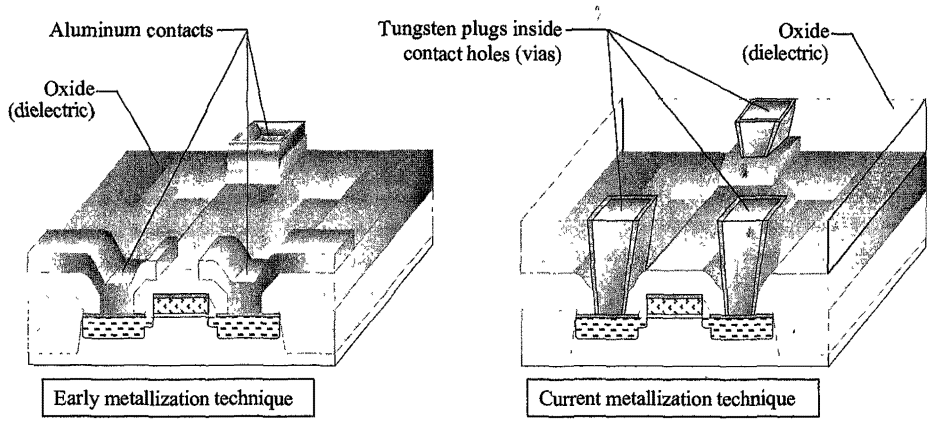


FIGURE 12.14 Tungsten Plug for Multilevel Metal Layers

METAL DEPOSITION SYSTEMS

The traditional metallization techniques used in semiconductor manufacturing fall under a category called *physical vapor deposition (PVD)*. PVD has been done with filament evaporation, followed by electron beam evaporation and, most recently, sputtering. Chemical vapor deposition (CVD) has recently become a frequently used technique to deposit metal films. Each change in the deposition system has brought improved control of thin film properties and quality.

During the SSI and MSI eras of semiconductor manufacturing, evaporation was the method for metallization. It was replaced by sputtering primarily because evaporation has poor step coverage. Evaporation is still used in research applications and for semiconductors using III–V technologies. It is also used in some specialized areas, such as for C4 bump deposition during packaging (see Chapter 20).

Electroplating technology has been employed in various applications. In recent years it has been used for thin film head metallization in the disk drive industry. However, electroplating is just now being introduced for semiconductor manufacturing as a copper deposition method. It is the early deposition method of choice for creating copper interconnects.

The different metal deposition systems used in traditional and dual damascene metallization are:

- Evaporation
- Sputtering
- Metal CVD
- Copper electroplate

Evaporation

In the early days of semiconductor manufacturing, all metal layers were deposited by the PVD method of *evaporation.* Since the late 1970s, sputtering has replaced evaporation in most silicon wafer technologies in order to gain improved step coverage, gap fill, and yield. It is beneficial to review evaporation to understand how it functions and why the silicon industry changed to sputtering.

The process of evaporation consists of placing the material to be deposited in a crucible and heating it inside a vacuum chamber until it vaporizes (see Figure 12.15 on page 314). The most typical method of heating was the use of an electron beam to heat the metal placed in a crucible. By maintaining a high vacuum in the evaporator, the mean free path of the vapor molecules is increased, and the vapor travels in a straight line in the chamber until it strikes a surface and condenses to form a film.

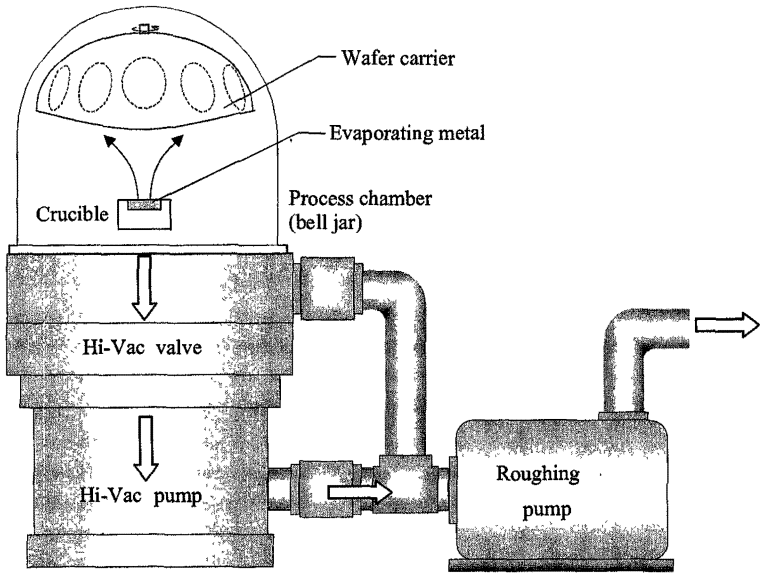


FIGURE 12.15 Simple Evaporator

The biggest drawback to evaporation was the inability to produce uniform step coverage. As the industry progressed to VLSI and ULSI technology, metallization needed to be capable of filling higher aspect ratio holes and producing conformal step coverage. Evaporators made some progress for step coverage by heating the wafers and rotating the wafers in the chamber with a hemispherical cage. However, evaporator technology is not able to form a continuous film over features with an aspect ratio greater than 1.0:1, and is marginal for an aspect ratio of 0.5:1 to 1.0:1. This drawback to evaporation led to its demise in IC production.

Another serious drawback of evaporation is its limitation for depositing alloys. The evaporator requires multiple crucibles for the different materials, which is a problem because of the different vapor pressures of the materials. This limitation makes it difficult to control the ultimate composition of the deposited alloy with any degree of accuracy.

Ultimately, the wafer fabrication industry quickly replaced evaporators with sputtering systems due to their improved step coverage capability. The evaporator is still used in research applications and for some compound semiconductor technologies that can actually use the poor step coverage of evaporators to their advantage during specialized processing. In addition, evaporators are still sometimes used in the chip packaging process to deposit solder bumps on the chip surface (see Chapter 20).

Sputtering

Sputtering, a form of physical vapor deposition (PVD), was discovered in 1852 by Sir William Robert Grove and developed as a thin film deposition technique in the 1920s by Langmuir.[33] As its name implies, sputtering is mainly a physical, rather than chemical, process. In the sputtering process, high-energy particles strike a solid slab of high-purity target material and physically dislodge atoms. These sputtered atoms migrate through a vacuum and eventually deposit on a wafer.

The advantages of sputtering are:[34]

1. Ability to deposit and maintain complex alloys.
2. Capability to deposit high-temperature and refractory metals.
3. Ability to deposit controlled, uniform films on large wafers (200 mm and larger).
4. Capability of multichamber cluster tools to clean the wafer surface for contamination and native oxides before depositing metal (referred to as *in situ sputter etch*).

While sputtering was a big improvement in gap fill over the earlier metallization method of evaporation, its limited capability for step coverage and filling high-aspect ratio gaps makes it insufficient for ULSI applications.[35] Over the years, improvements have been made to the sputtering

process to obtain better step coverage, including recent advances with ionized metal plasma PVD. For critical applications, such as the step coverage in tungsten plugs, deposition will usually be done with a CVD metal process. Sputtering continues to be used for depositing critical barrier and seed layers, such as tantalum/tantalum nitride used in copper metallization (when the aspect ratio is reasonable).

Basic Sputtering Steps ■ There are six basic steps to sputtering:[36]

1. Positive argon ions are generated in a plasma in a high vacuum chamber and accelerated toward a target material at a negative potential.
2. During acceleration, the ions gain momentum and strike the target.
3. The ions physically dislodge (sputter) atoms from the target, which has the desired material composition.
4. The dislodged (sputtered) atoms migrate to the wafer surface.
5. The sputtered atoms condense and form a thin film on the wafer surface with essentially the same material composition as the target, following the stages for thin film growth outlined in Chapter 11.
6. Excess material is removed from the chamber by vacuum pump.

To illustrate sputtering, Figure 12.16 shows the basic configuration of a simple, DC diode sputter deposition chamber with parallel plates. It consists of the solid *target* material, a substrate (wafer), and a vacuum enclosure. The target is electrically grounded and referred to as the *cathode,* while the substrate has a positive potential and is termed the *anode.* The anode and cathode are both called *electrodes.* The sputtering target is composed of the necessary material for deposition. Targets are manufactured to achieve a homogeneous composition, fine grain size, and specific crystallographic orientation, all of which contribute to achieving a uniform film deposition rate across the wafer.[37] Target purity levels of 99.999% (5 nines) or better are required for achieving acceptable film purity for sub-0.25 μm geometries.

A high density of positive ions from an argon gas glow discharge is strongly attracted to the negative target material, striking it at high velocity and dislodging the atoms to be deposited. The atoms are sputtered (knocked off) from the target material and scatter in the chamber; eventually some come to rest on the wafer or the chamber walls. This action necessitates cleaning of the chamber in some systems. The atoms on the wafer nucleate and grow a thin film. An important aspect of sputtering is that this process can be used to sputter-deposit alloys, in particular aluminum-silicon

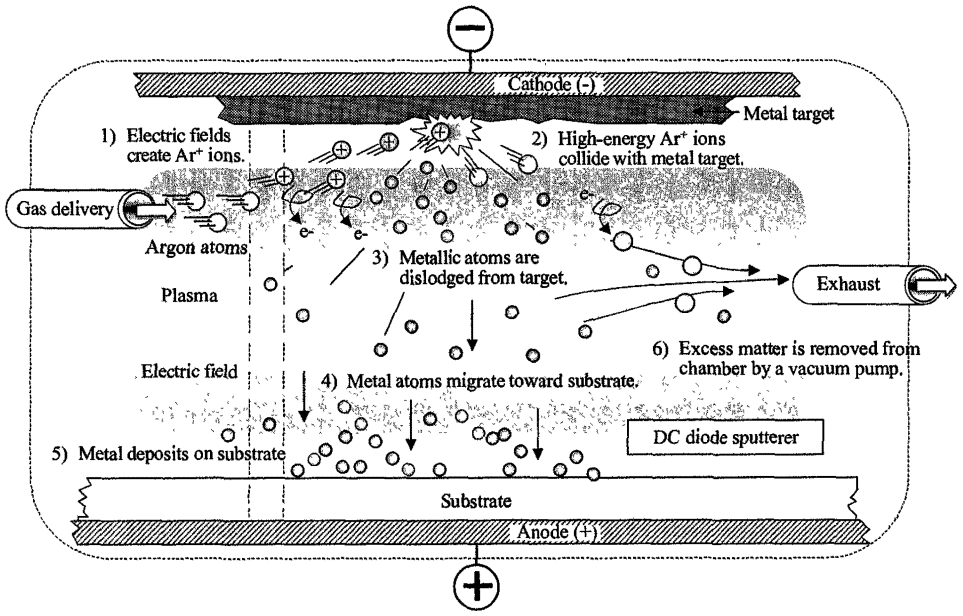


FIGURE 12.16 Simple Parallel Plate DC Diode Sputtering System

and aluminum-copper alloys. For instance, if the target material is aluminum with 1% copper, then the deposited film will be aluminum with 1% copper (at steady-state conditions).

The latest technology of sputtering equipment systems are multichamber cluster tools. The system has a loadlock that moves single wafers from the wafer cassette to its ultimate base vacuum pressure. The loadlock may consist of dual loadlocks to move the wafer through a staged vacuum sequence from ambient atmosphere to base vacuum pressure. A robot transfer system designed specifically for the cluster tool environment moves wafers from the various process chambers and a buffer chamber. It is important that the robot operate without generating particles and be extremely reliable. Cluster tool robots are often magnetically coupled, which permits the robot drive motor to remain on the outside of the process chambers to reduce particles.

Vacuum conditions are important in the sputtering chamber to create the plasma and maintain purity of the deposited films, with base vacuum typically at 10^{-7} torr during initial pump down. The rate that argon gas enters the process chamber is critical because it causes the pressure to rise in the chamber. With the argon and sputtered material in the chamber, the pressure rises to about 10^{-3} torr. The hot, high-vacuum environment of the process chamber promotes the formation of the sputtered atoms into thin films on the wafer surface.

Physics of Sputtering ■ An essential aspect of sputtering is the formation of the ionized argon gas into a plasma (see Chapter 8). Argon is used as the sputtering ion species because it is relatively heavy and is a chemically inert gas, which keeps it from reacting with the growing film or the target. If a high-energy electron strikes the neutral argon, the collision knocks off outer electrons and creates a positively charged argon ion. This energetic particle is used to strike the negatively charged target material to be sputtered.

Sputtering Mechanism. Positively charged argon ions in the plasma are strongly attracted to the negative potential of the cathode target. The charged ions accelerate and acquire kinetic energy (or energy of motion) as they pass through the voltage drop of the glow-discharge dark space (see Chapter 8 for a description of glow discharge). When the Ar ions strike the target surface, the momentum of the Ar ion transfers to the target material to dislodge one or more atoms. This action is referred to as *momentum transfer.* The ejected neutral atom or atoms move through the plasma (with a small chance of being ionized) and land on the wafer. The incident ion energy must be large enough to dislodge target atoms but not so large that the ion penetrates into the target material. Typical sputtering ion energies range from 500 to 5,000 eV.

An analogy for dislodging metal atoms from the surface of the target material during sputtering is hitting billiard balls in a pool game. Even though the cue ball in pool is travelling in one direction, the billiard balls can be knocked in some other direction. This same effect occurs in sputtering, where the argon ion impacts the target and dislodges one or more atoms from the surface of the target (see Figure 12.17).

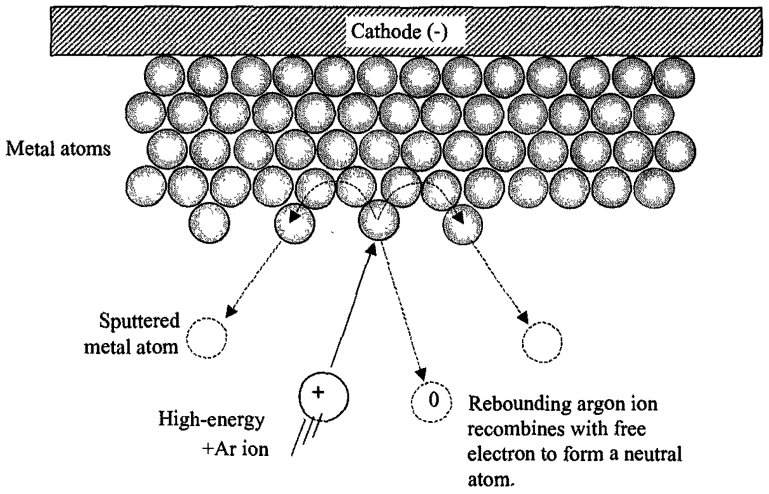


FIGURE 12.17 Dislodging Metal Atoms from Surface of Sputtering Target

The *sputtering yield* is defined as the number of atoms ejected by the target (cathode) per incident ion that strikes it. The yield largely determines the rate of sputter deposition. It varies from about 0.5 to 1.5. A sputter yield of 0.5 means that, on the average, two ions must strike the target for one atom to be ejected. The sputter yield depends on the following conditions:[38]

1. Incident angle of the bombarding ions.
2. Composition and geometry of the target material.
3. Mass of bombarding ions.
4. Energy of the bombarding ions.

A method of increasing the sputtering deposition rate is to confine the plasma to the region between the target and wafers. Due to the ionization process, there are dark spaces just to the front and the sides of the target. Dark space shields are placed on the sides of the target to prevent target material from being sputtered on the sides, since this material will never deposit on the wafers. Shields require periodic changing because sputtered target material builds up on the surface and causes particulate contamination.

Targets undergo erosion from the ion bombardment and are replaced when about 50% or more of the target material is removed.[39] Much of the energy consumed during the sputtering process is dissipated as heat in the target, or by emission of secondary electrons and photons by the target. For this reason, cooling of the target material is done to maintain a low target temperature.

In addition to being struck by the sputtered atoms, the substrate has other species landing on it (see Figure 12.18). These species cause heating of the substrate (up to 350°C) which leads to uneven film deposition. The high temperature can also create an undesirable aluminum oxide during aluminum deposition, which interferes with the sputtering process. The many species impinging on the wafer surface during diode sputtering also increase the possibility of damage due to radiation to sensitive devices.

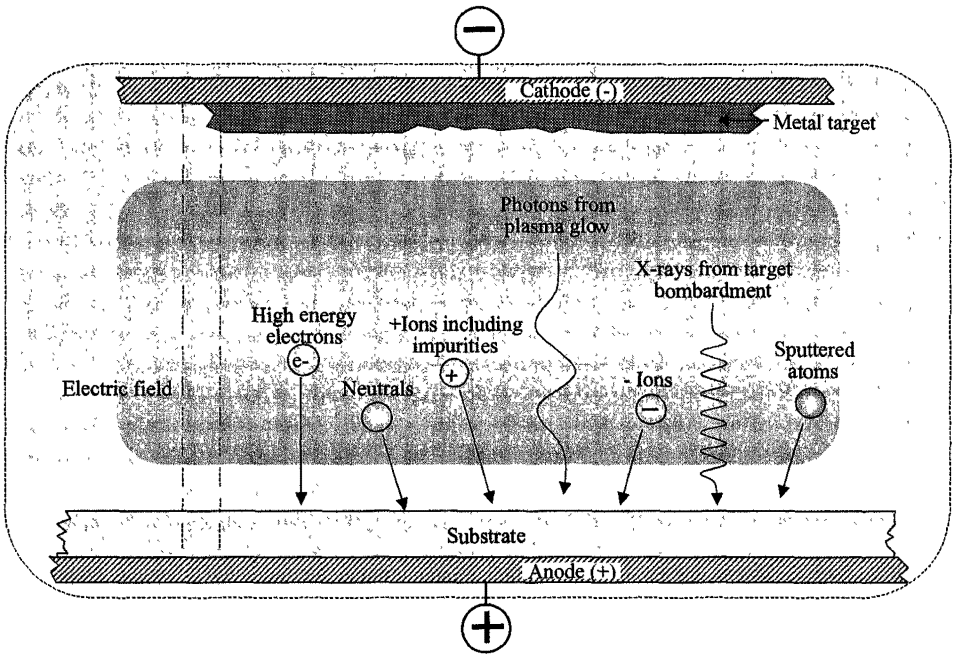


FIGURE 12.18 Different Species Landing on Substrate

Note the presence of the impurity gas atoms during diode sputtering. This is a problem if the impurity atoms are incorporated into the growing film on the substrate. The sources of contaminants are: (1) impure sputtering gas, (2) outgassing from the wafer holder, (3) outgassing from chamber walls, and (4) vacuum leaks.[40] It is necessary to use ultrahigh purity (UHP) sputter gases and to verify there are no chamber leaks.

The spacing between the cathode and anode must be optimized for each sputtering system. The goal is to have as many sputtered atoms as possible be deposited on the wafers and form the desired film.

The sputtering system described thus far is a simple DC diode system that has a DC voltage applied between two electrodes (cathode and anode). The DC diode sputtering system has serious drawbacks that limit its use in manufacturing. A DC diode type sputtering system cannot be used to sputter dielectrics because the electrodes become coated with the dielectric and the glow discharge cannot be sustained. The target rapidly builds up a positive charge, which repels incoming positive ions. DC diode type sputtering is also not capable of performing a sputter etch. *Sputter etch* (or *reverse sputter*) is a preclean step where the sputter process is reversed and argon atoms are used to remove thin native-oxide layers and remaining etch residues that contaminate contacts and vias. In other words, the wafer is sputtered instead of the target. This sputter etch preclean is important in multichamber cluster tools because of the benefit of cleaning followed by deposition without removing the wafer from the vacuum environment.

Three types of sputtering systems are covered in the following pages:

- RF (radio frequency)
- Magnetron
- IMP (ionized metal plasma)

The simple RF sputter system is not used in wafer fabrication because of its inherent inefficiency. The magnetron has traditionally been the most widely used sputter system, and ionized metal plasma (IMP) is becoming more common for sub-0.25 micron technology.

RF Sputtering ■ In an *RF sputtering* system, an RF field is used to create the plasma instead of the DC field described above. The RF frequency has commonly been 13.56 MHz. It is applied to the back surface of the target electrode and capacitively coupled to its front surface (see Figure 12.19). Both the electrons and ions in the plasma are exposed to the RF field, but due to the high frequency, the electrons respond most strongly. The chamber and electrode behave like a diode, creating a high amount of electron flow and resulting in a negative charge on the target electrode. This negative charge (produced by a self-bias) attracts positive argon ions, which sputter material from the insulator or non-insulator target.[41]

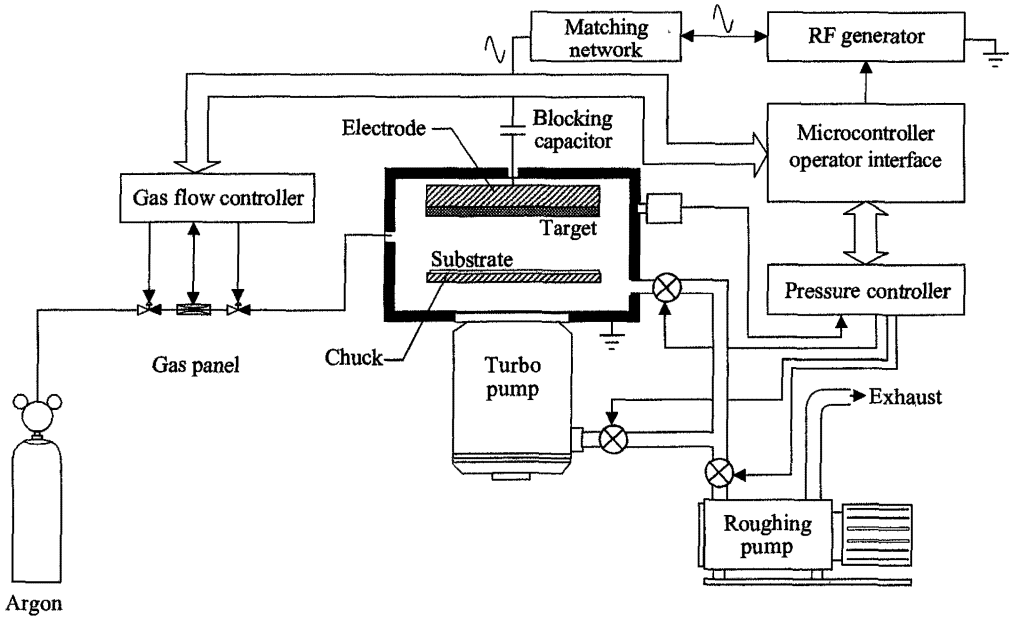


FIGURE 12.19 RF Sputtering System

The wafer can be electrically biased to put it at a different field potential than the argon. The bias applied to the wafer causes the argon atoms to strike the wafer directly. This RF biasing permits the exposed wafer surface to be sputter etched and cleaned.

In practice, an RF sputtering system is limited because it does not have a high sputtering yield, leading to a low deposition rate. Many of the secondary electrons emitted by the target pass through the discharge region without contributing to the creation of plasma. The sputtering rate of the target would be much higher if these secondary electrons were involved in ion collisions, leading to the creation of more ions to bombard the target. To overcome this inefficiency and to achieve high metal deposition rates necessary for wafer fabrication, the concept of magnetron sputtering was developed.

Magnetron Sputtering ■ *Magnetron sputtering* employs magnets configured around and behind the target to capture and restrict the electrons in front of the target (see Figure 12.20). This setup increases the ion bombardment rate on the target in order to produce more secondary electrons, which then increases the ionization rate in the plasma. The end result is that more ions cause more sputtering of the target, which increases the deposition rate of the system. The invention of magnetron sputtering to improve the deposition rate is one of the main reasons why sputtering has become the leading process for single-wafer deposition systems for aluminum and contact alloys used in metallization.

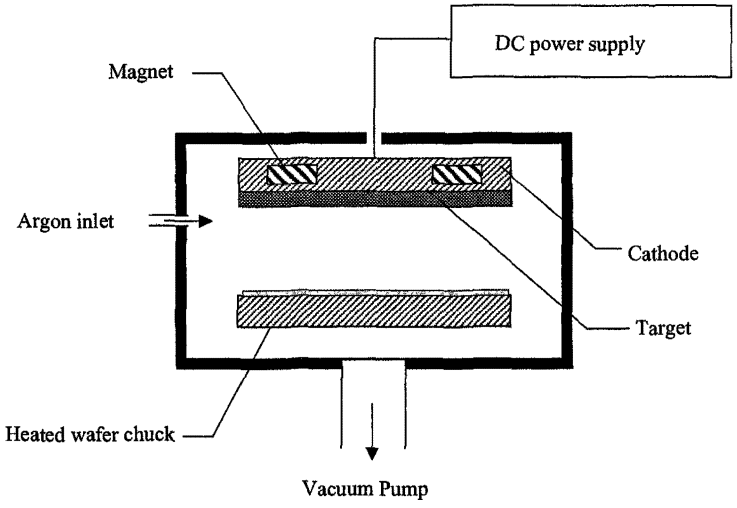


FIGURE 12.20 Magnetron Sputtering

The magnetron sputtering design requires a substantial amount of power supplied to the argon plasma to maximize the sputtering rate (about 3kW to 20 kW of power). Because the sputtering target absorbs most of this power and is in contact with the cathode, cooling of the cathode is required. An important challenge is uniformity of the sputtered deposit on large wafers. To achieve high deposition rates and film uniformity in single-wafer cluster tools, new cathodes have been developed with rotating, rare earth, high-strength permanent magnets.[42]

Step Coverage. Sputtering requires a vacuum environment with high-purity argon to avoid contamination from residual gas. The vacuum during the sputtering process is around 1 mtorr, with a mean free path of several centimeters. This is approximately the distance between the target and wafer. Because of this distance, atoms dislodged from the target essentially pass on to the wafer by following a line-of-sight path.

This line-of-sight path for the ejected atoms exiting at many different angles from the target causes poor coverage on steps and sidewalls in contacts and vias. Due to their geometry, the sidewalls and bottoms of high-aspect ratio contact windows and vias will receive only 10% or less of the metal deposited on the top surface.[43] For this reason PVD is not usually chosen when depositing a material on high aspect ratio steps and trenches.

Collimated Sputtering. To achieve increased coverage on the bottom and side of a contact or via, directional enhancement can be achieved by the use of *collimated sputtering* (see Figure 12.21). The collimator is configured so that it appears as the electrical ground for the plasma. In this manner, any neutral species sputtered from the target at a high angle is intercepted and deposited on the collimator. Other atoms that are ejected straight back from the target will pass by the collimator and deposit on the bottom of the contact hole. At the same time, a collimator reduces sidewall coverage in the contact or via.

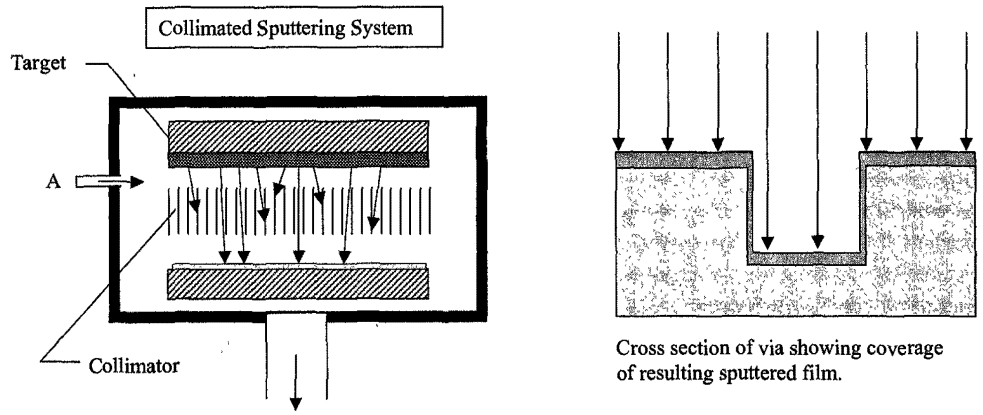


FIGURE 12.21 Collimated Sputtering

The use of a collimator means that a large portion of the sputtered material will not reach the wafer, since most of the sputtered material ends up on the collimator. This result lowers the sputter yield and increases the cost of deposition, since fewer sputtered atoms reach the wafer. At the same time, maintenance costs increase because the target is less efficient. Given these problems, it is more efficient to use ionized metal plasma (IMP) or CVD deposition over magnetron sputtering if step coverage is a critical factor.

Ionized Metal Plasma (IMP) ■ A problem for sputtering technology in wafer fabrication for high performance ICs is that as feature sizes shrink, the capability to sputter into high aspect ratio vias and narrow trenches becomes limited. A recent development in PVD to overcome this problem is *ionized metal plasma PVD (IMP* or *ionized PVD),* which was introduced in the mid-1990s. In this approach, the sputtered metal is ionized in an RF plasma at a pressure of 20 to 40 mtorr (see Figure 12.22). The positive metal ions then travel in a highly directional, vertical path toward a wafer configured with a negative voltage bias. This voltage bias can also be used to control the energy of the incoming metal ions, which helps reduce damage to the surface of the wafer.[44] Biasing the wafer enables a higher degree of film conformality in the bottom and corners of high aspect ratio gaps.

Ionized PVD achieves good hole-fill for Ti and TiN on structures with 0.25-micron contacts and vias and 6:1 aspect ratios, with 70% bottom coverage, 10% sidewall coverage, and excellent sidewall integrity. This coverage is critical in the bottom, bottom corner, and lower sidewall areas of holes with a typical 85° sidewall angle.

Ionized PVD is used in production to deposit titanium and titanium nitride barrier films for devices at the 0.25 μm and below generations with high-aspect ratio contact and via structures.[45] It is also capable of depositing tantalum, tantalum nitride, and copper, which are important metals for copper metallization. The development of ionized PVD has further improved sputtering to extend its usefulness in high-aspect ratio gap fill.

Metal CVD

The use of chemical vapor deposition (CVD) for metal deposition is increasing because of its superior conformal step coverage and void-free filling of high-aspect ratio contacts and vias. These factors are critical for wafer fabrication as feature sizes decrease to the 0.15-μm generation and

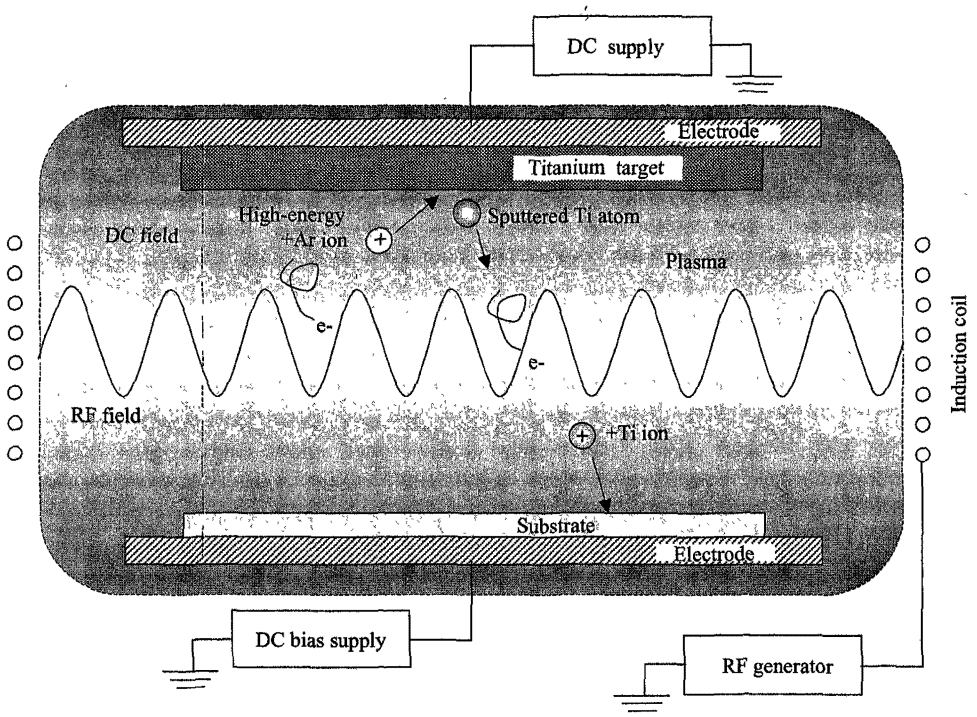


FIGURE 12.22 Concept of Ionized Metal Plasma PVD

beyond. The projected aspect ratio for vias in DRAM memory is projected to be 7:1 for the 0.15-μm generation, and 2.4:1 for logic.[46]

Tungsten CVD ■ *Tungsten (W) CVD* is used in multilevel interconnection technology as the process to deposit the via plug and to act as the contact plug between the first metal layer and the silicide contact. Tungsten CVD is also used as the local interconnect (LI) because of its low resistivity. Tungsten has performed reliably, with billions of W CVD plugs used on one microchip. However, its usefulness as a vertical interconnect may be limited with the introduction of the new damascene copper metallization scheme.

Tungsten CVD is common as a plug material due to:

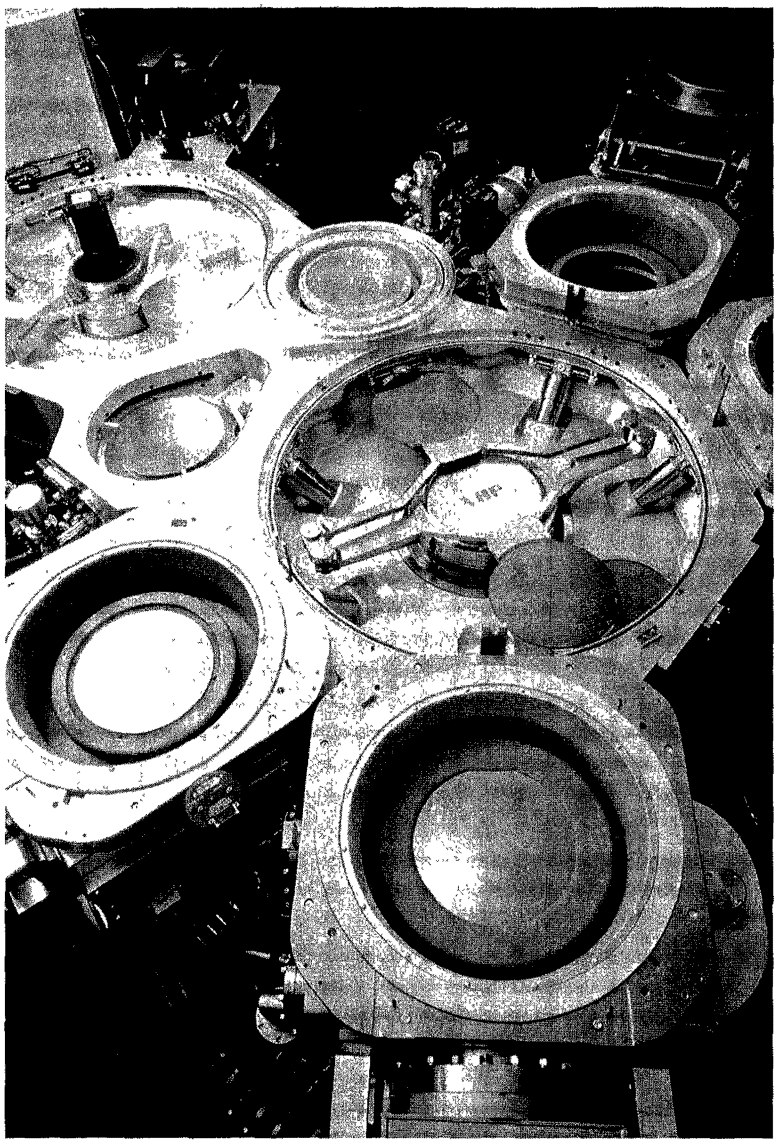
1. Excellent step coverage and gap-fill, particularly in high-aspect ratio vias.
2. High electromigration resistance.

Low-pressure CVD (LPCVD) is the usual method for depositing W plugs, along with plasma-enhanced CVD (PECVD). Reactors are both hot- and cold-wall and configured as batch or multichamber cluster tools. The most common source gas for W is WF_6.[47] Sputtering costs less to operate than CVD for W deposition but has traditionally poorer directional control. This coverage produces nonuniform tungsten deposition quality in vias, which is why CVD has been the preferred method. This may change with the recent development of ionized PVD.

Tungsten CVD is typically deposited in blanket films. Blanket deposition deposits tungsten nonselectively on the entire wafer surface, including the via pathways, and is the most common approach.

Blanket Tungsten CVD Deposition. Blanket tungsten deposition followed by chemical mechanical planarization (CMP) is the common method used to fill vias to connect between metal layers. As always in CVD, the properties of the film deposited with CVD depend on the chemical reactions that occur at the wafer's surface. For blanket tungsten CVD, the most frequently used reaction is the reduction of tungsten hexafluoride (WF_6) by hydrogen to produce tungsten and hydrogen fluoride:

$$\underset{\textit{tungsten hexafluoride}}{WF_6\,(\text{gas})} + \underset{\textit{hydrogen}}{3H_2\,(\text{gas})} \rightarrow \underset{\textit{tungsten}}{W\,(\text{solid})} + \underset{\textit{hydrogen fluoride}}{6HF\,(\text{gas})}$$



PVD Cluster Tool
(Photo courtesy of Applied Materials, Inc.)

Typically the first step in blanket tungsten CVD is to deposit a titanium/titanium nitride (Ti/TiN) barrier layer (see Figure 12.23). The Ti is deposited before TiN to react with the underlying material and lower the contact resistance. If this is a contact in the first interlayer dielectric (ILD-1), then Ti reacts with silicon to form $TiSi_2$ silicide. The TiN functions as a barrier metal and adhesion promoter (glue layer) for tungsten. It needs a minimum bottom thickness of about 50 Å with continuous sidewall coverage to be an effective barrier metal and keep the tungsten from attacking underlying layers.[48] The Ti/TiN barrier layer may be deposited by IMP PVD. The TiN barrier metal may also be deposited either by CVD using ammonia and one of a number of precursor molecules, including the tetrakis dimethylamino titanium (TDMAT) or tetrakis diethylamino titanium (TDEAT). CVD has usually been preferred for TiN because of improved step coverage.

It is necessary to remove the excess blanket tungsten above the dielectric layer. This was previously done by a tungsten etchback process to remove the excess tungsten and leave a planarized plug. In 0.25-μm device generations and below, planarization of tungsten by chemical mechanical planarization (CMP) is the preferred process (CMP and etchback are addressed in Chapter 18).[49] Barrier metals such as TiN and Ti are used to prevent diffusion between the tungsten and the silicon.

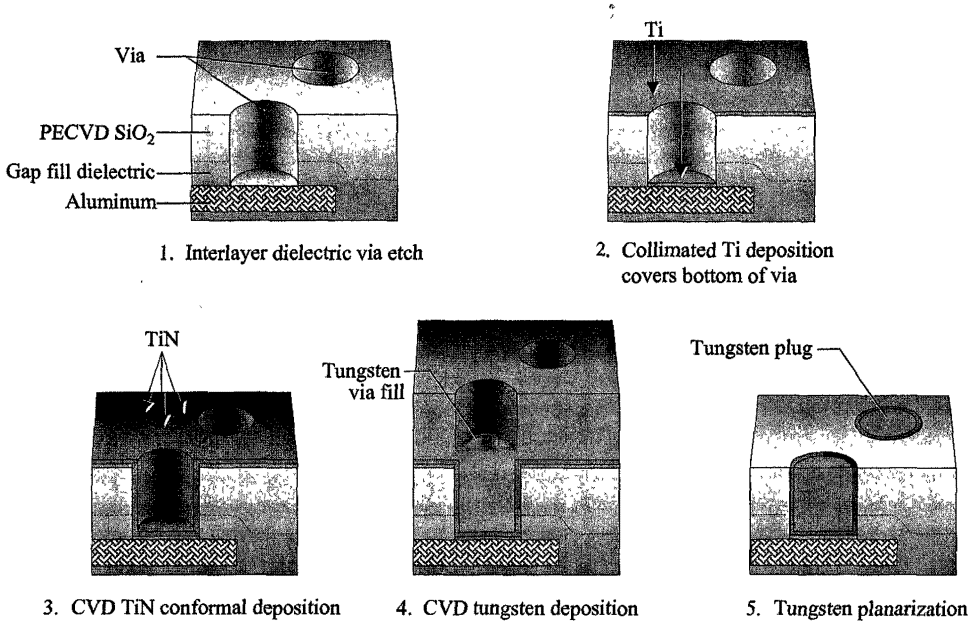


FIGURE 12.23 Blanket Tungsten CVD with Ti/TiN Barrier Metal

Copper CVD ■ *Copper (Cu) CVD* is a potential process for depositing a thin copper seed layer necessary for copper electroplating. The seed layer, or strike layer, is thin (about 500 to 1000 Å) and deposited on top of a diffusion barrier (most likely a Ta-based barrier metal). For successful electroplating, it is critical that this seed layer be continuous and free of pinholes and voids along the feature sidewalls and bottom. If a seed layer is not continuous, this can create a void in the electroplated copper. The excellent conformality of CVD is why this process has been investigated for the seed layer.

Cu Precursors. There are two precursors for CVD Cu, Cu(I), and Cu(II). Note that Cu(I) indicates a copper ion with a +1 positive charge and Cu(II) indicates a +2 positive charge. The most widely used Cu(I) precursor is Cu(hfac)(TMVS), chemical name trimethylvinylsilylhexafluoroacetylacetonate copper (I), chemical formula $C_{10}H_{13}CuF_6O_2Si$, and referred to by its trade name of CurpraSelect.[50] This molecule combines copper in a +1 oxidation state with TMVS and hfac ligands to make a clear yellow, liquid precursor (a ligand is a molecule or ion surrounding a central metal cation). TMVS is flammable, but the compound Cu(hfac)(TMVS) does not ignite easily. Cu(hfac)(TMVS) is compatible with industrial stainless steel delivery vessels. The CVD deposition proceeds by reducing a copper atom to metal while oxidizing another atom to $Cu(hfac)_2$ and releasing free TMVS as by-products. The chemical reaction is:

$$2Cu^{+1}(hfac)TMVS\ (gas) \rightarrow Cu^0\ (solid) + Cu^{+2}(hfac)_2\ (gas) + 2TMVS\ (gas)$$

The most common Cu(II) precursor is $Cu(hfac)_2$. For example, the Cu(II) reaction can proceed by reduction in H_2:

$$Cu(hfac)_2\ (gas) + H_2\ (gas) \rightarrow Cu^0\ (solid) + 2H(hfac)\ (gas)$$

Copper Electroplate

The transformation of IC fabrication to copper metallization is underway for all chip manufacturers. Initially, high-performance microprocessors and fast, static RAMs are being converted to copper technology. *Copper electroplate,* also known as *electroplating, electrochemical deposition (ECD),* or *electrofill,* is the first-generation deposition method to be used for copper metallization. Cost and performance will be important issues that affect when copper will replace aluminum as the mainstream interconnect metallization. Electroplating has not traditionally been used in semiconductor manufacturing. However, it has been an important metallization process in other areas of electronic manufacturing, such as thin film heads used in disk storage devices and copper wiring for printed circuit boards. Therefore, its process and equipment requirements are well understood.



Copper Electroplating Tool
(Photo courtesy of Novellus)

Electroplating Fundamentals ■ The basic principle of electroplating copper metal is to immerse a wafer with a conductive surface into a solution of copper sulfate ($Cu(SO_4)$) which contains copper ions to be deposited (see Figure 12.24). The wafer and its seed layer are connected electrically to an external power supply as a negatively charged plate, or cathode. A solid piece of copper is immersed in the solution and configured as the positively charged anode. Electrical current passes through the wafer, into the solution, and through the copper anode.[51] When current flows, the following reduction reaction occurs at the wafer surface to deposit copper metal:

$$Cu^{2+} + 2e^{-} \rightarrow Cu^{0}$$

Copper ions *Electrons* *Copper metal*

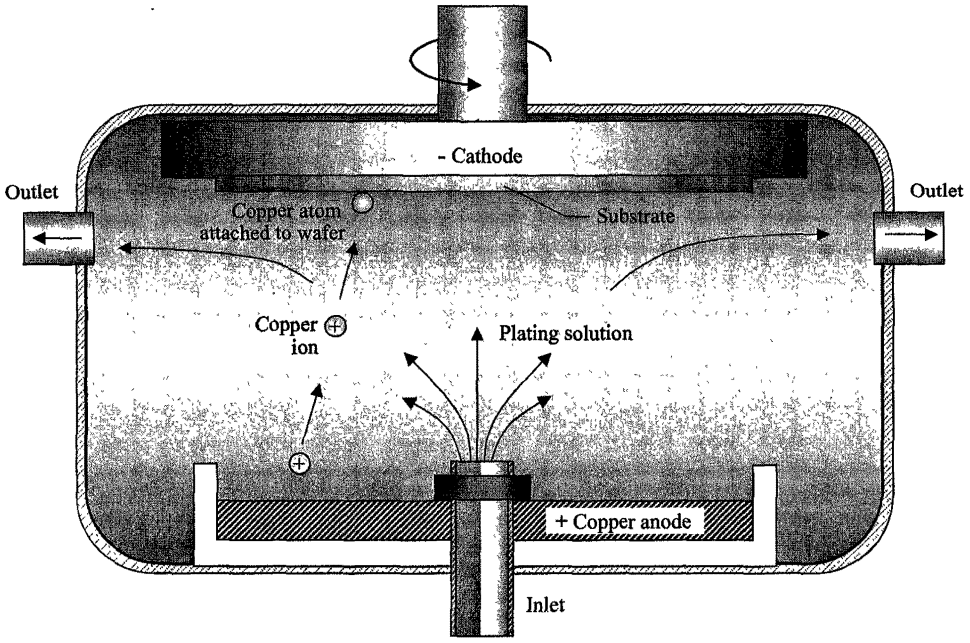


FIGURE 12.24 Copper Electroplating

Metal copper ions on the cathode wafer surface are reduced to metal copper atoms during electroplating. At the same time, an oxidation reaction occurs at the copper anode that balances the current flow in the cathode. This reaction maintains electrical neutrality in the solution.

The quantity of copper deposited is directly proportional to the current delivered to the conductive wafer surface (Faraday's law of electrolysis).[52] With this relationship, the basic parameters to control in electroplating are current and time. With no current flow, there is an equilibrium potential among the anode, cathode, and solution. When an external power source applies a voltage, a current flows between the anode and cathode and metal is deposited on the cathode proportional to the amount of current.

In practice, electroplating is complicated to control, especially the electroplating solution and how the current is applied. Given the high-aspect ratio holes and trenches that must be filled on high-performance ICs, electroplating is challenged to maintain a uniform current density in the holes. This condition is important because the plating rate is a direct function of current density. If there is a high current density at the top of the hole and a lower current density at the bottom of the hole, then copper will plate faster at the top.

Different types of voltage waveforms applied to the cathode/anode system can assist in plating high-aspect ratio holes. For instance, by applying an oscillating electric field and controlling the amplitude of the waveform, a deposit/etch (referred to as dep/etch) sequence is obtained. In this manner, the copper can be slightly removed (etched) in the high-density current regions to balance the copper gap-fill capability.

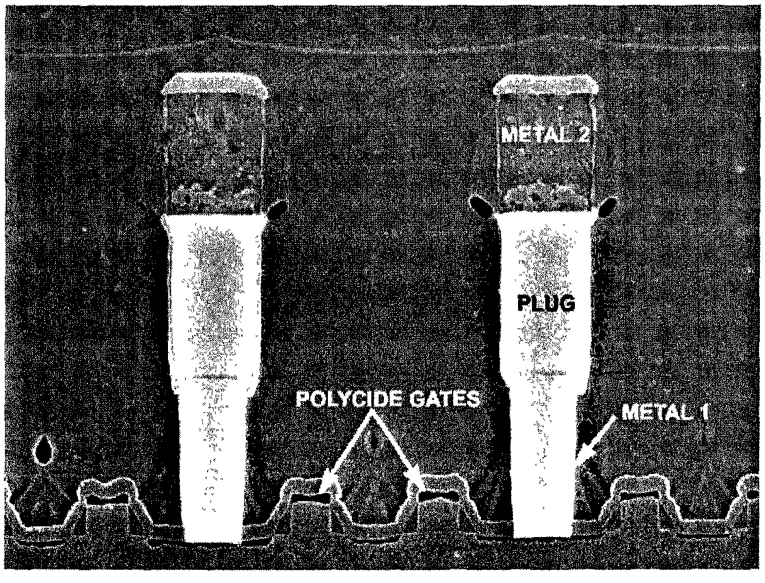
Because the copper electroplating metallization is a new process for semiconductor manufacturing, there are a host of other issues to address. For instance, it is important that no copper plate on the backside of the wafer. Electroplating also poses concerns for chemical control and disposal. It does simplify the metallization process somewhat by avoiding a high vacuum or complex heating of the wafer.

METALLIZATION SCHEMES

The strategy for metal interconnects for IC fabrication is undergoing dramatic change. After many years of development activities, copper is being implemented as the main chip interconnect conductor for a wide range of ICs.[53] Either aluminum or copper metallization is expected to be used for chip applications during a transition period until copper is fully implemented. The type of metallization scheme used depends on how the chip design is optimized for performance and price.

Traditional Aluminum Structure

The traditional interconnect metal has been an aluminum-copper alloy with an interlayer dielectric of SiO_2 to insulate between the metal layers. The process steps for the traditional aluminum



Blanket Aluminum Etch (Micrograph courtesy of Integrated Circuit Engineering)

interconnect technology for an ILD-2 and metal-2 (one complete level) are described in Steps 11 and 12 of the process flow described in Chapter 9. This traditional process can be viewed as an *aluminum subtractive process* because the aluminum is deposited as a blanket film and then etched away (subtracted) to form the circuitry. The critical step that defines the width and spacing of the aluminum interconnect lines is metal etch, while SiO_2 is deposited into the narrow gaps between the lines.

New materials and processes are being rapidly introduced into advanced wafer fabs to support the 0.18-μm technology node with traditional aluminum alloy metallization. This is a transitional period until copper metallization is eventually implemented. As a rule, most fabs will change only one major microchip material at a time (e.g., metal or insulator) to minimize the inherent risk to production. With these new fabrication materials and their stringent processing requirements, the general trend is for increased use of high-density plasma CVD (HDPCVD) for dielectrics, a shift from sputtered metals to metal CVD, and increased use of chemical mechanical planarization (CMP). The goals are to increase product performance and manufacturing productivity.

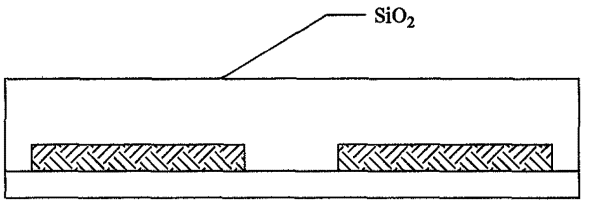
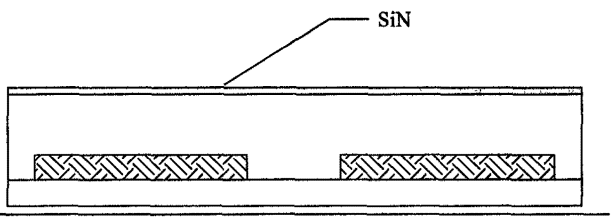
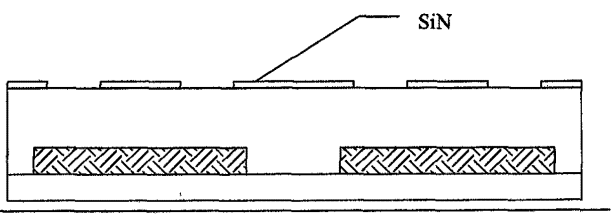
Copper Damascene Structure

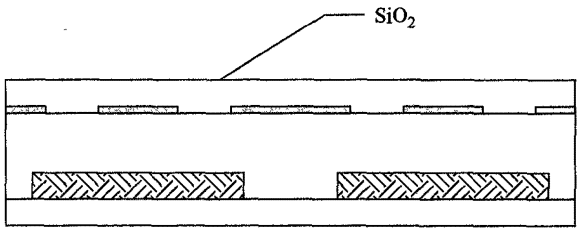
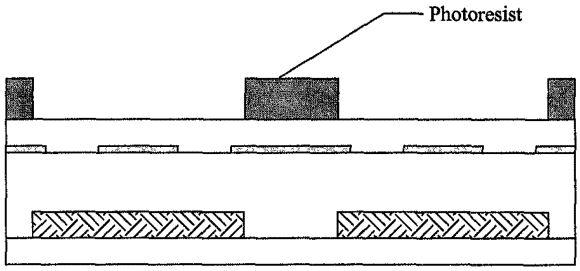
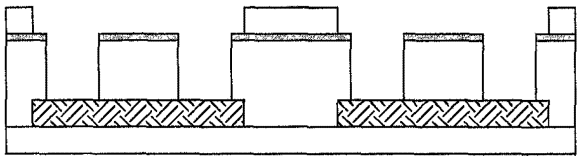
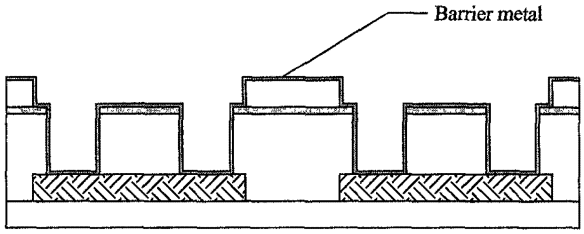
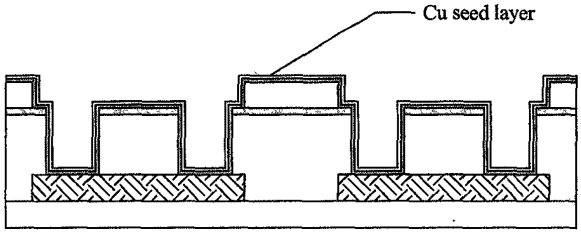
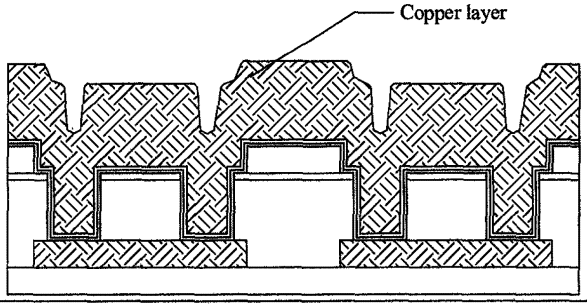
The semiconductor industry is implementing copper as the interconnect material for microchip fabrication. The traditional process flow will not be used for copper metallurgy because copper is not suitable for dry-etching. Copper also has special requirements to minimize its diffusion into silicon. To form copper interconnect wiring, a dual-damascene process is used to avoid copper etching. In the damascene process, it is no longer metal etch that defines the critical line width and spacing, but rather dielectric etch.

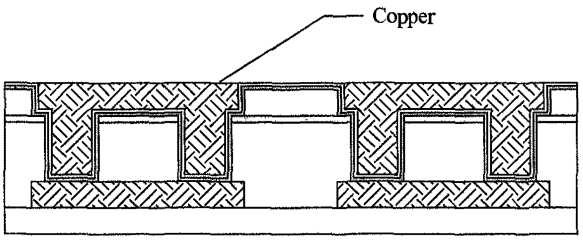
Dual Damascene Process ■ Since copper is difficult to etch, early developers of copper technology were forced to consider alternative approaches to creating metal lines. The *dual-damascene process* became the consensus process for copper metallurgy. It creates both the vias and lines for each metal layer by etching holes and trenches in the ILD, depositing copper in the etched features and using chemical mechanical planarization to remove the excess copper (referred to as *copper overburden*).

There are many possible scenarios for a dual-damascene process flow. Table 12.5 explains a basic process flow with the essential technologies that may or may not include all the steps performed by a particular manufacturer.

TABLE 12.5* Copper Metallization using Dual Damascene

Process Step	Description	Structure
1. SiO_2 deposition	ILD oxide deposition with PECVD to desired thickness for via. There is no critical gap fill; therefore, PECVD is acceptable.	
2. SiN etch stop deposition	Thin (250 Å) SiN etch stop is deposited on ILD oxide. The SiN needs to be dense and pinhole-free; therefore, HDPCVD is used.	
3. Via patterning and etch	Photolithography to pattern and dry etch via openings into SiN. Strip photoresist after completion of etch.	

Process Step	Description	Structure
4. Deposit remaining SiO_2	PECVD oxide deposition for remaining ILD oxide.	
5. Interconnect patterning	Photolithography to pattern SiO_2 trench with resist. Previously patterned via openings are located in trench.	
6. Etch trench for interconnect and hole for via	Dry etch trench in ILD oxide, stopping on the SiN layer. Etch continues to form via opening by passing through opening in SiN.	
7. Deposit barrier metal	Deposit Ta or TaN diffusion layer with ionized PVD on bottom and sidewalls of trench and via.	
8. Deposit Cu seed layer	Deposit continuous Cu seed layer with CVD. The layer must be uniform and free of pinholes.	
9. Deposit Cu fill	Deposit Cu fill with electrochemical deposition (ECD). Fill both via opening and trench.	

Process Step	Description	Structure
10. Remove excess Cu with CMP	Remove excess Cu using chemical mechanical planarization (CMP). This planarizes the surface and prepares for the next level. The resulting surface is a planar structure with metal inlays in the dielectric to form the circuitry.	

* P. Singer, "Making the Move to Dual Damascene Processing," *Semiconductor International* (August 1997): p. 79.

For dual-damascene processing, a two-tiered metal inlay is created that includes a series of holes (for contacts or vias) in addition to the interconnect trench for metal lines. The vias and metal line levels undergo copper fill concurrently, which saves a process step and eliminates the interface between the via and metal line.

There are different approaches to performing dual-damascene processing, each with their own set of challenges. Most of the manufacturing methods integrate the use of etch-stop materials to control how deep etching occurs for the via and trench. Etch stop is typically done by using a barrier metal such as silicon nitride (Si_3N_4 or Si_xN_y) as a hard mask at the bottom of the via or trench. Si_3N_4 etches at a much slower rate than the dielectric material, thus effectively stopping the etch process. Some approaches may even have two separate Si_3N_4 layers, which adds process complexity. In addition, Si_3N_4 has a high-k and increases the interlayer capacitance of the interconnect stack, which is why the Si_3N_4 layer needs to be thin.

The dual-damascene flow shown above has the via etch done in the same etch step as the trench, with the SiN etch stop to define the bottom of the trench. In other variations of dual-damascene, the via is etched first, then the trench is patterned and then etched. Trench structures created over topography from a previous via etch pose a significant challenge for photolithography.

The most important reason for using a damascene approach to copper metallization is to avoid metal etch. A second advantage to damascene processing is that there is no further need for dielectric gap fill between etched metal lines, since the dielectric is applied as a blanket and then etched.

The dual-damascene process in wafer fabrication potentially uses 20 to 30% fewer steps than traditional aluminum metal interconnect. Not only are there fewer manufacturing steps, but dual damascene also eliminates or reduces some of the most difficult steps of traditional aluminum metallization, including aluminum etch and many of the tungsten and dielectric chemical-mechanical polish steps.[54] Reducing process steps in wafer fabrication is important for improving the yield of the process, since fewer process steps translate to fewer sources of error that reduce assembly yield.

METALLIZATION QUALITY MEASURES

Wafer fabrication quality measures for metallization are provided in Table 12.6.

TABLE 12.6 Quality Measures for Metallization

Quality Parameter	Types of Defects	Remarks
1. Sputtered metal adhesion.	A. Metal layer does not adhere to substrate.	Parameters affecting film adhesion are: • Wafer contamination • Stress • Types of materials • Temperature of substrate • Argon pressure
2. Sputtered film stress.	A. Excessive stress leads to: • Cracks in the film surface. • Loss of film adhesion. • Increased resistivity in some materials.	Film stress can be caused by excessive temperature of the wafer substrate. Ways to reduce wafer temperature are: • Lower deposition rate • Increased backside cooling Cracks can cause: • Peeling of the film layer • Contaminant migration • Electrical opens or short. Measure film stress (bow of wafer surface) before and after deposition. Units for stress are dynes.
3. Sputtered film thickness.	A. Metal layer does not meet film thickness specifications (e.g., sheet resistance out of specification).	Parameters that affect film thickners are: • Incorrect recipe • Improper flow rates • Improper substrate temperature • Improper chamber pressure • Improper power supply energy • Wrong time setting
4. Uniformity of electroplated (electrochemical deposition, or ECD) metal film.	A. Nonuniform film thickness indicated by: • Inadequate gap fill and step coverage for high aspect ratio openings (nonuniform metal thickness at top and bottom of high aspect ratio holes). • Nonuniform deposition thickness across the wafer and from wafer to wafer. • Voids in film.	Critical parameters for electroplating film uniformity are: • Depositing a uniform CVD seed layer without voids. • Maintaining proper electroplating bath chemistry (composition and concentration) for organic additives, primarily brighteners and suppressors, to attain void-free deposition that fills bottom and sidewalls of gaps.

METALLIZATION TROUBLESHOOTING

Common troubleshooting problems for metallization are provided in Table 12.7.

TABLE 12.7 Common Metallization Troubleshooting Problems

Problem	Probable Cause	Corrective Actions
1. Degradation in step coverage of metal film.	A. Decrease in substrate temperature. For Al alloy sputter deposition, step coverage is dependent on the wafer temperature during deposition.	Heating the substrate to improve step coverage because surface mobility of deposited metal atoms improves.
	B. Increase in deposition rate.	Decrease deposition rate. Increase in sputter deposition rate may degrade step coverage due to reduced surface mobility because of arrival of more atoms at surface.
2. Vacuum chamber integrity.	A. Chamber cleanliness or moisture in chamber.	• Check for vacuum leaks or chamber outgassing. Residual gas in chamber that can change film reflectivity. • Check for O_2 or N_2 in metal films that can change resistivity and film stress. • Clean chamber and do H_2O bake-out.
	B. Outgassing or system leaks.	• Use helium leak detection to inspect for system leaks. • Assess the chamber conditions before deposition with residual gas analyzer.
3. Contamination of metal film.	A. Particles on surface of film.	Check the following common particle sources: • Dirty input wafers. • Problems in hand-off between robot and chamber mechanism. • Incomplete chamber cleaning between runs. • Dirty cassettes. • Contaminated nitrogen backfill.
4. Voids in copper trench fill after dual-damascene electroplating.	A. Excessive seed layer thickness on the wafer surface (field) that pinches off the via or trench, creating a center void in the film.*	Optimize the CVD Cu seed layer deposition by evaluating the electroplating current from the field to the bottom of vias and trenches. The goal is to deposit adequate seed layer at the bottom of a high-aspect ratio feature without increasing field thickness.
5. Excessive copper dishing after dual-damascene CMP.	A. Cu dishing for Cu CMP is often caused by tantalum diffusion barrier, which must be polished back in the presence of Cu.**	After Cu is polished, the Ta layer must be removed from the level. The Cu pad/slurry field does not effectively remove Ta. Options are: • Optimize the pad/slurry for Ta. • Minimize the Ta field level thickness.

* R. Jackson et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998): p. 56.
** Ibid.

SUMMARY

Metallization deposits a thin metal film to form the interconnect wiring and contact or via connections on the chip. There are six metal categories used in wafer fabrication, with specific requirements for optimum performance. Aluminum has been the traditional metal for interconnect lines. An ohmic contact is the low-resistance contact between silicon and the interconnect metal. Aluminum is sometimes alloyed with silicon to reduce junction spiking at the ohmic contact. Aluminum is often alloyed at 0.5 to 4% copper to minimize electromigration reliability problems. New interconnect metallization is based on copper metallurgy to decrease the metal resistivity. Copper combined with low-k dielectric will decrease the chip interconnect delay, but copper does require a new processing approach since it does not etch well. Barrier metals are used when joining metals. Different barrier metals have optimum properties depending on the application. A silicide is a refractory metal alloyed with silicon to improve contact resistance and adhesion. A salicide is a particular silicide that is aligned to the source, drain, and gate structure. Metal layers are connected with vias filled with tungsten plugs. The most widely used metal deposition system is sputtering. Sputtering physically bombards a target to dislodge atoms and deposit the atoms as a thin film on the wafer surface. The three most common types of sputtering are RF, magnetron, and ionized metal plasma. RF sputtering is inefficient, while magnetron has a higher deposition rate but limited gap-fill capability. Ionized metal plasma has improved directionality for high-aspect ratio deposition. CVD metal has the best conformality and is used for tungsten plug fill and copper seed layer. Copper electroplate is the process chosen by major semiconductor manufacturers for copper wiring deposition. Traditional aluminum metallization deposits a blanket aluminum film and then etches line patterns to form interconnect wiring. Copper metallurgy will use a dual-damascene process that etches vias and trenches in dielectric, deposits copper fill, and then removes excess copper through chemical mechanical planarization.

KEY TERMS

interconnect
wiring
contact
vias
plug
eutectic temperature
anneal or sinter
ohmic contact
junction spiking
electromigration
hillocks
copper interconnect
barrier metal
refractory metals
silicide
polycide
salicide
tungsten plugs
evaporation
physical vapor deposition (PVD)
target
momentum transfer
sputtering yield
sputter etch (reverse sputter)
RF sputtering
magnetron sputtering
collimated sputtering
sputtered aluminum
ionized metal plasma PVD (IMP or ionized PVD)
tungsten (W) CVD
copper CVD
electroplate (electroplating, electrochemical deposition (ECD), or electrofill)
evaporation
aluminum subtractive process
dual-damascene process

REVIEW QUESTIONS

1. Define the following terms: interconnect, contact, via, and plug.
2. What metal has traditionally been used for interconnect wiring, and what is its replacement?
3. List and describe the seven performance requirements for metals used in wafer fabrication.
4. List the categories of metals and metal alloys used in semiconductor manufacturing.
5. Explain why aluminum has been the interconnect metal of choice for microchips.
6. What is an ohmic contact and what is its advantage?
7. Describe junction spiking and list the two primary ways it was resolved.
8. Discuss electromigration and how it affects reliability. How is electromigration controlled in aluminum wiring?
9. List and discuss the five benefits for introducing copper metallization.
10. What are the three major challenges for converting to copper for interconnects?
11. Explain what a barrier metal is and the essential properties of a barrier material. What metals are commonly used as a barrier metal?
12. List the special requirements for a copper barrier metal.
13. Define silicide and explain why refractory metal silicides are important in wafer fabrication.
14. What are the benefits of titanium silicide? Explain the C49 and C54 phase structures.
15. What silicide is promising for future wafer technology? Why?
16. What is the salicide process?
17. Describe a tungsten plug, and discuss how it is used in multilayer metallization.
18. Why was evaporation replaced as a metal deposition system?
19. Provide a short explanation about sputtering and how it works.
20. Is sputtering suitable for alloy deposition?
21. What are the advantages to sputter deposition?
22. List and explain the six steps to sputtering.
23. Explain the physics of sputtering.
24. Describe sputter yield.
25. What is the purpose of sputter etch?
26. Describe an RF sputter system. What is its major limitation?
27. Discuss how a magnetron sputtering system improves the deposition rate.
28. Explain why sputtering has poor step coverage. How does collimated sputtering improve step coverage?
29. Describe ionized metal plasma. How does this process improve high-aspect ratio gap fill?
30. What is the typical method for creating a tungsten plug in advanced ICs?
31. Why could CVD be used for the copper seed layer?
32. Explain the fundamental process of copper electroplate.
33. Why is traditional aluminum metallization a subtractive process?
34. What is the primary reason for implementing the dual-damascene process for copper?
35. List ten steps for a dual-damascene metallization process.

METALLIZATION EQUIPMENT AND MATERIALS SUPPLIERS' WEB SITES

Angstrom Sciences Inc.	http://www.angstromsciences.com/
Applied Materials	http://www.appliedmaterials.com/products/
Genus Incorporated	http://www.genus.com/
Materials Research Corporation	http://www.materialsresearch.com/
Nordiko USA Inc.	http://www.nordiko.com/
Novellus Systems Inc.	http://www.novellus.com/index.htm
Process Materials Inc	http://www.processmaterials.com/
TEL, Tokyo Electron Ltd.	http://www.teainet.com/
Veeco-CVC	http://www.veeco.com

REFERENCES

1. C. Weber, D. Jensen, and E. Hirleman, "What Drives Defect Detection Technology?" *Micro* (June 1998): p. 60.
2. M. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *Solid State Technology* (September 1996): p. 105.
3. S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed. (New York: Wiley, 1994): p. 548.
4. K. Bachmann, *The Materials Science of Microelectronics,* (New York: VCH Publishers, 1995), p. 484.
5. Compiled from multiple sources, including: S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology,* 2nd ed., (Sunset Beach: Lattice Press, 1986), p. 399.
6. L. Peters, "Advanced Aluminum Interconnect Technology," *Semiconductor International* (November 1998): p. 83.
7. R. Jaeger, *Introduction to Microelectronic Fabrication,* (Reading, MA: Addison-Wesley, 1988), p. 136.
8. S. Campbell, *The Science and Engineering of Microelectronic Fabrication,* (New York: Oxford University Press, 1996), p. 406.
9. Ibid., p. 412.
10. E. Amerasekera and F. Najm, *Failure Mechanisms in Semiconductor Devices,* 2nd ed., (New York: Wiley, 1997), p. 103.
11. J. Yue, "Reliability," Sze, *ULSI Technology,* ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 672.
12. K. Rodbell, "Reliability," *Handbook of Semiconductor Interconnection Technology,* ed. G. Schwartz, K. Srikrishnan, and A. Bross (New York: Marcel Dekker, 1998), p. 480.
13. S. Campbell, *The Science and Engineering of Microelectronic Fabrication,* p. 413.
14. J. Yue, "Reliability," p. 663.
15. P. Singer, "Tantalum, Copper and Damascene: The Future of Interconnects," *Semiconductor International* (June 1998): p. 91.
16. "Technology News," *Solid State Technology,* (February 1998): p. 26.
17. P. Singer, "Tantalum, Copper and Damascene," p. 91.
18. Ibid.
19. A. Sethuraman, J. F. Wang, and L. Cook, "Copper vs. Aluminum: A Planarization Perspective," *Semiconductor International* (June 1996): p. 178.
20. X. Lin and D. Pramanik, "Future Interconnect Technologies and Copper Metallization," *Solid State Technology* (October 1998): p. 63.
21. P. Singer, "Tantalum, Copper and Damascene," p. 94.
22. Semiconductor Industry Association, *National Technology Roadmap for Semiconductors: Technology Needs,* (San Jose: SIA, 1997), p. 101.
23. C. Ryu et al., "Barriers for Copper Interconnections," *Solid State Technology* (April 1999), p. 53.
24. B. Chin et al., "Barrier and Seed Layers for Damascene Copper Metallization," *Solid State Technology* (July 1998), p. 141.
25. C. Ryu, "Barriers for Copper Interconnections," p. 56.
26. A. Braun, "Copper Electroplating Enters Mainstream Processing," *Semiconductor International* (April 1999), p. 64.
27. R. Jackson, et al., "Processing and Integration of Copper Interconnects," *Solid State Technology* (March 1998): p. 56.
28. S. Ghandhi, *VLSI Fabrication Principles,* p. 576.
29. D. Campbell, "Semiconductor Contact Technology," *Handbook of Semiconductor Interconnection Technology,* ed. G. Schwartz, K. Srikrishnan and A. Bross (New York: Marcel Dekker, 1998), p. 176.
30. K. Rodbell, Reliability," p. 480.
31. L. Peters, "Is the 0.18 μm Node Just a Roadside Attraction?" *Semiconductor International* (January 1999), p. 50.
32. R. Liu, "Metallization," *ULSI Technology,* ed. C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 410.
33. S. Campbell, *Science and Engineering,* p. 292.
34. R. Liu, "Metalization," p. 379.
35. X. Lin and D. Pramanik, "Future Interconnect Technologies," p. 74.
36. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology,* 2nd ed., pp. 443–50.
37. M. Rittner, "Growth Predicted for Sputtering Target and Sputtered Film Markets," *Solid State Technology* (January 2000): p. 26.
38. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology,* p. 445.
39. A. Braun, "Sputtering Targets Adapt to New Materials and Shrinking Architectures," *Semiconductor International* (June 1998): p. 130.
40. S. Wolf and R. Tauber, *Silicon Processing for the VLSI Era,* vol. 1, *Process Technology,* p. 449.
41. Ibid., pp. 348–51.
42. R. Liu, "Metallization," p. 382.
43. Ibid., p. 384.
44. B. Chin et al., "Barrier and Seed Layers," p. 141.
45. A. Braun, "Sputtering Targets," p. 128.
46. Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors: Technology Needs,* (San Jose: SIA, 1997), p. 101.
47. G. Schwartz, "Metallization," *Handbook of Semiconductor Interconnection Technology,* ed. G. Schwartz, K. Srikrishnan and A. Bross, (New York: Marcel Dekker, 1998), p. 320.
48. J. Baliga, "Depositing Diffusion Barriers," *Semiconductor International* (March 1997): p. 77.
49. K. Wijekoon, et al., "Tungsten CMP Process Developed," *Solid State Technology* (April 1998): p. 53.
50. B. Zorich and M. Majors, "Safety and Environmental Concerns of CVD Copper Precursors," *Solid State Technology* (September 1998): p. 101.
51. P. Singer, "Tantalum, Copper and Damascene," p. 94.
52. Ibid.
53. Ibid., p. 91.
54. Ibid.

CHAPTER 13

PHOTOLITHOGRAPHY: VAPOR PRIME TO SOFT BAKE

The essence of photolithography is to print temporary circuit structures on a wafer that will later be used to assist etch and ion implant processes. The structures are first created as patterns on a quartz template that is referred to as a mask. Ultraviolet light passes through the mask to transfer the pattern to a light-sensitive film on the wafer surface. As with common photography work, the photographic film is developed and the image appears on the wafer. Later a chemical etching process is used to image the film pattern on the underlying wafer surface, or the wafer is sent to the ion implant area where selective doping can be performed through the patterns on the wafer. The various patterns transferred to the wafer define the numerous device features, vias, and wiring interconnects necessary for the device layers as well as determine areas of the silicon to be doped.

Photolithography is closely related to the cost and performance of a microchip. The processing cost of a wafer is largely independent of the number of chips on a wafer. That is, a wafer with few chips costs almost as much as the same wafer with many chips since the number of processing steps, amount of materials, wafer handling, and so on, are nearly identical for the two wafers. More chips at the same cost means that the overall cost of each chip is lower, simply by putting more chips on the wafer.

Since the beginning of semiconductor manufacturing, photolithography has been recognized as the driving force behind the integrated circuit fabrication process. This trend continues today as the industry strives to pack more devices and the associated circuitry on a chip. More than any other single fabrication technology, photolithography has contributed to the revolutionary advancement of chip performance, affirming Moore's law over the fifty-plus years of semiconductor manufacturing.

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the basic concepts for photolithography, including process overview, critical dimension generations, light spectrum, resolution, and process latitude.
2. Discuss the difference between negative and positive lithography.
3. State and describe the eight basic steps to photolithography.
4. Explain how the wafer surface is prepared for photolithography.
5. Describe photoresist and discuss photoresist physical properties.
6. Discuss the chemistry and applications of conventional i-line photoresist.
7. Describe the chemistry and benefits of deep UV (DUV) resists, including chemically amplified resists.
8. Explain how photoresist is applied in wafer manufacturing.
9. Discuss the purpose of soft bake, and explain how it is accomplished in production.