Compendium TFE4180

Hans Henrik Urdahl

May 16, 2013

Contents

1 Jenkins

1.1 Chapter 1.3 - Positions, directions and planes in crystals

The position of an atom in a unit cell is determined using vectors, as in materials science. Planes are determined by Miller indices (hkl). Find the intercepts of the plane on the crystal axes, take the reciprocals of the numbers, and reduce eventual fractions so that all numbers are integers. Families of planes are denoted as {hkl}.

1.2 Chapter 1.4 - The major crystal structures of semiconductors

The simplest materials to discuss are silicon and germanium, with covalent bonds, tetrahedral orientation and diamond structure. These structures are quite open, which leads to low density. GaAs and CdTe are binary compounds with zinc-blende structure.

1.3 Chapter 2.3.5 - Epitaxial growth of semiconductors

1.3.1 Vapour-phase epitaxy

VPE is a process where epitaxial solid film is formed between gaseous solids. Also known as chemical vapor deposition (CVD). The driving force is the change in free energy due to the chemical reaction. The most common form of VPE is the open-flow VPE, or the CVD. Reactant gases flow through the chamber at T_0, and the substrate is at a higher temperature, T_s, which allow chemical reactions to happen, and an epitaxial layer is formed. By-products are transported out of the chamber. VPEs can be either cold-wall or hot-wall. Cold-wall reactors can create uneven layers, due to differences in temperature. Difficulties in gas transport, and equilibrium problems can be solved by multiple chambers and different gas compositions. The rate of gas flow can also affect the results. Slow gas flow creates turbulence, and uneven layers. High gas flow is preferred. Exhaustion of reactants can also create uneven layers. Tilting of the substrate can improve results.

1.3.2 Molecular beam epitaxy

A thermally controlled evaporation process in an ultra-high vacuum chamber. Neither collisions or reactions between the particles in the beam before they reach the substrate. Transport of mass is by molecular flow, not viscous flow as in CVD. Parameters must be controlled carefully, and the substrate must be clean and defect-free. The growth rate is relatively slow, about an atomic layer per second. The result of this is the ability to control the composition of the sample at an extremely high level. The sample is placed at the center of a steel vessel, cooled by N_2 to provide low gas-pressure. Knudsen cells (effusion cells) contain the different materials we want to deposit on the sample, and they are heated by resistive techniques. The cells can be shuttered, so that we choose which material to deposit. The flux is controlled by the temperature of the cells. The surface of the sample has to be clean, and is often lightly sputtered to remove impurities. A RHEED (reflective high-energy electron diffraction) gun is used to monitor the structure of the outer layers of the sample during growth.

1.3.3 Liquid phase epitaxy

LPE involves growing an epitaxial layer by placing a saturated solution in contact with the substrate. The solution is cooled, which leads to deposition of material on the sample. The following steps are involved. A supersaturated solution is heated above saturation temperature to obtain a homogenous solution. It is cooled to a starting temperature, and brought into contact with the substrate. The temperature is lowered, while growth happens, and the substrate is removed to prevent further growth.

1.4 Chapter 2.5.2 - Chemical vapor deposition

Originally used to produce oxide layers. Gases are carried to the reaction chamber by inert gases as nitrogen. CVD grow films which tend to be non-uniform and porous. This can be countered by introducing certain gases by separate ports.

1.5 Chapter 3.2.1 - Electron states in a pure semiconductor

The energy states available to electrons in a solid is important to understand in order to explain the electrical properties of a semiconductor.The total energy of a electrons in a solid is composed of potential energy and kinetic energy. Potential energy arises from interaction between the electrons and the nuclei of the atoms. Near the nucleus the potential energy is a deep well. Core electrons are tightly bound and strongly localized. The valence electrons, on the other hand, are less bound and localized, which means that their energy levels are affected by neighboring atoms. By regarding these free electrons as a simple "gas" within the body of the semiconductor, one can derive Ohm's law from the Maxwell-Boltzmann distribution,

$$J = \sigma E \tag{1}$$

where

$$\sigma = \frac{ne^2\tau}{m}.$$

n is the number of electrons per volume unit, $e$is the electronic charge and τ is the characteristic time or relaxation time for the semiconductor. Another important quantity is the mobility of charge carriers,

$$\mu = \frac{v_d}{E} \tag{2}$$

where v_d is the drift velocity of the electrons in an applied field, and E is the electrical field.

1.6 Chapter 3.2.5 - Semiconductor-semiconductor junctions

The homojunction is an abrupt junction between a n-doped and p-doped semiconductor. Electrons will flow from the n-doped side to the p-doped side, and , which creates a Coloumb barrier. Other junctions are possible, but will not be discussed here.

1.7 Chapter 3.3 - Electrical conductivity, mobility and the Hall effect in semiconductors

When an electrical field is applied to mobile carriers in the conduction and valence band, and a magnetic field is applied, the carriers will experience a force,

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} \tag{3}$$

This is the Lorentz force on a charged particle. This force produces an acceleration, and gives rise to the Hall effect. If we assume that the electrical and magnetic fields are perpendicular to each other, a charge gradient perpendicular to the applied electrical field will create a transverse electrical field. In a steady state situation, this induced field will balance the force from the magnetic field. The field that achieves this is known as the Hall field ($\mathbf{E_H}$). More about this can be found in Jenkins.

1.8 Chapter 6.2 - The interaction of ions with semiconductors

When a beam of energetic charged ions incident on the surface of a semiconductor, two types of interactions can take place. The ions can interact with the nuclei of the atoms or with the electrons. The most important effects that occur after a bombardment of a semiconductor surface are as follows.

- The atoms of the surface may be given enough momentum (or energy) to be sputtered. Some of the sputtered atoms may be ionized, which gives rise to the technique of secondary ion mass spectrometry (SIMS).
- Ions may be reflected from the surface. Such ion scattering can be analyzed.
- The ions may be captured by the nuclei in the solid, and nuclear reactions can happen. This leads to radiation, which can be analyzed.
- The ions may transfer energy to inner electron shells of the atoms. These electrons will emit X-rays, which can be analyzed.

1.9 Chapter 6.3 - Secondary ion mass spectrometry

When an energetic ion collides with the surface of a semiconductor, a surface atom may receive enough energy to be ejected from the surface, also called sputtered. A small fraction of the sputtered ions become ionized. These ions emerge from the top monolayers of the surface, and cane analyzed with a mass spectrometer. The rate of sputtering for a given incident ion flux j is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{jS}{N} \tag{4}$$

where N is the number of atoms per unit volume of the target, S is the number of atoms sputtered per incident ion, also called the sputtering yield. This factor is dependent on many factors, as the mass and energy of the ion, the binding energies of the target atoms and different surface parameters. The rate of sputtering also relies on the incident primary ion flux. This leads to two regimes.

- *Static SIMS*: Low primary ion flux, and low primary ion energy. The result is an almost negligible sputter rate, so that a nearly undisturbed surface monolayer can be analyzed.

- *Dynamic SIMS*: Higher primary ion flux and energy, Usually, more than one monolayer is removed from the surface. This sputtering allows us to produce depth profiles of semiconductors, where a focused ion beam (FIB) is raster-scanned over an area, greater than its diameter. A flat-bottomed crater is created, which can be measured by a profilometer.

The choice of primary ion depends on the sensitivity required. Noble gas ions, as argon, produce the least disturbance of the surface, while reactive ions, like O_2^+ and O^- provide higher sensitivity. A problem with SIMS analysis of insulators is that the surface becomes charged, which can affect the secondary ion yield. Therefore *fast bombardment SIMS* (FABSIMS) is used. *Laser ionization mass analysis* (LIMA) can also be employed.

1.10 Chapter 7.1 - Optical microscopy and imaging

The unaided human eye can resolve points with a separation of about 0.2mm. Therefore, optical microscopy is very important in many areas, e.g. characterization of semiconductors. If a microscope has a magnification on the order of 100, the eye can resolve points with a separation of 2 μm. To increase the apparent size of an object, we have to bring the object closer to the eye. However, the *least distance of distinct vision* for the eye is about 25cm. A simple microscope is made up of a single convex lens, which allows us to bring the object closer to the eye by creating a virtual image that seems further away from the eye. The magnification (m) is given by

$$m = \frac{v}{u} \tag{5}$$

where v is the image distance, and u is the object distance. For a thin lens with focal length f, this can be written as

$$m = v/f - 1 \tag{6}$$

where v is a negative number because the image is virtual (it cannot be displayed on a screen). The magnification of such a system is very small. To increase the magnification, a *compound microscope* is used. This microscope consist of at least two stages, and can provide magnification up to 1500X.

The simplest form consist of two lenses, the *objective*, closest to the object, and the *eyepiece*. The objective has a very small focal length, and forms a real, inverted image. The image is further magnified by the eyepiece, and the total magnification is the product of the two lenses' magnifications. The light source in such microscopes depend on the characteristics of the sample (opaque, transparent or solid), and one can use a condenser lens to intensify the light if the sample is transparent.

This simple description of a microscope ignores the wave nature of light. Light from a point source does not produce a point image, due to diffraction of light in the lens system. The light is spread out, but most of the light is concentrated in a central region, known as the *Airy disc*. The radius of the central region of the Airy disc of a point image is

$$R = \frac{1.22\lambda D}{2f} \tag{7}$$

where λ is the wavelength of the light, D is the effective diameter of the lens, and f is the focal length of the lens. A microscopes power to resolve separate points is known as its *resolving power*. As the points get closer to each other, the Airy discs will overlap. When two separate images of two point objects are just resolved, the separation of the objects are

$$D = \frac{0.61\lambda_0}{n_0 \sin\theta_0} \tag{8}$$

where $n_0 \sin\theta_0$ is called the *numerical aperture* of the lens (NA), and λ_0 is the wavelength of the light in vacuum. The NA is a measure of the lens' power to gather the light. Equation 8 describes the theoretical limit, and in practice, the resolution will be lower. The maximum NA of a top quality lens is about 1.6. We can increase the resolution by increasing the wavelength of the light used, and if we use electrons, in an *electron microscope*, the wavelength is shorter than for light, which results in a greatly increased resolution.

To distinguish details in an object, they have to stand out from their surroundings. This is called *contrast*. Normally in microscopy, contrast is provided because of local differences in absorption, scattering and reflection. If a feature causes small changes, they can be hard to detect due to sample illumination entering the optical system. This is the problem with so-called *bright field illumination*. *Dark field illumination* can fix the problem. The technique is based on placing an opaque disc over the center of the condenser. The edges of the sample is illuminated, and features as defects will diffract some light into the objective, and we get a bright image against a dark background. Due to oblique illumination, no light enters the optical system directly.

A technique used for detecting defects in semiconductors is called *decorating*. Impurity atoms are diffused into the crystal, and the crystal is allowed to cool slowly. This causes the impurity atoms to diffuse to the locations of the errors. The impurity precipitates can be detected in transmitted light. The host material must be transparent. Infrared light (IR) can be used. Polarized light can also be used to identify dislocations.

The *Nomarski microscope* is an important microscope in the class of microscopes called *phase contrast microscopes*. This microscope detects very small differences thickness or refractive index. The key component is the *Nomarski prism*, which splits a single beam of light into two beams of in-phase light. If the surface is not perfectly smooth and reflective, the light that passes through the prism after reflections will have a different phase, and the polarization of the recombinant light beam will be elliptically polarized. The analyzer of the microscope is set to only transmit the elliptically polarized light, and so we get a picture where the surface relief is bright against a dark background. This microscope can show step heights as small as 3nm.

The *confocal microscope* passes light through a pinhole, which results in "point images", where the surroundings are excluded. In ordinary microscopes, the surroundings are blurry, but in the confocal microscope, they are not included. This means that in order to create an image, the microscope has to "scan" the surface. This also means that one can build a "three dimensional" image of the sample. This is known as *optical sectioning*.

1.11 Chapter 7.2 - Electron microscopy

Optical microscopy limits the resolution to about 200nm. The wavelength of an electron accelerated by a voltage of 100keV, the wavelength of the electron is about 0.004nm. However, the equivalents of lenses in electron microscopy (electrostatic and magnetic lenses) have very low NA (around 0.01), so the theoretical resolution is somewhere around 0.25nm for normal electron microscopes.

1.11.1 Chapter 7.2.1 - The transmission electron microscope (TEM)

An electron source (often a tungsten filament which is heated) sends electrons through a condenser lens onto the specimen, in a beam. The sample is very thin, around 100nm, so that most electrons pass through it. Variations in the atomic composition or thickness of the sample

produces changes in the electron scattering from the main beam. If the sample is crystalline, strong Bragg diffraction of the beam will occur in specific directions. As in optical microscopy, one can include or exclude the original beam to create bright or dark field imaging. Therefore, the TEM can also show lattice images.

1.11.2 Chapter 7.2.2 - The scanning electron microscope (SEM)

The SEM focuses the electron beam to a small spot, and scans the entire sample. The microscope detects secondary electrons scattered back from the surface. The SEM does not look at transmitted electrons, and therefore, the sample can be quite thick. The process used in the SEM can result in a charged surface, if the sample is not vary conducting. Therefore, the surface of such samples are covered in a thin layer of metal, because charge build up can alter the final image. The contrast in the SEM is produced from the difference in electrons reaching the detector from different regions of the spectrum. The simplest form of contrast comes from irregularities in the surface of the semiconductor. The number of electrons reaching the detector depends on the angle between the surface and the incident electron beam. The brightness depends on the surface inclination, in such a way that the image depicts a three-dimensional effect. The spot size of the electron beam decides the resolution of the microscope, and the spot size is decided by the Airy pattern obtained from diffraction at the aperture of the final focusing lens. However, reducing the spot size reduces the secondary electron emission, and at some point, the signal from the secondary electrons will be overcome by noise. The SEM can also provide contrast in crystalline semiconductors via electron channeling between low-index planes. If there are interstitial atoms, the backscattering will be large. Another interesting phenomena in the SEM analysis, is the excitation of an atom, which leads to emission of X-rays. These rays can be detected and analyzed using *energy-dispersive spectrometers* (EDX spectrometers), which can determine which elements are present in the sample, provided they have an atomic number over 10.

1.12 Chapter 7.3 - Scanning tunneling and atomic force microscopy

The *scanning tunneling microscope* (STM) does not need lenses, light or electron sources. The STM uses an atomically sharp, conductive tip which is held fractions of a nanometer over the surface. This creates a current because of the surface electrons of the semiconductor, due to quantum mechanical tunneling of electrons. When scanning the surface, differences in current flow determines the distance between the tip and the sample, and a true image is created. The tip is mounted on a system of piezoelectric transducers coupled with a computer, which controls the tip dynamically. Horizontal resolution of 0.1nm is routinely available with commercial instruments. The topographic images depend on the variations of electron density over the surface. The STM can only analyze conducting or partially conduction samples. This problem is overcome with the *atomic force microscope* (AFM). The sensor in an AFM consist of a flexible micro fabricated cantilever with a tip. When the sample pushes against the tip, interatomic repulsion causes the cantilever to deflect, and this deflection is measured by an optical lever technique. A laser beam is reflected off the cantilever, and onto a photodiode. When the cantilever is deflected, the angle of the laser beam detected is changed, and a computer detects the changes. The sample is mounted on a piezoelectric transducer, and can be moved up and down to return the cantilever to the original state. The voltage needed to return the probe to its original state is the measure of the local topography. The AFM can sense tip displacements as small as 0.01nm.

2 Quirk & Serda - Semiconductor Manufacturing Technology

2.1 Chapter 2 - Characteristics of Semiconductor Materials

Energy-band theory of solids describe semiconductors as materials with an intermediate energy band gap between the valence band and conduction band of the solid. *Conductors* have overlapping valence and conduction bands, or a very small band gap, which means that valence electrons require little energy to move from the valence band to the conduction band, and such materials are capable of carrying an electrical current. *Insulators* are materials with high resistivity, and are sometimes referred to as *dielectrics*. They oppose the flow of electrical current. The property to carry a current is called the *conductivity* of a material, C. From this property, resistivity (ρ), is defined as

$$\rho = \frac{1}{C} \tag{9}$$

Resistance (R) is the opposition of current, and is defined as

$$R = \frac{\rho l}{\text{area}} \tag{10}$$

where l is the length of the conductor and area is the cross-sectional area of the conductor.

Capacitance is the storage of charge on two conductive plates separated by a dielectric medium. There is a difference in potential between the plates, and an electrostatic field exists between them.

2.1.1 Silicon

One of the most used semiconductors in IC fabrication. Silicon is a semiconductor because of its four valence electrons. Pure silicon is called *intrinsic silicon*, with no impurities. It is strongly covalent bonded. Pure silicon is a poor conductor. It was chosen for semiconductor fabrication because of the following reasons.

- Abundance of silicon.
- High melting point for wider processing range.
- Wider temperature range of operation.
- Natural growth of silicon dioxide.

Silicon dioxide is a high-quality dielectric and a good chemical barrier. SiO_2 has similar properties as pure silicon.

By doping silicon, the conductivity can be greatly improved. Doped silicon is also known as *extrinsic silicon*. Dopants are either n-type (with five valence electrons) or p-type (three valence electrons). This modifies the number of free electrons, and which kind of charge carriers dominate (electrons or holes).

Alternative semiconductor materials are used for specific applications. These are called *compound semiconductors*, and one example is GaAs (which is a III-V compound). One can also create II-VI compounds. GaAs has greater electron mobility than silicon, so the majority carriers move faster. This results in faster ICs. One of the primary drawbacks of GaAs is the lack of natural oxides, as well as the cost. The extreme toxicity of arsenic must also be considered.

2.2 Chapter 3 - Device Technologies

2.2.1 The pn Junction Diode

A pn junction diode is formed whenever a region of n-doped semiconductor is adjacent to a p-doped region. The pn junction diode is made of a single crystal semiconductor material, one region is heavily doped with donor dopant, while the other region is heavily doped with acceptor dopant. Initially, both regions are electrically neutral, but with time, holes and free electrons will diffuse across the junction, creating a *carrier-depletion region*. The areas close to the interface will become charged, and the net effect of difference in charges create a potential difference across the junction. This results in a *barrier voltage* that has to be overcome to operate the diode. By reverse biasing or forward biasing the diode, we can create the effect of an open circuit or a closed circuit.

2.2.2 CMOS IC Technology

CMOS (complementary metal-oxide semiconductor) revolves around the improvements of the *FET (field-effect transistor)*. The FET is a voltage-amplifying device. The FET has low voltage and power requirements. It turns on as a result of an electric field when voltage is applied to the *gate*. There are two types of FETs: the junction FET (JFET) and metal-oxide FET (MOSFET). The reason for calling it MOSFET is that the gate is insulated from the electrodes by a metal-oxide layer, called *gate oxide*. MOSFETs can be p-type (pMOS) and n-type (nMOS). This distinction is made because the major current carrier in the two types is either electrons or holes. The MOSFETs are made of the gate, the *source* and the *drain*. The latter ones are heavily doped with either p- or n-dopants (for a nMOS, the source and drain are p-doped and vice versa). The MOSFET can be biased for conduction or insulation. When a gate voltage is applied, an electric field is established, and the charge carriers on the other side of the insulating metal-oxide layer will diffuse away, effectively closing the circuit between the source and drain.

2.3 Chapter 7 - Metrology and Defect Inspection

IC metrology is a necessary means for measuring the performance of fabrication processes to ensure specific quality standards are upheld. Yield is defined as the ratio of acceptable products to the total number of products.

2.3.1 Unpatterned Surface Defects

Unpatterned wafers are bare wafers or wafers with various blanket films used as test wafers. These wafers are used for monitoring processes. Typical defects inspected for on unpatterned wafers are particles, scratches, slip lines and material defects. Defect-detection equipment for these types of defects are generally *dark field and bright field optics*. Dark field optical detection is often better for small defects. *Color interference contrast* splits a beam of light into a direct and a reference beam. The direct beam is altered by the sample, and recombined with the reference beam, creating an image based on interference.*Light scattering defect detection* uses light or laser light to scan the surface of the wafer, and identify defects based on the principle of dark field optics. *Particle per wafer per pass (PWP)* is a measure of the number of defects created in one process run.

2.3.2 Patterned Surface Defects

Because the patterned surface is not uniform, defects become harder to detect. Most defects are identified using light scattering techniques and optical microscopy. Light scattering techniques have to be modified to be able to distinguish the pattern on the surface from defects. **Critical dimension** measurements are important in order to achieve precise control over all line widths of the product. In CMOS, the transistor's gate structure is very critical, and measurement tools need a precision under 2nm to control the CD of the smallest structures. The tool for this task is the *SEM*. The CD-SEM is often automated, in order to achieve high wafer throughput.

2.3.3 Analytical equipment

The following equipment is used for analysis of wafers.

- *SIMS*: Secondary-ion mass spectrometry uses ions, created in a plasma, to sputter the surface of the sample, and analyze the secondary ions in a mass spectrometer. This method gives an accurate measure of the composition of the sample. The SIMS is, more or less, a destructive method of analyzing the composition of the surface. The ions used to sputter the surface can be focused into a small spot (called an ion microprobe), or can be used to flood the surface (known as ion microscope). The rate of sputtering determines if the technique is *dynamic (high sputter rate) or static (low sputter rate)*. The SIMS can only detect a very small amount of the backscattered, secondary ions. Therefore another technique, known as *time of flight SIMS (TOF-SIMS)* can be utilized. The TOF-SIMS measures the time a secondary ion uses to travel a fixed distance to determine the element. TOF-SIMS uses a very low sputter rate, and is therefore essentially non-destructive.
- *AFM*: The AFM uses a surface profiler to created a true 3D surface profile. The AFM measurements are very slow. The simplest AFM technique uses a laser beam, reflected off the surface of the cantilever that scans the surface. A more advanced method uses an oscillating cantilever, and measures shifts in phase due to can der Waals forces.
- *TEM*: The TEM measures transmitted electrons through an ultra thin slice of the desired sample. The resolution is high, but the TEM requires carefully prepared samples.
- *FIB*: The FIB is a destructive technique similar to SEM in design and operation, but uses Ga^+ ions in stead of electrons. The ions are focused into a beam by a set of lenses. The atoms of the sample are sputtered off. The FIB can create small voids of precisely controlled shape, size and depth (called FIB milling). However, the FIB is time-consuming.

2.4 Chapter 8 - Gas Control in Process Chambers

2.4.1 Objectives

- Explain why process chambers are used in semiconductor manufacturing.
- Describe the benefits of a vacuum, vacuum ranges and appropriate pumps.
- Explain the need for gas flow in process chambers, and how it is controlled.
- Explain what an RGA is, and why it is beneficial in process chambers.
- Describe plasma and how it is obtained.

- Discuss the effects of contamination and how to minimize it.

In present-day wafer fab, the *process chamber* is an important part of the equipment. The process chamber is a controlled vacuum environment where intended chemical reactions occur under controlled conditions (also called reactor). Some of the functions of the process chamber are: Controlling how gas chemicals flow into and react in the chamber in close proximity to the wafer, maintaining a prescribed pressure, removing undesired moisture, air and byproducts, creating an environment for chemical reactions such as plasma creation, controlling heating and cooling of the wafer. Today, the process chambers are ordered in a *cluster tool*, where multiple process chambers are clustered around a central transfer chamber.

2.4.2 Vacuum

The definition of vacuum is an enclosed volume where there is less pressure than in the surrounding atmospheric pressure. The benefits are as follows.

1. Create clean environment.
2. Low molecular density.
3. Extend the Mean Free Path of gas molecules (in order to i.e. create plasma).
4. Accelerate reactions (by lowering vapor pressure).
5. Create a force (such as vacuum pickup).

There is a range of different vacuums, defined by pressure.

- *Low/Rough vacuum*, 760-1 Torr: Gas flow is by viscous flow, true mechanical pressure gauges can be used. Used where gas-phase chemical reactions, momentum transfer between molecules, and/or high rate of interactions between gas and surfaces is needed.
- *Medium vacuum*, 10^0-10^{-3} Torr: The transition between low and high vacuum.
- *High vacuum*, 10^{-3}-10^{-6} Torr: Characterized by molecular flow of gas molecules. Results in very clean wafer surfaces.
- *Ultrahigh vacuum*, 10^{-6}-10^{-9} Torr: Continuation of high vacuum.

The different types of mass transport (viscous flow or molecular flow) is determined by the Mean Free Path of molecules, which is inversely proportional to gas pressure.

2.4.3 Vacuum pumps

Pumps can be categorized into two general types: roughing pumps and high vacuum pumps.

Roughing pumps are used when pumpdown starts, and remove 99.99% of air from the chamber. Two types will be discussed. *Dry mechanical pumps* use mechanical devices to remove gases, like the rotary claw pump. Nonmetallic materials are often used for moving surfaces to avoid the need of sealing or lubrication. *Blower or booster pumps* require no lubricants, and provide high gas throughput. Desirable in systems where a high volume of gas must be exhausted. An example is the lobe pump.

High vacuum pumps are used to further lower the pressure, and is often coupled with roughing pumps. The *turbomolecular pump, or turbo pump* works by mechanical compression. A number

of high speed rotating blades positioned between fixed blades impart momentum and direction to the gas molecules. It cannot pump at viscous flow. Magnetically levitating bearings are used for the blades to avoid lubrication. The most common reason for failure is dumping the pump (sudden exposure to atmospheric pressure). The *cryopump* captures gases from the chamber by lowering temperatures until they are frozen, and captured by the pump. The cryopump is extremely clean, with no oils or moving parts. The two main parts are the gaseous helium compressor and the pump module. The helium compressor works like a refrigerator, by expanding helium. The temperature of the gas is lowered down to 10K, and this cools the cryoarrays. Gases in the chamber condense and freeze on these arrays. This means gases accumulate on the arrays, and the gases must be removed periodically through *regeneration*, where the pump is warmed to room temperature or higher, and the gases are vented out. In integrated tools, loadlocks are used to load wafers, and lower the gas pressure from atmospheric to medium vacuum, before the wafers are further processed in the cluster tool.

2.4.4 Process chamber gas flow

Process chamber requirements for gas flow are:

- The ability to handle a wide variety of bulk and special gases (can be corrosive or toxic).
- Control of gas flow into process chambers.
- Control of gas mix proportions during the process run.
- Materials used in the process chamber must not be affected by the gases, and must not introduce contaminants to the gas stream.

Thoughput of gas is defined as the volume of gas flowing through a certain point of the vacuum system during a fixed period of time (the flux of gas). *Pump speed* indicates how effectively a pump can remove gases, and is expressed in the same units as throughput.

Mass flow controllers (MFCs) controls the gas flow into the process chamber. MFCs use the heat-transfer properties of gas to measure mass flow, through a thermal sensor. *Residual gas analyzers* (RGAs) identify the types of gas molecules remaining in an evacuated system. The RGA measures the partial pressure of each gas, and the total pressure of all gases. There are four basic parts to an RGA: an ionizer, and aperture, a quadrupole analyzer and a detector. The ionizer creates ions from the gas by bombarding the gas with electrons. The ions are directed towards the aperture by a different in potential. The quadrupole mass analyzer (QMA) only let through ions with a certain mass-to-charge ratio, and can therefore separate the ions, before they hit the detector. The RGA can provide real-time information about the cleanliness and stability of the process chamber during pump down.

2.4.5 Plasma

Plasma is a neutral, highly energized, ionized gas, which consists of neutral molecules, positively charged ions and free electrons. Ionization can occur when exposing the gas to strong electromagnetic fields, or by bombarding the gas with electrons. In a weakly ionized plasma, also called *glow discharge*, free electrons collide with neutral atoms, and excite them. These atoms emit light, which gives the plasma a characteristic glow. *Radicals* can be created by bombarding neutral atoms with electrons. This happens when a bond is broken, without adding or removing an electron.

In order to minimize contamination, water being the most significant source of contamination, the following steps can be taken.

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 and leaks.
5. Use low particle-generating gas-handling components.

2.5 Chapter 9 - IC Fabrication Process Overview

2.5.1 Objectives

1. Be able to 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.

Wafer fabs are generally divided into six distinct areas: Diffusion (including oxidation, film deposition and doping), photolithography, etch, thin films, ion implant and polish. The test/sort area is often located in other areas, while assembly and packaging is located in other facilities.

The diffusion bay is the area where high-temperature processing and film deposition happens. The primary tools are a *high-temperature diffusion furnace* and a *wet cleaning station*. The diffusion furnaces can operate at temperatures up to 1200°Can run oxidations, diffusion, deposition, anneals and alloy processes. The wet cleaning station is used to clean wafers and remove contamination and native oxide.

The photolithography bay is notably different, due to the yellow light. Here, the wafer is sent through the coater/developer track, a cluster tool, that primes the wafer, coats it with photoresist, spins the wafer, bakes it and performs exposure and alignment in another part of the cluster tool, the stepper. After the exposure, the wafer is sent back to be developed, rinsed and baked.

The etch process creates a permanent pattern in the wafer in areas not protected by photoresist. Common tools are plasma etchers, plasma resist strippers and the wet cleaning station. Dry plasma etch is the most used etching process. After the etch, the plasma stripper uses ionized oxygen to strip the photoresist from the wafer. Finally, the wafer is cleaned.

The ion implanter is used to dope wafers. Gases carrying the desired dopants are ionized inside the implanter, and accelerate towards the wafer to penetrate the surface in areas not protected by photoresist. After implant, the photoresist is stripped, and the wafer is cleaned.

The thin films bay is responsible for depositing dielectric and metal layers onto the wafer. Some of the systems used here are CVD, PVD, SOG (spin-on-glass system), and RTP (rapid thermal processor system), as well as the wet cleaning station.

The chemical mechanical planarization (CMP) process is referred to as the polish process. The aim is to level the surface of the wafer, and the CMP process utilizes both chemical etching and mechanical abrading to remove a desired amount of the upper layer of the wafer.

2.5.2 CMOS manufacturing steps

There are 14 steps in the typical CMOS process.

1. *Twin well process*: First, the active regions are defined. The twin-well consists of a p-well and a n-well adjacent to each other. The dopants are introduced through a retrograde implant technique to obtain optimum parameters. First, an epitaxial layer (epilayer) is grown, lightly doped with a p-dopant. Secondly, oxide is grown on the epilayer. Then the wafer undergoes photolithography to create a mask for the n-well implant. Openings in the photoresist layer allows for ion implant in the desired areas. The photoresist is then stripped, and the wafer is annealed. The p-well is made next to the n-well, through the same procedure.

2. *Shallow trench isolation process*: STI is an alternative method for creating isolation regions between active transistor regions on a substrate. There are three major steps: STI trench etch, oxide fill and oxide polish. The trench etch consists of first growing a barrier oxide on the wafer, then depositing a nitride layer. After this, the wafer is photolithographed to prepare it for the trench etch, which is done with dry plasma. After this, the photoresist is stripped, and the wafer is wet cleaned. The oxide fill is done by depositing oxide using CVD. Finally, the oxide is removed through CMP, and the nitride is stripped, leaving the trenches filled with oxide.

3. *Poly gate structure process*: The transistor gate is grown on the n- and p-well. Polysilicon (referred to as poly) is the desired material. First, the wafer is cleaned and prepared, and gate oxide is grown. Secondly, polysilicon is deposited on the surface, and doping of the polysilicon is possible if desired. The wafer is sent back to photolithography, where the pattern for the poly gates is transferred, following an etch back, leaving only the gate structures. This step is one of the most critical steps, because the gate is often the smallest structure. Therefore various quality measurements are required.

4. *Lightly doped drain (LLD) implant process*: The channel length (the region below the gate) is small, and this creates the possibility for charges to punch through the transistor, and cause undesirable channel leakage current. Therefore, the LLD implant process is done to define the source and drain regions. The LLD implant process is done with heavier dopants, at low doses, to create an amorphous region in the upper layer of the silicon. This helps maintain a shallow junction. The wafer is first photolithographed to shield the wanted areas, before the LDD implant is done.

5. *Sidewall spacer formation*: Sidewall spacers are used alongside the gates to prevent higher S/D implant to penetrate too close to the channel where S/D punchthrough could occur. The process involves depositing spacer oxide by CVD and etching the oxide with dry plasma until only the sidewall oxide layer remains.

6. *Source/Drain (S/D) implant processes*: The final step of the retrograde implant technique involves photolithographing a mask onto the wafer, and a medium-energy ion implant, that penetrates deeper than the LLD implant. The spacer oxide prevents the dopant from entering the narrow channel. This is done separately for the n- and p-transistors.

7. *Contact formation*: The wafer is cleaned, and Ti is deposited on the wafer by sputtering (PVD). The wafer is inserted in the RTP tool, and the high temperature triggers the formation of tisilicide ($TiSi_2$). Finally, the unreacted titanium is etched away, leaving tisilicide over the active regions.

8. *Local interconnect (LI) process*: The LI process involves creating metal connecting lines between the transistors and other tisilicide contacts. First a barrier layer of silicon nitride is deposited by CVD. Thereafter, LI oxide is deposited by CVD. SiO_2 is used as oxide, and the oxide is lightly doped to increase dielectric properties. An RTP step is done to allow the glass to flow and smooth the surface. CMP is done to planarize the surface, and the wafer is photlithographed, and etched back, to form the trenches where LI metal will be deposited. Ti is deposited in a thin film along the bottom and walls of the trenches by PVD, and titanium nitride is deposited as a barrier metal for the tungsten. Then, tungsten is deposited and fills the trenches and the entire wafer. Finally, the tungsten is polished down to the upper surface of the LI oxide.

9. *Via-1 and Plug-1 formation*: The interlayer dielectric (ILD) is the insulator between each metal layer. The ILD will have many small *vias*, which are small openings in the ILD, connecting the metal layers. The process starts with depositing ILD oxide by CVD. The oxide is polished using CMP. A step of photolithography and etch is done to create the via holes. The next step is depositing a thin layer of titanium by PVD as a glue to the ILD oxide, and a layer of TiN as barrier metal by CVD. Tungsten is deposited on the wafer and in the via holes by CVD, and the wafer is polished down to the upper surface of the ILD oxide.

10. *Metal-1 interconnect formation*: The metal interconnect is a three-layer metal film that connects all the vias, and is referred to as metal stack or sandwich. First, a Ti layer is deposited by PVD, because Ti bonds well with W, and with the dielectric layer. An aluminum-copper alloy is deposited with PVD, and a thin film od TiN is deposited by PVD as an antireflective coating (ARC). The wafer is patterned with photoresist, and the metal layer is etched using a plasma etcher.

11. *Via-2 and Plug-2 formation*: The formation of the ILD-2 is similar to ILD-1, except for the filling of the small voids created when etching the Metal-1 layer. This is done by either SOG and etch back, or HDPCVD (*high-density plasma chemical vapor deposition*). After the gaps are filled, the ILD-2 layer is deposited, and a patterning with photoresist and etch is done to form the Via-2 holes. New plugs are formed the same way as Plug-1.

12. *Metal-2 interconnect formation*: This procedure is repeated to form all subsequent metal stacks. Metal-2 is deposited, the same way as Metal-1. After the Metal-2 etch, ILD-3 gap fill is done. ILD-3 oxide is deposited by PECVD, and polish is done. Via-3 openings are etched, and formed the same way as Via-2 and Via-1.

13. *Metal-3 to pad etch and alloy*: After repeating the layering process for layers 3 and 4, and completing Metal-4 etch, ILD-5 layer is deposited. CMP is not necessary for this layer. ILD-5 is etched to allow Metal-5 deposition. This layer is deposited thicker than the previous layers, and etched as necessary to form bonding pads, and to remove metal

from areas where it is not needed. The final steps include one more oxide layer, ILD-6 and a final layer of silicon nitride, called the *passivation layer*. This layer is to protect the product from moisture, scratches and contamination. A final low-temperature anneal is done to improve metallurgical bonds.

14. *Parametric testing*: The wafer is tested twice to ensure its product-worthiness - once when the wafer completes the first metal etch, and aging after the completion of the final wafer fab process step. Certain electrical parameters are measured. The testing procedure is called *wafer electrical test* or WET. The wafer is also tested at the test/sort.

2.6 Chapter 11 - Depostion

2.6.1 Objectives

1. Describe multilayer metallization. Discuss the acceptable characteristics of a thin film. State and explain the three stages of thin film growth.
2. Provide an overview of the different film deposition techniques.
3. List and describe the eight basic steps in a CVD reaction, including 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 deposition.
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.

2.6.2 Film Layering Terminology

Multilevel metallization refers to the metal and dielectric layers need to interconnect the densely packed devices on a wafer. Adding metal levels is costly. *Aluminum metallization* is the use of aluminum alloy for interconnect wiring. The Al is deposited on the entire surface and etched to define the width and spacing of the interconnect lines. There is a transition to copper alloys at the moment (12 years ago). The layers of metal or ILD can be referred to as Metal-1, Metal-2 or ILD-1 and ILD-2. *Critical layers* are layers in which line widths are etched to the critical dimension (typically lower layers). These are particularly sensitive to contaminates. *Noncritical layers*, usually the upper layers, have much wider line widths. They are less sensitive to contamination.

Dielectric layers are the ILDs. The first interlayer dielectric (ILD-1) is also called the pre metal dielectric (PMD). Typically a doped silicon oxide, or glass. The important function of ILD-1 is to isolate transistor devices electrically from the metal interconnect layer and physically from contamination..

2.6.3 Film Deposition

A thin film is a thin, solid layer of material created on a substrate. The film can be either conduction, insulating or semiconducting. Some important characteristics of thin films are as follows.

- *Film-step coverage*: The ability of the thin film to maintain uniform thickness. This is desirable.
- *High aspect ratio gaps*: A small gap is characterized by its aspect ratio, which is the ratio of depth to width. The ability to fill very small gaps or holes is one of the most important film characteristics for sub-0.25 micron technology. Examples of such gaps are bias, and trenches for STI. High aspect ratio gaps make it difficult to uniformly deposit the film.
- *Thickness uniformity*: Acceptable thin films are conformal with good thickness uniformity, meaning they follow the topography of the substrate. Resistance of the layers vary with thickness. Thinner films tend to have more defects.
- *Film purity and density*: High film purity means that the film does not have any unwanted chemical elements or molecules. High density of the films an indicator of the film quality, because it means that the film is free of pinholes and voids.
- *Stoichiometry*: A desirable film will have uniform composition. The goal of the deposition is to have the correct ratio of molecules in the reaction so that the deposited film approaches the ratio of the elements in the nominal chemical formula of the incoming gases.
- *Film structure*: The film structure is critical, especially related to grain size. If the grain size varies in a film, then the film will have varying electrical and mechanical properties.
- *Film adhesion*: Adhesion to the substrate is important to avoid delamination and cracking. Adhesion is determined by cleanliness, and by the type of material the film can alloy to.

Film growth is characterized by three distinct stages. The first stage of film growth is *nucleation*, where clusters of stable nuclei form on the substrate. The second stage is *nuclei coalescence*, also referred to as island growth. Randomly oriented islands clusters grow. They continue to until they reach the third stage, which is a *continuous film*, where they meet and form a solid sheet that spreads across the substrate surface. The size of individual clusters is related to the surface mobility and nucleation rate. High mobility and/or low nucleation rate promote the formation of large clusters, and may lead to a polycrystalline film. The opposite case favors amorphous film growth. Low deposition temperatures also favors amorphous film growth. Deposited films can be amorphous, polycrystalline or single crystalline. Semiconductors are used in all three forms. Metals and dielectrics are usually amorphous or polycrystalline. Epitaxial single-crystal films are necessary for reliable semiconductor properties. To obtain such films, the film has to be deposited on a single-crystal wafer substrate.

2.6.4 CVD

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 to provide the needed energy to drive the reaction. The essential aspects are as follows.

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

There are five basic chemical reactions in CVD. *Pyrolysis* involves a compound dissociating with application of heat. *Photolysis* involves the application of radiant energy. *Reduction* is a chemical reaction that involves reacting a molecule with hydrogen. *Oxidation* involves reacting a molecule with oxygen, and *Reduction-oxidation (redox)* is a combination of the two former reactions. The choice of a particular reaction is usually defined by different parameters as required temperature, film properties etc.

The CVD reaction takes place on or very close to the wafer surface, and is a *heterogenous reaction*. This is desirable, because if the reaction takes place in the gas phase (a *homogenous reaction*), the resulting film has poor properties. The fundamental steps of the CVD reaction are as follows.

1. Gas transport to the deposition zone.
2. Formation of film precursors, through gas-phase reactions.
3. Film precursors are transported to the wafer growth surface.
4. Precursors are adsorbed to the surface.
5. Precursors diffuse on the surface to the film growth sites.
6. Surface chemical reactions happen, leading to film deposition and byproducts.
7. Byproducts are desorbed from the surface.
8. Byproducts are removed from the reactor in the bulk gas-flow away from the deposition zone.

The *rate limiting step* of the CVD reaction is the "bottleneck" in the process. The rate of the CVD reaction cannot proceed more rapidly than the mass-transport rate to the surface. This is true no matter what the temperature is. In such a situation, the reaction is said to be *mass-transport limited*. If the temperature is sufficiently low, the rate at which the reactions on the surface happen, will be lower than the mass-transport rate. In this situation, the deposition rate is *reaction-rate limited*. For these reactions, uniform temperature is especially important to achieve a uniform deposition rate across the wafer surface.

Gas-flow dynamics are important in a CVD reaction for uniform film deposition. The gas flow near the surface is zero or near-zero, and therefore it is assumed that the dominant mass-transport mechanism here is diffusion. This condition creates a boundary layer of gas flow, that increases from zero at the surface to some given gas flow in the bulk area. If the boundary layer is narrow, it can be treated as not moving, and is referred to as a *stagnant layer*.

The *pressure in CVD* is important. If the reactions happen at low pressure, the diffusivity of gas increases drastically, and the rate of deposition becomes reaction-rate limited. This means that multiple wafers can be processed simultaneously, by stacking them closely in the reactor.

Doping during CVD is desirable because it leads to important benefits. There are several dopants that will produce different results.

- *Phosphine* (PH_3): Phosphosilicate glass (PSG) is deposited. PSG creates a planar surface, with excellent gap-fill properties. Popular for ILD-1.
- *Diborane* (B_2H_6: Borosilicate glass is deposited.
- *Borophosphosilicate glass*: Adding boric oxide and phosphorous pentoxide leads to deposition of BPSG. PSG is more often used.
- *Fluorosilicate glass*: FSG is investigated as a first gen. low-k dielectric for ILD deposition.

It is important to note that adding these dopants to the silicon oxide is not the same as doping silicon. These dopants only modify the physical properties of silicon oxide. There is no donation or acceptance of electrons involved.

2.6.5 CVD deposition systems

The CVD reactor can be either *hot-wall* or *cold-wall* as mentioned earlier. The hot-wall reactor has deposition on the walls of the reactor, as well as the substrate, and requires frequent cleaning.

- *APCVD (Atmospheric pressure CVD)*: Operates in the mass-transport limited regime, which means the system must have optimum gas flow to every wafer. The system is simple, and allows for high deposition rates, however, films deposited with APCVD often exhibit poor step coverage. The APCVD is most often used for deposition of SiO_2, either with silane or with TEOS-Ozone ($Si(C_2H_5O)_4$ with O_3). TEOS is an abbreviation for tetraethylorthosilicate or tetraethyoxysilane. APCVD oxides are often doped whit phosphorous or boron.
- *LPCVD (Low pressure CVD)*: More common than APCVD because of lower cost, higher production throughput and superior film properties. Operates at a medium vacuum. The LPCVD reactors typically operate in the reaction-rate limited regime. This means that the gas-flow conditions in the reactor are unimportant. The step coverage is usually good. Hot-walled reactors are favored, so that uniform temperature control is achieved. This also means that particles will deposit on the walls, and therefore regular cleaning must be done. This can be done in situ with plasma-generated fluorine gases or with chlorine trifluoride and elevate temperatures. In situ cleaning is favored due to shorter downtime, lower particle contamination and lower personnel risks. LPCVDs can be used for a number of applications. Depositing SiO_2 can be done with TEOS or silane. LPCVD can also be used to deposit silicon nitride with good film properties, which is used as a final passivation layer. Polysilicon is usually deposited with LPCVD by the pyrolysis of silane. By adding diborane the reaction rate is raised. Silicon oxynitride films exhibit improved properties compared to silicon nitride, and can be deposited by LPCVD.
- *Plasma-assisted CVD*: This type of CVD relies on both plasma energy and thermal energy in order to initiate the chemical reactions needed for CVD. The advantages are:
    1. Lower processing temperature.
    2. Excellent gap-fill for high aspect ratio gaps.
    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 temperatures.

The PACVD happens when RF power break the bonds of gas molecules, to form radicals. These radicals readily bond to other atoms to form a film at the wafer surface. Gaseous byproducts are removed by the vacuum pumping system. The wafer is usually heated to assist surface reactions and reduce the level of undesirable contaminants. PACVD reactions on the surface of the wafer are very complex, and the specific intermediate reactions are not well understood. Therefore, deposited films are often found to be not very stoichiometric. There are two types of PACVD: PECVD and HDPCVD.

- *Plasma-Enhanced CVD (PECVD)*: Uses plasma energy to create and sustain the CVD reaction. This is a natural extension to LPCVD. The key difference is the lower processing temperature in PECVD. PECVD is performed in a vacuum chamber between parallel conducting plates. The wafer is mounted on the grounded bottom plate, and RF power is applied to the top plate. The gas flowing into the chamber develops into a plasma. Exhaust gases are pumped out at he center of the bottom electrode.The PECVD reactor is typically cold-walled. Both silicon oxide, silicon nitride and silicon oxynitride can easily be deposited using PECVD. However, PECVD does not exhibit great gap fill for high aspect ratio gaps, and may lead to voids.
- *High-Density Plasma CVD (HDPCVD)*: The HDPCVD is a high-density mixture of gases at low pressure directed towards the wafer in the reaction chamber. HDPCVD can deposit films to fill high aspect ratio gaps with relatively low processing temperatures. The reaction involves two or more gases forming gas precursors, and is used for ILDs STI, etch-stop layers and deposition of low-k dielectrics. The wafer is biased, pulling ions out of the plasma and towards the wafer. This gives directionality, and is the main reason for the HDPCVDs ability to fill gaps. One of the problems of the method is to optimize the technology for high-volume wafer fab. Due to the RF bias, the wafer may be overheated, which is undesirable. The HDPCVD uses *simultaneous deposition and etching* to fill the high aspect ratio gaps. The ratio of deposition to etching is called the *dep:etch ratio* and is approximately 3:1. The dep:etch procedure consists of five steps:

  1. Ion-induced deposition: Ions are pulled from the plasma to produce the gap-filling phenomena.
  2. Sputter etch: Energetic argon and reactant ions sputter atoms from the surface.
  3. Redeposition: Atoms are dislodged from the bottom of the gaps, and usually redeposit on sidewalls.
  4. Hot-neutral CVD: A minor contribution to deposition.
  5. Reflection: Ions reflecting off sidewalls are then deposited. This is also a minor contribution.

  The dep:etch of HDPCVD is a beneficial byproduct of the directionality of the plasma. Low pressure is important, and high gas-flow.

2.6.6 Dielectrics and performance

The *dielectric constant (k)* represents the materials effectiveness at storing potential electric energy. The lowest possible k is 1.0. Reducing the k of dielectrics reduces capacitative losses

between adjacent conductors, meaning higher speed performance. Low-k dielectrics become increasingly important for lower line widths.

For some uses (as in DRAM), a high-k dielectric is wanted, to obtain the wanted charge storage for smaller components. These dielectrics are also needed for the gate oxide, because silicon dioxide is subject to tunneling currents when the layer is too thin. This leads to current leakage and circuit failure.

Device isolation is important in MOS fab. There are two basic techniques. *Local oxidation of silicon (LOCOS)* uses patterned islands of silicon nitride to define areas for oxide growth. This technique is not acceptable for deep sub micron scaling, because of the lateral growth of oxide during silicon oxidation. STI is the preferred isolation process. It is more costly and complex, but the benefits outweigh the higher cost.

2.6.7 Spin-On-Dielectrics

Referred to as SOD or SOG (for spin-on-glass). SOGs can be organic or inorganic. After being spun on, they are cured. This process may lead to cracking and stress build up due to shrinking. SODs are considered a cost-effective alternative to CVD. New SODs exhibit good properties, and may be used in the future.

2.6.8 Epitaxy

Epitaxy is the deposition of a thin layer of single-crystal material upon the surface of a single-crystal substrate. The layer is called an epilayer. During epitaxy, doping is possible, but *autodoping* and *out-diffusion* may cause less abrupt doping transition between substrate and epilayer than desired. Epitaxy can be both homo- and heteroepitaxy. Three methods for creating epilayers on silicon wafers are VPE, MBE and MOCVD. The latter is *metalorganic CVD*, most used for ultra-thin, doped or unroped semiconductor heterolayers. It can also be used for III-V compounds.

2.7 Chapter 12 - Metallization

2.7.1 Objectives

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 using copper metallization in wafer fab. 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.

2.7.2 Types of metals and requirements

- The requirements for a successful metal material are:
  1. Highly conductive and capable of high current densities.
  2. Good adhesion to underlying layer and easily connected to external contacts and low contact resistance.
  3. Readily deposited with uniform structure and composition by a relatively low-temperature process. Deposition into high aspect ratio gaps for damascene metallization.
  4. High-res patterning or ease of planarization (damascene).
  5. Relatively soft, and ductile for reliability.
  6. High resistance to corrosion.
  7. Resistance to mechanical stress.
- *Aluminum*: The earliest interconnect metal. Used for thin films in interconnects. It has low resistivity, and adheres well to silicon oxide, by forming a layer of aluminum oxide. It is easily deposited, and etches well in solutions that do not attack underlying films. In order to form a contact between aluminum and silicon, the interface has to be heated in a process called a low temperature anneal or *sintering*. This forms an electrical interface known as an *ohmic contact*. This contact has very low resistance, and obeys Ohm's law. The resistance of the contact is however inversely proportional to the area of the interface. Therefore, in smaller circuits, high contact resistance may happen. One of the problems encountered when using pure aluminum on silicon was *junction spiking*. Both adding some silicon to the aluminum (to prevent further diffusion into the silicon) and barrier metallization was tried to prevent this. The latter was most effective, because silicon alloying in aluminum can lead to nodule formation.
- *Aluminum-Copper alloys*: Aluminum was also subject to another problem, known as *electromigration*, which lead to voids and *hillocks* in the aluminum layers. Electromigration in interconnects can be controlled by alloying the aluminum with between 0.5% and 4% copper. This increases the current-carrying capabilities of the aluminum. By adding too much copper, electromigration will actually increase. Care must also be taken when etching the metal. Copper does not etch as easily, and residual copper remaining after etching will promote corrosion. Electromigration problems can therefore be solved by barrier metallization.
- *Copper*: Copper interconnects are desirable because copper has lower resistivity than aluminum. It consumes less power due to the possibility of shorter line widths. This leads to a tighter packing density. Copper is resistant to electromigration, and fewer process steps are needed (due to the damascene process). By using copper and a low-k dielectric in wafer fab, interconnect delay can be greatly decreased. However, *there are three challenges with using copper*. It diffuses quickly into oxides and silicon. It cannot easily be patterned using regular plasma etching techniques. Copper oxidizes quickly in air at low temperatures. These challenges are overcome by using the dual-damascene process, and barrier metals.
- *Barrier metals*: A reliable method of creating ohmic contacts to shallow junctions. A barrier metal is a thin layer of deposited metal or metals designed to prevent intermixing

of the materials above and below the barrier. The essential properties of an acceptable barrier layer metal are:

1. Good diffusion barrier properties.
2. High electrical conductivity with low ohmic contact resistance.
3. Good adhesion between the semiconductor and the metal.
4. Resistance to electromigration.
5. Stability at high temperatures and when thin.
6. Resistance to corrosion and oxidation.

The *refractory metals* are often used (Ti, W, Ta, Mo, Co, Pt). For aluminum TiN and TiW are good barrier metals. TiN does not produce a good contact to silicon, so a very thin layer of Ti can be deposited first. In copper metallurgy, barrier metals are essential. Cu requires complete encapsulation by a thin-film barrier metal. These metals have to prevent copper diffusion, have low film resistivity, adhere well to both Cu and dielectrics, be compatible with CMP, be continuous and conformal with good step coverage, and allow for minimal thickness to allow copper to occupy the maximum cross-sectional area. Ta is a good candidate, as well as TaN and TaSiN. Copper barrier layers can be deposited by HDPCVD, or ionized metal plasma PVD.

- *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/metal interface. If the metal is reacted with polysilicon, it is called a polycide. They provide extremely good metallurgic contacts. To form a silicide, the refractory metal is deposited on the silicon wafer, and a high temperature anneal is done. $TiSi_2$ was the most common silicide used, serving as a contact between the silicon and the tungsten plug. Tisilicide forms two phases, C49 and C54. C54 is the desirable phase. For future, smaller technology cobalt silicide seems promising, due to smaller grain sizes. Silicides are not barrier metals. A *salicide* is a self-aligned silicide. Because salicide only forms on silicon surfaces, the need to pattern and align is avoided.
- *Metal plugs*: Usually made of tungsten (W). Tungsten uniformly fills high aspect ratio gaps by CVD methods. It is resistant to electromigration, and serves as a barrier. Aluminum would be desirable because of lower resistivity, but cannot fill high aspect ratio gaps in the same way.

2.7.3 Metal deposition systems

There are four different metallization methods used for traditional and dual damascene metallization.

Evaporation is a PVD process, today replaced by sputtering. The material to be deposited is placed in a crucible, and heated in a vacuum chamber until it evaporates. An electron beam could be used for heating. Maintaining a high vacuum means the Mean Free Path of evaporated molecules is greatly increased, and the molecules can travel freely into the chamber until they hit a surface, where they condense to form a film. This does not produce a uniform step coverage. Alloys are not easily evaporated, and the resulting film is not stoichiometric.

Sputtering, a form of PVD, took over. It is mainly a physical process. The advantages are: The ability to deposit and maintain complex alloys, the capability to deposit high-temperature refractory metals and controlled, uniform films on large wafers. The basic steps are:

1. Positive argon ions are created in a plasma in a high vacuum chamber and accelerated towards a target at a negative potential.
2. The ions gain momentum during acceleration.
3. The ions physically dislodge (sputter) atoms from the target, with the desire material composition.
4. The sputtered atoms migrate to the substrate surface.
5. The sputtered atoms condense and form a thin film on the substrate, with essentially the same composition as the target.
6. Excess material is removed from the chamber by vacuum pump.

Vacuum conditions are important to create the needed plasma, and maintain purity of the deposited film. The physics of the process are quite simple. Positive argon ions are created in the plasma, and drawn towards the target. The atoms of the target are sputtered due to the momentum transfer. Argon is used because it is heavy, and inert. The sputtered atoms travel towards the surface, and form a thin film as mentioned. The *sputtering yield* is defined as the number of atoms ejected per incident ion that strikes the target. Normal values are between 0.5 and 1.5. This number depends on the incident angle of the bombarding ions, the composition and geometry of the target and the mass and energy of the bombarding ions. The target is eroded, and must be changed periodically. The sputter gas has to be ultrahigh purity, so that no impurity gas ions are incorporated into the growing film. Three types of sputtering will be described here.

- *RF Sputtering*: An RF field is used to create plasma instead of the DC field above. The RF sputtering system is limited due to low sputtering yield, and therefore low deposition rate.
- *Magnetron Sputtering*: Magnets configured around and behind the target capture and restrict the electrons in front of the target. This increases the bombardment rate on the target, which produces more secondary electrons, which again increases ionization of the plasma. The result is more sputtering of target, and higher deposition rates. Strict vacuum conditions are necessary to ensure contamination. This leads to a Mean Free Path of sputtered atoms long enough for atoms to travel linearly to the surface of the wafer, and therefore sidewalls and bottoms of high aspect ratio gaps are not uniformly covered. *Collimator sputtering* can solve this, by introducing a collimator in the reactor. Any sputtered atoms with high angles are collected by the collimator, which means the atoms will travel straight to the substrate, and the bottom of gaps will be covered. Sidewalls will not be covered in the same way.
- *Ionized Metal Plasma (IMP*: Sputtered metal is ionized in an RF plasma at low pressure. The ionized metal particles travel towards the substrate with high directionality. The substrate is configured with a negative voltage bias. This leads to better results in high aspect ratio gaps.

Metal CVD gives superior conformal step coverage and void-free filling of high aspect ratio gaps. Tungsten CVD is deposited using LPCVD or PECVD with excellent step coverage and gap-fill.

Tungsten is deposited in a blanket, and CMP is done. Copper CVD has been investigated in order to create a conformal seed layer in order to do copper electroplating.

Copper Electroplating is also known as *electrochemical deposition (ECD)*. The basic technique involves immersing the wafer in a solution of copper sulfate, and a voltage is applied to the wafer in order to deposit copper. In practice, this technique is difficult to control, because the current density has to be uniform across the surface in order to obtain a uniform film. By applying an oscillating electric field a dep/etch sequence can be achieved, which can provide better uniformity.

2.7.4 Metallization schemes

The traditional aluminum metallization technique is described in chapter 9. For copper, the *dual-damascene process* is used. The basic process steps are:

1. SiO_2 is deposited as ILD oxide.
2. SiN is deposited on the surface of the oxide as an etch-stop.
3. The SiN is patterned for via formation by photolithography and etching.
4. Another layer of oxide is deposited.
5. Interconnect patterning is done by photolithography.
6. The trench for interconnect metal and hole for vias are dry etched, stopping on the SiN layer.
7. A barrier metal is deposited across the entire surface.
8. A copper seed layer is deposited by CVD.
9. Copper fill is deposited by ECD. Both via and trenches are filled.
10. Excess copper is removed by CMP, preparing the surface for the next level.

The most important reason for the damascene process with copper, is the avoidance of metal etching. The process also results in fewer process steps, and eliminate some of the most challenging steps of the traditional process, including aluminum etch and many of the tungsten and dielectric CMP steps.

2.8 Chapter 13 - Photolithography: Vapor Prime to Soft Bake

2.8.1 Objectives

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 photoresist.
3. State and explain the eight (ten) basic steps of photolithography.
4. Explain how the wafer surface is prepared for photolithography.
5. Describe photoresist and discuss photoresist physical properties.

6. Discuss the chemistry and application of conventional i-line photoresist.
7. Describe the chemistry and benefits of deep UV resists, including chemically amplified resists.
8. Explain how photoresist is applied in wafer fab.
9. Discuss the purpose of soft bake, and explain how it is accomplished in production.

Photolithography produces as three-dimensional pattern on the surface of the wafer using a light-sensitive photoresist material and controlled exposure to light. The process is central to wafer fab, because the wafers repeatedly flow in and out of the photolithography step. It is often considered the most critical step in the IC process. A *reticle* is a quartz plate that contains the pattern for one or several dies to be reproduced on the wafer. A *mask* has the entire array of patterns required for one an entire wafer layer.

Critical dimension generations are for instance the 0.18μm-generation or the 0.1μm-generation. The critical dimension in photolithography is often used to describe device technology generations.

The electromagnetic spectrum, or *light spectrum* is used to introduce the UV light spectrum of most interest in photolithography. Different wavelengths describe different types of UV light used for photolithography. The important UV wavelengths are summarized in Table 1.

Table 1: Important UV wavelengths for photolithography exposure.

UV wavelength (nm)	**Wavelength name**	**UV emission source**
436	g-line	Mercury arc lamp
405	h-line	Mercury arc lamp
365	i-line	Mercury arc lamp
248	Deep UV (DUV)	Mercury arc lamp or KrF excimer laser
193	DUV	ArF excimer laser
157	Vacuum UV (VUV)	F_2 excimer laser

An important performance measurement is the *resolution* of each image. The minimum feature size dimension is the CD. Resolution is important for CD. The wavelength of the exposing light has to be at most about the same length as the CD:

Overlay accuracy is a measure of the degree of alignment between the pattern on the mask and the existing features on the wafer. *Process latitude* describes the capability of the photolithography process consistently produce products that meet the specified requirements.

There are basically two types of photolithography processes. *Negative lithography* prints a pattern on the wafer opposite to the pattern on the mask. This means that the photoresist exposed to light becomes hard and insoluble through cross linking. These photoresists are called negative resists. The dark portion of the mask is made of chrome. *Positive lithography* prints a pattern which is the same as the pattern on the mask. Exposed areas of the photoresist are made soluble. Masks can be described as *clear-field* or dark-field depending on the amount of mask covered by chrome. If a positive resist requires a certain clear-field mask, then a dark-field mask of the same pattern is required for a negative resist.

2.8.2 Eight basic steps of photolithography

The eight steps are summarized in Table 2.

Table 2: The eight steps of photolithography.

1. Vapor prime
2. Spin coat
3. Soft bake
4. Alignment and exposure
5. Post-exposure bake (PEB)
6. Develop
7. Hard bake
8. Develop inspect

Vapor prime: This is the first step of the photolithography process, and consists of three minor steps. The aim of this process step is to prepare the wafer for the subsequent process steps. First, the wafer is cleaned, often using acetone and ethanol. Secondly, the wafer is dried during a dehydration bake, to drive off residual moisture. This promotes good adhesion between the substrate and the photoresist, which is applied later. The dehydration bake can be done in a convection oven, or on a hot plate, and the temperature is normally kept between 200 and 250°C. After the dehydration bake, the wafer can be primed using *hexamethyldisilazane* (HMDS), to promote adhesion between the substrate and the photoresist. This step is not always done.

Spin coat: The application of the photoresist is done by spin coating, a technique where the liquid photoresist is spun off the wafer, to ensure a uniform layer with desired thickness. The *photoresist* is an organic compound that, during exposure to ultraviolet (UV) light, changes solubility in a developer liquid. There are mainly two types of photoresist: positive and negative photoresist. Positive photoresist is initially not soluble in the developer liquid, but becomes more soluble after exposure to UV light. Negative photoresist is initially soluble, and hardens through cross linking during the exposure. These two types of photoresist exhibit different physical attributes, and depending on the requirements for the final product, one can choose the optimal photoresist. The method of spin coat can be summarized in four steps.

1. **Dispense**. The liquid photoresist is dispensed onto the wafer.
2. **Spin-up**. The wafer is spun to spread the photoresist uniformly over the wafer. This is done by quick acceleration to a high rpm speed.
3. **Spin-off**. Excess photoresist is spun off to obtain a film which is as uniform as possible.
4. **Solvent evaporation**. The wafer is spun at a constant rpm until most of the solvent is evaporated, and the photoresist is nearly dry.

Soft bake: After the spin coat is done, the wafer is normally heated in a process called *soft bake*. This process step is done to drive off the residual solvent from the photoresist, improve the adhesion between the substrate and the photoresist, and relive stresses in the resist film. The soft bake can be done on a hot plate, and the temperature is often kept between 85 and 120°C.

Alignment and exposure: During this process step, the wafer is placed in the optical system

used to expose the wafer to UV light. These systems consist of a light source, an optical system (i.e. different lenses and mirrors), a mask with the desired pattern to be transferred to the wafer and an alignment system. This process step is very important, due to the fact that most ICs consist of multiple layers, and errors in the alignment can easily lead to failure. Different light sources can be used to achieve different dimensions of the printed pattern. In general, the shorter the wavelength of the light used, the smaller dimensions can be achieved.

Post-exposure bake (PEB): After the wafer has been exposed to the UV light, a short PEB can be done to catalyze the chemical reactions needed in the photoresist, in order to make it soluble in the developer liquid. This is not always required, but must in general be done when using the newer chemically amplified photoresists. The PEB also improves adhesion between the substrate and the photoresist.

Develop: A liquid chemical developer is used to dissolve the soluble regions of the photoresist on the wafer. This is done to accurately replicate the initial pattern on the wafer. The developing process must be controlled carefully to avoid patterning problems. If the wafer is under- or overdeveloped, the resulting pattern may lead to a failure in the final product.

Hard bake: After development, a thermal bake (referred to as *hard bake*) is done to evaporate any residual solvent and harden the remaining photoresist. The hard bake also improves the adhesion between the wafer and the photoresist. This is beneficial for the subsequent process steps. Hard bake is usually done on a hot plate, with temperatures between 130 and 150°C, depending on the kind of photoresist used (negative photoresist often requires higher temperatures than positive photoresist).

Develop inspect: This is the final step of the photolithography process, where the printed pattern is inspected to find potential defects. If a wafer is defective, it should not be further processed. If a defective wafer is detected, it can be stripped and cleaned, and the photolithography process can be repeated. The post-development inspection can be done with optical microscopy, if the dimensions of the sample are not too small, in which case, other methods (i.e. electron microscopes) must be utilized.

The *vapor prime* step is important because wafers are required to be clean throughout the entire wafer fabrication process. Contaminants on the surface prior to photolithography can lead to poor adhesion, and lifting during development. Particulate contamination can lead to pinholes and uneven resist coating. Priming of the wafer can be done in different ways, the most common being *puddle dispense and spin* and *spray dispense and spin*. An advantage of the spray approach is that the spray will assist in particle removal from the surface.

2.8.3 Spin coat and photoresist

The goals of photoresist are transferring the mask to the resist top layer on the wafer, and protecting underlying layers during subsequent processing. Improvements in photoresists provide better resolution, adhesion uniformity and increased process latitude.

Photoresist physical properties: A particular resist is chosen on the basis of the following properties.

- Resolution. The ability to differentiate between closely spaced pattern features.
- Contrast. The sharpness of the transition from exposure to non-exposure in photoresist. High-contrast is desirable to produce vertical resist sidewalls.

- Sensitivity. The minimum amount of light needed at a certain wavelength to produce a good pattern. This amount is called the *exposure dose*.
- Viscosity. Describes the flow characteristics of the resist. By having a viscous resist, uniformity and better etch resistance is achieved, but the resist will be thicker.
- Adhesion. The resist must adhere to many types of surfaces.
- Etch resistance. The resist must maintain adhesion through etching, either dry or wet.
- Surface tension. Intermediate surface tension to ensure good flow and wafer coverage, as well as the resist molecules holding together during subsequent processing.
- Storage and handling. The photoresist is sensitive to energy, and must be stored and handled thereafter. Improper treatment of resist will result in poor properties of the resist.
- Contaminants and particles. The purity of the resist material is important.

Conventional i-line photoresists are optimized for i-line UV light. They can be positive or negative. The positive resists are composed of a resin, a sensitizer, a solvent and sometimes other additives. The resin is an inert organic polymer matrix, while the sensitizer is the photosensitive component. In negative photoresists, the resin is cross linked when the sensitizer is exposed to light (the sensitizer creates radicals that initiate the cross linking reaction). One certain positive photoresist is called *DNQ-novolak*.

Deep UV (DUV) photoresists are based on *chemical amplification (CA)* of the photoresists. This means that the resist is more sensitive to light than the standard i-line resists. The CA resists are boosted by adding a sensitizer called *photoacid generator (PAG)*, that produces an acid upon exposure to light. This acid makes the resist soluble in the developer. DUV resists are very sensitive to contaminations, especially amines, which are found in the ambient atmosphere. Therefore, the air in all processes involving CA DUV resists, must be chemically filtered.

Photoresist dispensing methods: The most common method is spin coating. This method as four steps, as mentioned above. The goal is to have a uniform film coating on a wafer, and to achieve repeatable film thickness from wafer to wafer. The spin coat procedure is highly automated in wafer fab. A wafer is placed on the chuck, which uses vacuum to hold it. The dispense nozzle dispenses resist. If the resist is dispensed while the wafer is still, it is called static dispense. Otherwise, the process is known as dynamic dispense. The wafer is spun, first slowly to distribute resist, and then it is accelerated to achieve desired thickness (typically to around 4000 rpm). Resist thickness is found to be vary as

$$\text{Resist thickness } \alpha \ \frac{1}{(\text{RPM})^{1/2}} \tag{11}$$

There are several properties to control in order to maintain the correct thickness and uniformity. Acceleration ramp, temperature, humidity, exhaust and particulate contamination are some. When wafers are spun, the resist is forced towards the edges, and onto the backside. This ridge is called the edge bead, and must be removed. Normally, some solvent is sprayed on to remove the edge bead. Care must be taken, so that not too much resist is removed.

Soft bake: The goals of soft bake have already been stated. If soft bake was not done, the following problems would occur:

1. The resist film would be sticky and susceptible to particulate contamination.

2. Inherent resist stresses from spin coating would lead to poor adhesion.
3. High solvent level would lead to inadequate distinction between exposed and unexposed resist.
4. Outgassing from resist due to high solvent level could contaminate the optical system during exposure.

The preferred equipment for soft bake is heat conduction from a wafer on a vacuum hot plate.

2.9 Chapter 14 - Photolithography: Alignment and Exposure

2.9.1 Objectives

1. Explain the purpose of alignment and exposure in photolithography.
2. Describe the properties of light and exposure sources important for optical lithographic.
3. State and explain the critical aspects of optics for optical lithography.
4. Explain resolution, describe its critical parameters, and discuss how it is calculated.
5. Discuss each of the five equipment eras for alignment and exposure.
6. Describe reticles, explain how they are manufactured, and discuss their use in microlithography.
7. Discuss the optical enhancement techniques for subwavelength lithography,
8. Explain how alignment is achieved in lithography.

The optical system in photolithography basically consists of an UV light source, an optical system, a reticle with the die pattern, an alignment system and a wafer covered with photoresist. The wafer stepper has three purposes: To focus and align the wafer surface to the quartz plate reticle, reproduce a high-res reticle image on the surface through exposure and produce an adequate number of wafers per unit time.

2.9.2 Optical lithography

Optical lithography has traditionally been the limiting factor of the CD in IC fab. There are other methods to achieve smaller CD, but they will not be discussed.

A light source is needed for optical lithography. Due to the wave nature of light, *interference of light waves* must be taken into account. Interference can be either constructive or destructive. *Optical filters* use light interference to block unwanted incident light through both reflection and interference, to obtain light with a certain wavelength. The two light sources most used for UV light are the mercury arc lamp and excimer laser.

- *Mercury arc lamp*: This lamp is used for the conventional i-line steppers. An electrical current is passed through a high-pressure tube of mercury-xenon gas to create a discharge arc. The characteristic light spectrum emitted has intensity peaks corresponding to g-line, h-line and i-line UV light. An important aspect of light sources is *light intensity*, which determines the amount of time needed to deliver a certain amount of radiant energy to the target. Lower intensity means higher exposure time. It is also important to modify

the absorbance of the resist, so that light intensity at the bottom of the photoresist is not significantly lower than at the top.

- *Excimer laser*: The main benefit of excimer laser has been to provide more light intensity at wavelengths shorter than i-line (DUV and so on). The excimer laser comes from the expression *excited dimer* laser, which uses an unstable complex of a noble gas and a halogen to emit light. The complex emits light through decomposition. A high-pressure mix of gases are excited by applying a high-voltage pulse discharge across two flat-pad electrodes. The excimer lasers create short pulses of light with high peak intensity. This can damage optical material. Longer pulses are desired. The light from excimer lasers are not very spatially coherent. This can cause diffraction, but is controlled to avoid diffraction.

Optics: The *law of reflection* says that the angle of incident for a light beam is the same as the angle of reflection, $\theta_i = \theta_r$. When light passes through one transparent medium into another, it changes direction. This is called *refraction*. The relative index of refraction, n, represents how much the light bends when it passes through the interface of two media, based on changes in velocities. The absolute index of refraction compares the speed of light in vacuum to the speed in the chosen medium. n is defined as

$$n = \frac{\sin\theta_i}{\sin\theta_r} \tag{12}$$

A *lens* is an optical element that refracts light from an object passing though it to form an image of the object. The lens will make the light either converge (convex lens) or diverge (concave lens). Lenses are originally made of glass, but for DUV and VUV light, other materials must be used because ordinary glass absorbes at these wavelengths. *Lens compaction* can happen if the absorption of light is significant. The material of the lens becomes denser, and the refractive index of the lens is changed. Such usage flaws are called *aberrations*. *Diffraction* of light can happen when light passes through narrow slits, or past a sharp edge. Interference patterns are created. Diffraction can be a problem for small features in photolithography. The ability of a lens to capture diffracted light is called its *numerical aperture*. It can be approximated as the index of refraction times the radius of the lens divided by the focal length of the lens.

Antireflective coating (ARC) is a coating applied to prevent reflection of light off the surface of the wafer. This prevents *reflective notching* and the effects of *standing waves*. There are basically two types of ARCs: bottom ARCs deposited beneath the resist, and top ARCs deposited on top of the resist to prevent secondary reflections from the surface of the resist. The bottom ARC is most effective. The BARCs can be organic or inorganic. The organic BARCs absorb light, while the inorganic BARCs work by phase-shift cancellation of light. The BARCs are dielectrics spun on before the photoresist. ARCs must also be easy to remove, if needed. The TARCs use destructive interference to eliminate secondary reflection from the surface of the resist.

Resolution is defined as the ability to discern pairs of closely spaced features on the wafer. The formula for resolution is given in Jenkins ($R = k\lambda/NA$). The formula shows that decreasing wavelengths lead to improved resolution. However, shorter wavelengths lead to a decrease in *depth of focus (DOF)*. DOF is the range around the focal point in which the image is continuously in focus. The center of focus (COF) is the point from the center of the lens where the best imaging occurs. The equation to describe the DOF is

$$\text{DOF} = \frac{\lambda}{2(\text{NA})^2} \tag{13}$$

. Resolution and DOF are competing parameters. If the wafer is not very flat, or the equipment vibrates, the results will be poor.

2.9.3 Photolithography equipment

The five lithography eras are defined by the equipment used.

1. **Contact aligner**: The primary method used until the 1970s. The mask for a contact aligner has all the die patterns needed to process a wafer. The wafer is coated with photoresist, and both the mask and wafer are viewed simultaneously in a microscope having split vision optics. The mask pattern is aligned to the wafer, with an alignment stage through manual operation. The mask is brought in contact with the wafer, before exposed. This system was prone to contamination, and the mask had to be replaced every 5-25 process run. Overlay accuracy was also a problem. The contact aligner produces good image resolution, because the mask and wafer are so close. Reduced image distortion. Very operator dependent.

2. **Proximity aligner**: Similar to the contact aligner. The mask does not make contact with the resist, but is placed close to the wafer (2.5-25 μm). Less contamination, but light scattering leads to lower resolution.

3. **Scanning projection aligner**: Also called scanner. It projects a full mask with 1:1 image ratio onto the wafer using a mirror system.The UV light is focused onto the wafer through a narrow slit, allowing a uniform source of light. The mask and wafer are mounted on a scanning carriage, moving in unison across the narrow beam of light. A major challenge was making a good 1X mask containing all the dies for sub micron features.

4. **Step-and-Repeat aligner (Stepper)**: The stepper projects only one exposure field (may be one or more chips), before stepping to the next location on the wafer to repeat exposure. The stepper uses a reticle. One of the advantages is the ability to use a reduction lens, meaning the reticle are 4X, 5X or even 10X larger than the patterned image. At each step, the stepper will focus the wafer and reticle to the projection lens, align the wafer and expose it, before stepping to the next location. Compensations for variations in wafer flatness and geometry can easily be done because only a small portion of the wafer is exposed at a time. Steppers expose small areas at a time, and are costly.

5. **Step-and-scan system**: A hybrid tool from scanning projection aligners and steppers. A focused slit of light is scanned across the reticle and wafer, and a reduction lens is used. After one scan, the system steps to the next location. This gives increased exposure field for large wafers. The reticle is scanned through a reduction lens, allowing for smaller optical systems. Because the reticle is large, more die can be placed on a single reticle. Focus can be adjusted through the scan, permitting compensation for lens defects and wafer flatness. This yields improved CD control. The major challenge is is the increased demand on mechanical tolerance. The system has to precisely move the both wafer and reticle simultaneously in opposite directions. Position tolerances are within tens of nanometers.

Reticles: A reticle is a transparent plate with a pattern image that will be transferred to the wafer. The reticle is used for the stepper and step-and-scan system, and must be perfectly manufactured. If the reticle has defects, all dies will have the same defects. The primary material for submicron lithography is fused silica, because of the high transmission in DUV. It is expensive, and has low thermal expansion. Chrome is usually used for the opaque areas.

The common size ratio is 4:1 or 5:1 for reticle:die. The common method for creating the reticle pattern is by e-beam lithography. The method is similar to optical lithography. Chrome is deposited, and lithography is used for creating the etch pattern. E-beam is more reliable, and has high-res and better dimensional control. It is time consuming and complex, and is therefore not used for wafers in general. The sources of reticle damage are usually dropping or scratching the reticle, electrostatic discharges or particles of dirt. A pellicle film can be introduced to keep contaminants away.

Optical enhancement techniques: Subwavelength lithography permits a wafer to be patterned with a resolution slightly below the light exposure wavelength. There are three main techniques:

- *Phase-shift mask (PSM)*: The reticle is modified with an additional transparent layer so that alternating clear regions diffract the light 180°. The light diffraction underneath the opaque area is reduced. Absortive phase shifters can be used on the chrome mask.
- *Optical proximity correction (OPC)*: Introducing alterations into the reticle pattern to compensate for optical proximity effects (as diffraction). Producing reticles with this level of control is very complicated.
- *Off-axis illumination (OAI)*: By having the incident light strike the mask at an angle, the diffraction fringes can be aligned with the lens. This reduces the resolution limit and increases DOF.

Print bias is the difference between the feature dimension on a reticle, and on the printed wafer. By, for instance, adding serifs to the mask, round corners can be avoided.

Alignment is very important. First the reticle is aligned to the system, known as BLC (baseline compensation), thereafter the wafer is aligned to the reticle. Overlay accuracy is important in order to create functioning systems. For steppers and step-and-scan systems each reticle pattern is aligned and exposed at multiple locations. Alignment marks are used, and the system is automated.

During alignment and exposure, several parameters are important to control, as temperature, humidity, vibration, pressure and particle contamination.

2.10 Chapter 15 - Photolithography: Photoresist Development

2.10.1 Objectives

1. Explain why PEB is done for conventional and DUV resist.
2. Describe the negative and positive resist development process for conventional and CA DUV resist.
3. List and discuss the two most common resist development methods and the critical development parameters.
4. State why a hard bake is done after resist development.
5. Explain the benefits of a post-develop inspect.

2.10.2 Post-exposure bake

The PEB is done to catalyze critical resist chemical actions in CA resists. It also improves adhesion, and reduces standing waves. For the CA DU resists, it is during the PEB that the resist becomes soluble. The acid produced during exposure causes the deprotection reaction during PEB. The requirements are critical. Temperature uniformity is very important, and PEB time. It is done on a hot plate, and the parameters depend on the resist. Early DUV resists were sensitive to PEB delay. If the time from exposure to PEB was more than a few minutes, the acid would be neutralized by ambient amines in the atmosphere. Recent resists permit a delay up to 30 minutes. It is desirable to complete PEB as soon as possible after exposure. For i-line resist, PEB is done to reduce standing waves.

2.10.3 Develop

The primary goal of development is to accurately reproduce the pattern from the mask in the photoresist. If the CD meets the specifications, all other features are assumed acceptable. There are three main problems with development. Underdevelopment, incomplete development and overdevelopment. Underdevelopment can result in sloping sidewalls. Incomplete development results in residual resist on the substrate where it should have been removed. Overdevelopment removes too much resist, causing narrower features and poorly-defined lines.

Very little chemical reaction is needed for *negative resist development*. The procedure is mainly a solvent wash of the unexposed resist. The solvent is typically organic. Swelling and distortion of the cross linked resist may happen. This is the main reason why negative resist is not used for geometries below 2μm.

Positive resist development involves chemical reactions between the resist and the developer. The rate at which resist dissolves is known as the *dissolution rate*. High rate is desirable, but not too high, because it may lead to problems. High *developer selectivity* is also desirable. This means that the developer reacts fast with the unexposed resist relative to the exposed resist. The positive developers are based on strong alkaline diluted with water. TMAH is a good developer (tetraethyl ammonium hydroxide). There is no problem with swelling. For conventional i-line resists, carboxylic acid is produced during exposure. The TMAH neutralizes this acid and dissolves the exposed resist.

There are two common *development methods*.

- *Continous spray development*: A single wafer is positioned on a vacuum chuck, and spun at a slow speed, while one or more nozzles spray developer on the resist covered surface. The developer is dispensed in a fine mist. This is a good method for controlling temperature. Spray pattern and speed of rotation are important parameters to control to achieve uniformity and repeatability.
- *Puddle development*: The same basic equipment is used. An appropriate amount of developer is dispensed as a puddle on the surface. Excess developer can be spun off. Multiple dispenses of developer can be done to replenish chemical agents. Low flow of developer is kept to reduce variations. This method minimizes temperature gradients and permits control of the variables affecting uniformity.

The critical development parameters are:

1. **Developer temperature**: Optimum temperature is between 15°Cand 25 °C. For positive resists, faster development happens at lower temperatures, opposite of negative development.
2. **Developer time**: The developer continues to react with the resist until it is removed. In automized wafer processing, an in-situ *dissolution rate monitor (DRM)* is used to monitor the reaction. Interferometric signal data is collected, and this gives information on the end point (EP).
3. **Developer volume**: The amount is critical. An inadequate volume can give *scumming*, which is a residue film on the wafer surface. Excessive use is a cost issue.
4. **Normality (N)**: Refers to the stoichiometry in a solution.
5. **Rinse**: Typically done with DI water. It serves to stop the development process.
6. **Exhaust flow**: Developer mist in the wafer track machine can result in overdevelopment.
7. **Wafer chuck**: The wafer has to be held perfectly level, to ensure uniform coverage during puddle development.

2.10.4 Hard bake

Post-development thermal bake to evaporate any residual solvent and harden the resist. The temperature can be elevated to just below the boiling temperature of the resist, effectively drying the wafer. Usually, the temperature is held between 130-150°C. If the temperature is too high, the resist may flow, and cause deformations in the pattern. For DUV resists and cross linked resists can withstand higher temperatures, up to 210°C. This is beneficial for plasma etching and ion implant processes, where temperatures can exceed 150°C.

2.10.5 Develop inspect

In order to identify certain defects, inspection is done. If a wafer has a defective pattern, and is further processed, it becomes scrap. Otherwise, it can be stripped and sent through photolithography again. Traditionally it has been a manual, operator-intensive inspection with optical microscopy. Automated inspection is more and more common in advanced fabs. The goal is zero defects, but many fabs have about 2% rejects. If this is higher than 4%, there is a quality problem that requires corrective action.

2.11 Chapter 16 - Etch

2.11.1 Objectives

1. List and discuss nine important etch parameters.
2. Explain dry etch, including its advantages, and discuss how etching takes place.
3. List and describe the equipment systems for seven dry etch reactors.
4. Explain the benefits of high-density plasma (HDP) etch, and discuss four types of HDP reactors.

5. Give an application example for dielectric, silicon and metal dry etch.
6. Discuss wet etch and its applications.
7. Explain how photoresist is removed.
8. Discuss etch inspection and its related important quality measures.

Etch is the process of selectively removing unneeded material from the wafer surface by using either chemical or physical means. The goal is to accurately reproduce the mask features on the resist-coated wafer. The resist is not attacked significantly. *Dry etch* exposes the wafer to a plasma created in the gaseous state. This is the primary method for etching wafers with sub micron features. Dry etch is used for *dielectric etch, silicon etch and metal etch.* The etch process can be patterned or unpatterned. Patterned etch involves a masking layer, and unpatterned etch is mainly done in the stripping of layers.

2.11.2 Etch parameters

- **Etch rate**: The speed at which material is removed during etching. Given by Eq. 14

$$\text{Etch rate} = \frac{\Delta T}{t} \tag{14}$$

  Where ΔT = the amount of material removed in Ångstrom, and t = the time elapsed during etching, usually in minutes or seconds. Loading effects refer to depletion of etchants if the area of etching is large, or surplus of etchants for small areas, which lead to differences in etching rates.

- **Etch profile**: Refers to the shape of the sidewall in the etched feature. Two basic profiles: isotropic and anisotropic. Isotropic etch profiles etch in all directions at the same rate, leading to undercut sidewalls. Usually exhibited in wet etch. Can lead to undesirable etch profiles. Anisotropic etch profiles are achieved when the etching is only perpendicular to the wafer surface, with very little lateral etching. Achieved with most dry etches. Directionality improves the etch profile, especially in high aspect ratio gaps. Achieved with HDP etching.

- **Etch bias**: A measure of the change in line width of a CD after performing an etch process. Usually caused by undercutting. Etch bias is given by Eq. 15.

$$\text{Etch bias} = W_b - W_a \tag{15}$$

  where W_b = the original line width in the photoresist before etch, and W_a = the final line width in the etched material after resist removal.

- **Selectivity**: Represents how much faster one film etches relative to another under the same etch conditions. High selectivity means that only the desired layer is etched. Selectivity, S_R, is defined as

$$S_R = \frac{E_f}{E_r} \tag{16}$$

  where E_f = the etch rate of the film undergoing etch, and E_r = is the etch rate of the masking layer. Poor selectivity can be for instance 1:1, whereas good selectivity can be 100:1.

- **Uniformity**: A measure of the capability of the process to etch evenly across the entire surface, the wafer lot and from lot to lot. Closely related to selectivity. In small features, etch rate can be limited. Referred to as *aspect ratio dependent etching (ARDE)* or micro loading. The goal is to minimize ARDE.
- **Residues**: Unwanted material remaining on the wafer surface after etch. Can come from contaminants in the chamber or in the film, improperly chosen etch chemicals and nonuniform dopant distribution. Sometimes an over etch is done at the end of the etch process to strip all residues.
- **Polymer formation**: A polymer formation is sometimes intentionally deposited on the sidewalls of the etch features to form an etch-resistant film on the sidewalls, increasing the directionality of the etch. The polymer has to be removed after etch. This can be difficult. The process chamber must also be stripped.
- **Plasma-induced damage**: Nonuniform plasma can create trapped charges on the gate electrode, causing a breakdown of the gate oxide. Energetic ion bombardment of the gate oxide. Can be removed through anneals or wet etching.
- **Particle contamination**: Can come from plasma near the wafer surface. Controlled by optimized tool design, proper tool operation and shutdown.

2.11.3 Dry etch

Advantages are given in Table 3.

Table 3: Advantages of dry etch over wet etch.

1.	Etch profile is anisotropic with excellent control of sidewall profiles.
2.	Good CD Control
3.	Minimal resist lifting or adhesion problems.
4.	Good etch uniformity within wafer, from wafer to wafer, and lot to lot.
5.	Lower chemical costs for usage and disposal.

The disadvantages are poor selectivity, risk for device damage from plasma and expensive equipment.

The process involves a low-pressure plasma discharge to remove material. The major actions are as follows.

1. Etchant gases enter chamber.
2. Dissociation of reactants by electric fields.
3. Recombination of electrons with atoms to create plasma.
4. Reactive +ions bombard the surface creating anisotropic etch action.
5. Adsorption of reactive ions on surface.
6. Surface reactions of radicals and surface film, creating isotropic etch action.
7. Desorption of by-products.

8. Removal of by-products by vacuum pump.

Etch action is achieved by physical processes, chemical processes or a combination. Physical etching is analogous to sputtering, and is highly anisotropic. Chemical etch is isotropic in profile. A combination of the two is often used, because it produces good CD control with fair selectivity. Dry etch systems can be configured to create isotropic or anisotropic etch profiles. The *plasma potential distribution* refers to the dark space potential difference.

2.11.4 Plasma etch reactors

The basic components of a plasma dry etch include a reaction chamber, an RF power supply source, a gas flow control system and a vacuum system.

- **Barrel plasma etcher**: A cylindrical design with almost pure chemical isotropic etching at low pressure. The wafers are mounted vertically in a quartz boat with small separation between the wafers. RF power is applied from each side of the cylindrical etch tunnel. The RF field is parallel to the wafers, to create isotropic chemical etching. Minimal plasma-induced damage because there is no physical sputtering. Primarily used for stripping photoresist, using oxygen plasma. High selectivity.
- **Parallel plate (planar) reactor**: Two parallel plates, symmetrical in shape and size and position. A wafer is placed on the grounded electrode, with RF signal to the upper electrode. This is the plasma etch mode, with energetic ion bombardment. If the wafer is placed on the RF electrode, the wafer is in direct contact with the plasma, and is in the reactive ion etch (RIE) mode. A reactive gas is needed, as fluorine or oxygen. Low pressure and high RF power is used to control the etch rate.
- **Downstream etch systems**: Exposure to ion bombardment can damage the device. To locate the wafer etch region away from the plasma can minimize damage. The plasma is created separately at low pressure, and transported to the process chamber, where it is uniformly distributed across a heated wafer surface. No ions to create directionality, which is why downstream reactors exhibit chemical etching.
- **Triode planar reactor**: Adds a third electrode to attain control of the ion bombardment. Typically used in single-crystal silicon trench etching.
- **Ion beam milling**: Also called ion beam etching (IBE). Physical etch mechanism with strong directionality. Plasma is generated by an RF source, and electrons are emitted by a hot filament. An electromagnet surrounds the plasma chamber, causing electrons to travel in circular paths, creating a high number of positive argon ions (argon is used in the process). The ions are drawn from the plasma, sputtering the surface. A neutralizing filament by the wafer prevents charging of the wafer. Operated at low pressure, and used for difficult materials as gold, platinum and copper. The major problems are low selectivity and low etch rate.
- **Reactive ion etch (RIE)**: RIE is similar to the parallel plate reactor, but the wafer is placed at the RF powered electrode. Both a physical and chemical etch process happens. The DC self-bias on the wafer creates extra directionality, and improved anisotropic sidewall profiles. No sputtering at the anode.
- **High-density plasma (HDP) reactors**: Most predominant method for advanced ICs. The pressure is lowered to 1-10 mtorr, to increase the mean free path of gas molecules and

ions. This gives greater profile control, but lowers etch rate. To compensate, a high density plasma is needed. HDP refers to the number of active species relative to the nonactive ones. Ordinary plasma has about 0.1% reactive species, while HDP can have as much as 10 %. A magnetic field is applied because it is more efficient at creating HDP, with high-directional, low-energy ions entering high-aspect ratio gaps and less wafer damage. The DC bias on the wafer is reduced, leading to less ion bombardment. The *electron cyclotron resonance (ECR)* reactor was one of the earliest HDP reactors. it produces very high density plasma. Uses microwave excitation and a magnetic parallel to the microwave field. Electrons spiral, causing more reactive species. The most efficient transfer of energy happens when the cyclotron frequency of the electrons is equal to the frequency of the microwave field. The ions travel towards the wafer. Another high-density, low-pressure reactor is the *inductively-coupled plasma (ICP)* reactor. A spiral coil separated from the plasma by a dielectric plate creates the plasma. The wafer is located away from the coil, so it is not affected by the electromagnetic field. The wafer can be RF biased.

Etch system review: A wide variety of dry etch systems exhibit different capabilities. Stringent control is needed in order to be successful.

Endpoint detection: Dry etch differs from wet etch in that the selectivity is normally not good. A form of endpoint detection is therefore needed. Systems for this measure changes in etch rates, the types of etch products removed or change in active reactants. *Optical emission spectroscopy (OES)* measures the emitted light from excited atoms and molecules in the gas discharge. In this method, the endpoint detector can measure when the etch reaction has etched the desired material, and when the etching of the underlying layer has started. OES can also be used to perform etch reactor diagnostics, i.e. if a leak is present, nitrogen can be found in the chamber.

The vacuum system must handle the etch byproducts well, and produce low pressures of 1mtorr or less. Typically, turbo pump is used, as well as toughing pumps.

2.11.5 Dry etch applications

The requirements for successful dry etch are:

1. High selectivity to avoid etching material that not to be etched.
2. Fast etch rate, to achieve high wafer throughput.
3. Good sidewall profile control.
4. Good etch uniformity across the wafer.
5. Low device damage.
6. Wide process latitude for manufacturing.

Dielectric dry etch: The most complicated etch process due too critical dimensions being etched. Common for etching contact holes and vias. Required selectivity can be be 50:1. Fluorocarbons are used as sources for fluorine. The fluorocarbon gases dissociate to create fluorine radicals that attack the oxide. Higher carbon-to-fluorine ratio often means higher selectivity and lower etch rate. One of the challenges is to get the high selectivity in comparison to the underlying materials. Adding oxygen or hydrogen can help. Adding a etch stop layer is effective, but increases the number of process steps. Adding a polymer to the gas can passivate

the silicon. Selectivity to photoresist layers is a challenge when free fluorine is produced in the plasma, because it attacks the organic photoresist. DUV resists are generally less resistant than i-line during plasma processing. The sidewall profiles must be anisotropic. If the photoresist sensitivity is too low, erosion of the photoresist can lead to tapered sidewalls. During nitride etch, different fluorocarbons can be used to etch it.

Silicon dry etch: The two major layers for plasma etching are polysilicon gate formation and single-crystal silicon trench creation. Originally, fluorocarbons were used to etch, but now chlorine or bromine chemistry. Both have high selectivity, though bromine is the best. By mixing HBr and chlorine, with oxygen, increased etch rate, selectivity and a polymer formation is achieved. During poly gate etch, CD control is critical. The etch must have high selectivity and must be anisotropic. Etching poly silicon or silicon is a three step process. The steps are: *breakthrough*, which involves removing native oxide. This is followed by the *main-etch step* where most of the poly silicon is removed without damaging the gate oxide. The last step is the *overetch step*, that involves removing the remaining residues. Very high selectivity is needed. Fluorine gases, as well as chlorine and bromine is used to achieve the necessary results. Etching of single-crystal silicon requires precise dimensional control as well. A multistep etching is done, as before mentioned. During the silicon trench etching, fluorine, bromine and chlorine gases are used. Bromine gases are extremely corrosive to the gas delivery system. This problem is being addressed.

Metal dry etch: The major requirements are high etch rates (¿1000nm/min), high selectivity to masking layers, ILD and underlying layers, high uniformity with excellent CD control and no micro loading, no device damage, low residue contamination, fast resist strip and no corrosion. *Aluminum and metal stacks* are etched with chlorine and polymers for directionality. A challenge is the complexity of the multilayer stack, with barrier layers, ARC layers etc. Also, aluminum oxidizes very quickly when exposed to air. The typical steps for metal etching involve a breakthrough step, an ARC layer etch, main etch, overetch step, a barrier layer etch, optional residue removal and resist removal. Corrosion control is important to device performance after etching. The corrosive byproducts must be removed. By controlling water vapor and oxygen, most of the work is done, but in addition, the resist strip can be used to remove corrosive compounds. *Tungsten* is a common metal for vias. Tungsten etch can be done with fluorine or chlorine compounds. Tungsten etchback is one step in the formation of tungsten plugs. The etch involves two steps: First, 90% of the tungsten is etched away with excellent uniformity at a high rate. Then, a gas chemistry with high selectivity to the TiN barrier is used. Anisotropic etching is not needed, but minimal residue and plug loss is important. This process has been replaced by CMP in many fabs. Removing the refractory metals used to create silicides can be done with fluorinated or chlorinated gas chemistries.

2.11.6 Wet etch

The benefits of wet etch to dry etch is good selectivity. no risk of plasma damage and the use of a simple process. The following parameters must be controlled: Concentration of etchant, the time of etching, the temperature, the agitation of the solution bath and the number of runs. Wet etch is unsuitable for geometries below 3μm.

Types of wet etch: The *wet oxide etch* is done by a HF bath. The bath is a dilute solution of HF with ammonium fluoride, known as a buffered oxide etch, or buffered HF. Silicon oxide etches readily in buffered HF, but the etch is isotropic, which is not beneficial. *Wet chemical strips* are sometimes used to remove surface layers, like photoresist. Removal of nitride layers

is often done with wet etch with hot phosphoric acid.

2.11.7 Photoresist removal

Also called photoresist stripping. Done by wet etch, or dry etch. *Plasma ashing* is the dry removal of resist with oxygen, by reacting oxygen atoms with the resist material in a plasma environment. Plasma damage can occur during ashing. This has been remedied by using a "downstream" technique. Newer ashing methods include residue stripping. Wet cleaning can still be employed to remove residues, because the elevated temperatures in the ashing method may harden the residues.

2.11.8 Etch inspection

Quality measures must be ensured. Manual microscopes have been used, but today, automated inspection systems are mostly used. Similar to develop inspection.

2.12 Chapter 17 - Ion Implant

2.12.1 Objectives

1. Explain the purpose and applications for doping in wafer fabrication.
2. Discuss the principles and process of dopant diffusion.
3. Provide an overview of ion implantation, including its advantages and disadvantages.
4. Discuss the importance of dose and range in ion implant.
5. List and describe the five major subsystems for an ion implanter.
6. Explain annealing and channeling in ion implantation.
7. Describe different applications of ion implantations.

Doping is the process of adding a dopant (impurity) to a semiconductor material to modify its electronic properties. There are two techniques: thermal diffusion and ion implantation. **Doped regions** is where dopants reside within a semiconductor. A wafer is uniformly doped during crystal growth to create a p-type or n-type semiconductor, and the wafer is selectively doped in specified regions during production to create the devices on the individual wafers. A doped region is characterized by its *dopant profile*. A doped region can be of the same or opposite type as the wafer itself. The region where the doping changes from p to n is the on-junction, and the depth is called the junction depth, x_j. This is where the net dopant concentration is zero.

2.12.2 Diffusion

High temperature diffusion is a method used to introduce dopants in semiconductor fabrication. There are three steps involved: pre deposition, drive-in and activation. During predep, wafers are wafers are loaded in a high-temperature diffusion furnace, and the dopant atoms are transferred to the furnace. The dopants are introduced as a very thin layer on the wafer. A

thin layer of oxide is grown on the surface, called *cap oxide*. The total number of dopant atoms is called Q. This establishes the concentration gradient needed for diffusion, and Fick's laws describe the diffusion. The sec on part, the drive-in, involves a higher temperature to move the dopants into the wafer. Finally, the temperature is raised even higher, to achieve activation.

Each dopant has a particular *diffusivity*, which represents the rate at which the dopant moves in the wafer. The dopants move interstitially or substitutionally. Only when the dopants are activated will they contribute to the electric properties of the semiconductor. Only a fraction of the dopants are activated. During diffusion, the dopants move in all directions. Lateral diffusion happens, but is undesirable. There are eight steps required to properly perform diffusion in wafer fab,

1. Run qualification test to ensure the tool meets production quality criteria.
2. Verify wafer properties.
3. Download process recipe with desired diffusion parameters.
4. Set up furnace and temperature profile.
5. Clean wafer and perform a HF deglaze.
6. Perform predep.
7. Perform drive-in and activation.
8. Measure, evaluate and record junction depth and sheet resistivity.

Dopant sources used are often gaseous or liquid compounds containing the desired dopants. Pure dopants are not used, because they are difficult to deposit.

2.12.3 Ion implantation

A purely physical process, and preferred in most processes because of its repeatability and control. Done by the *ion implanter*, which has an ion source. The ions are extracted and separated in a mass analyzer to form a beam of the desired dopant ions. The ions are focused into a beam, and accelerated onto the wafer. The beam scans the wafer to provide a uniform layer across the wafer surface. A thermal anneal is done to repair the semiconductor crystal and activate the dopants. The two goals are to introduce a uniform amount of a specific dopant and to place the dopants at a desired depth.

Table 4: Advantages of ion implantation.

1.	Precise control of dopant concentration.
2.	Good dopant uniformity.
3.	Good control of dopant penetration depth.
4.	Produces a pure beam of ions.
5.	Low temperature processing.
6.	Ability to implant dopants through films.
7.	No solid solubility limit.

The disadvantages are *radiation damage*, which can be fixed by a thermal anneal, and the complexity of the process. The benefits overshadow the disadvantages.

Ion implant parameters: The two important parameters are:

- *Dose (Q)*: Refers to the number of implanted ions per unit area of wafer surface. Q is calculated as
$$Q = \frac{It}{enA}$$
where I is the beam current in ampere, t is the implant time in seconds, e is the electronic charge, n is the charge per ion and A is the area implanted.

- Range: The ion range is the total distance an ion travels in the silicon during implantation. The range is proportional to the kinetic energy of the ion, which is given by
$$KE = nV$$
where KE is the kinetic energy, n is the charge state of the ion and V is the voltage difference of the field accelerating the ion. The *projected range*, R_p, depends on the kinetic energy, but not all ions stop at the projected range. The distribution of distances traveled is referred to as *straggle*, ΔR_p. The straggle represents the spread of the implanted species around the projected range. As kinetic energy increases, R_p increases, but also ΔR_p. Implanted ions can be stopped by electronic stopping or nuclear stopping. Heavier dopants are stopped by nuclear stopping, and produce many displacements in the crystal.

2.12.4 Ion implanters

An ion implant tool consists of the following five major subsystems.

- **Ion source**: Positive ions are formed from a dopant gas, or by vaporizing a solid. Using solids permits higher maximum beam currents, but the setup time is long and more frequent maintenance is needed. Gases are diluted, which is why the beam current is lower. The ions are created by bombardment with electrons. In general, gases are made into plasma, in order to extract the positive ions.

- **Extraction and ion analyzer**: The ion extracter collects all the positive ions, and sends them through a slit in the ion source. The chamber is positively biased, repelling the ions, while the extraction assembly is grounded. A negatively biased suppression electrode is used to focus the ion beam. The ions extracted have different masses. The mass analyzer separates the desired dopant ion from the main body of ions. The mass analyzer works as a mass spectrometer, allowing only the correct ion to pass through.

- **Acceleration column**: To achieve additional acceleration, the ions are accelerated in the accelerate column after the analyzer magnet. An additional beam focusing is done in the post-accelerator. A quadrupole focuses the ion beam. A high-current high-energy beam accelerator uses alternating series of high-voltage electrodes. The voltage is matched to the arrival of the ions, so that the need for extreme voltages is avoided. Normal acceleration columns use linear accelerator technology. In the high-current low-energy beam case, a decelerator is used after the beam is focused. Space charge neutralization is used to minimize beam blow-up. Secondary electrons neutralize the positive charges. Finally, a neutral beam trap is placed right before the sample, so that neutralized dopant ions won't be implanted.

- **Scanning system**: The ion beam has a small diameter, and must be scanned over the sample. This can be done by moving the beam over a stationary sample, or vice versa.

Electrostatic scanning uses electrodes to deflect the beam, and scan the beam over a stationary wafer. The wafer is tilted to ensure a laminar flow of ions. In mechanical scanning, the wafers are moved, and the beam is stationary. In hybrid scanning, the wafers are loaded on a disk, rotated to scan in the y-axis, while the ion beam scans in the x-direction. Parallel scanning uses the same method as electrostatic scanning, but uses magnets to bend the ion beam right above the surface to avoid beam angles and shadowing. The energy of the ion beam is converted to heat, therefore wafer cooling is needed, either by gas cooling, or elastomeric cooling. The latter means covering the metal plate with a thin layer of elastomer, which is in contact with the wafer, maximizing heat conduction. Control of wafer charge buildup is important. Secondary electron flow was earlier used, but now a plasma electron flood system is used, where the wafer is held in a HDP environment, which is filtered so that only secondary electrons reach the wafer.

- **Process chamber**: The ion implant process happens in a process chamber. There is a high vacuum in the chamber, so load locks are used to loading and unloading wafers. The dose control is measured by measuring the ion beam with a sensor know as a Faraday cup. The sensor is placed in the ion beam path and measures the current. After implant, an anneal is done because of the damage to the crystal structure, and to activate the dopants.

Anneal can be done in a furnace for about 30 mins. This repairs the crystal, but the time and temperature can damage the IC and create undesirable diffusion. *Rapid thermal anneal (RTA)* anneals the wafer by using an extremely fast ramp and short dwell time at the target temperature. This minimizes transient enhanced diffusion of the dopants, and is optimal to achieve acceptable junction depth control in shallow implants.

Channeling happens when dopants pass through interstitial areas of the semiconductor crystal. This means that the implant depth is much higher than for those who collide with lattice atoms. This can be prevented by wafer tilting, a screen oxide layer, preamorphization and using dopants with greater amu's.

If there is particle contamination on the wafer surface, the incident ion beam can cause improper implantation. Therefore contamination must be controlled during this process.

2.12.5 Ion implant trends in process integration

Examples of different implant process requirements for advanced MOS wafer fab are:

- *Deep buried layers*: Implanted in silicon with high energy implanters. A triple well has a buried implanted well beneath the standard regions. This controls latch-up, and shunts current to the ground plane.
- *Retrograde wells*: The MOS transistor is placed in a well of the opposite conductivity type to form the semiconductor junction. The retrograde well has the peak implanted dopant profile buried at a certain depth. It is also called a vertically modulated well.
- *Lightly doped drain*: the LLD implant is used when defining the source and drain regions. The LLD creates complex lateral and vertical doping profile sin the interface region at the channel edge. The reduced doping of the LDD decreases the electrical field between the junction and the channel regions, and prevents the creation of hot carriers.
- *Source/Drain implants*: S/D implants form highly doped regions, that interface with the

lightly doped active channel and the well regions. the ion implant method is excellent because of the high degree of control.

- *Polysilicon gate*: Must be doped to render it conductive. Originally just n-doped for ease of fabrication, but his can pose som electrical performance problems. A solution is to create a dual-polysilicon gate structure, where a separate p-layer is created for the p-channel devices to complement for the n-dopant for the n-channel. To do this, unroped polysilicon is deposited and patterned,and then doped when the source and drain are doped.
- *Trench capacitor*: A replacement for the planar storage capacitor. Trenches are etched, and then doped with ion implant at an angle.
- *Ultrashallow junctions*: When features shrink, so must the junctions. Ultrashallow junctions need high control of the doping.
- *Silicon-On-Insulator (SOI)*: To form single crystal silicon on oxide is difficult. A method known as SIMOX (Separation by IMplanted OXygen) implants oxygen atoms into the wafer, and a a high-temperature anneal causes the atoms to react with the silicon to form silicon oxide buried in the wafer.

2.13 Chapter 20 - Assembly and Packaging

2.13.1 Objectives

1. Describe the general trends and design constraints of assembly and packaging.
2. State and discuss the traditional assembly methods.
3. Describe the different traditional packaging options.

The packaging and assembly process consists of two steps: The *IC final assembly*, which involves separating each good die from the wafer and attaching it to a metal lead frame or substrate, and *IC packaging*, which involves enclosing the die in a protective packaging. The protective packaging is normally plastic, and has four important functions: Protection from environment and handling damage, interconnections for I/O signals, physical support for the chip and heat dissipation.

2.13.2 Packaging levels

The *1st level packaging* is the chip assembly and packaging. The 2nd level assembles the IC component into a system of many components and connectors, in many cases the IC component is assembled onto a printed circuit board (PCB).

2.13.3 Traditional assembly

Consists of four basic steps.

1. **Backgrind**: The first operation done in final assembly. Wafers are thinned before assembly because thinner wafers are easier to dice into chips, and have better thermal

dissipation. The backgrind is done fully automatically. Sometimes, a metal film is applied to the backside of the the wafer, normally gold.

2. **Die separation**: During this step, the dies are cut from the wafer using a diamond-blade dicing saw. The wafer is placed on an adhesive film and sliced using the diamond-blade. Today, most wafers are cut fully. Earlier, wafers were partially cut, and then broken apart.

3. **Die attach**: Each good die is individually picked from the adhesive backing, and physically attached to the substrate or lead-frame. The die is physically attached using either epoxy attach, eutectic attach or glass grit attach. *Epoxy attach* uses an epoxy dispensed in the center of the lead-frame to attach the chip. The die is placed on the epoxy, and a thermal cycle is done to cure the epoxy. The epoxy can be formulated with silver flakes if a thermally conductive epoxy is needed. *Eutectic attach* involves depositing a thing film of gold on the backside of the wafer, which is alloyed to the substrate. The alloys used are of eutectic composition to ensure that the chip adheres to the substrate at the lowest possible temperature. *Glass frit attach* consist of a mixture of silver and glass particles in a suspension of organic medium. The silver and glass particles soften and bond to the ceramic substrate during annealing. Only used for ceramic substrates.

4. **Wirebonding**: The most common method for connecting the metal bonding pads to the lead terminals (sometimes called posts). A high-speed process. A fine diameter wire is spooled and bonded from the metal pads to the posts. The three methods of wirebonding are: *Thermocompression bonding*, where thermal energy and pressure are used to form the wirebond to the chip frame and the posts. *Ultrasonic bonding* uses ultrasonic energy and pressure. Pressure and rapid mechanical vibration makes the bonding wire form a metallurgical bond. *Thermosonic ball bonding* is a technique combining ultrasonic vibration, heat and pressure to form the bond. The tip of the wire is melted, forming a ball, and is bonded using ultrasonic energy and pressure. The wafer is connected to the posts by thermocompression.

To ensure quality of the wire bonds, visual inspection and pull tests are done.

2.13.4 Traditional packaging

The two most used types of packaging are plastic packaging and ceramic packaging

Plastic packaging uses an epoxy polymer to completely encapsulate the wire bonded die and lead-frame. The shapes of the package offers great flexibility for the shape of the leads. Plastic packaging also offers low material cost and low weight. The cross-linked polymer is dimensionally stable, ionically clean and resistant to processing temperature up to 250°C. Once encapsulated, only the leads necessary for the 2nd level assembly protrudes. A deflashing step is done to remove excess package enclosure. Component lead forming is done after molding. The different types of plastic packages are:

- Dual in-line package (DIP): has two rows of pin-in-hole (PIH) leads bent downward to attach to the circuit board.
- Single in-line package (SIP): an alternative to the DIP, used to reduce circuit board space occupied by the IC component body.
- Thin small outline package (TSOP): has gull-wing surface mount technology leads along two sides to attach to the corresponding pads.

- Quad flatpack (QFP): Like the TSOP, but with gull-wing leads on all four sides to achieve a high lead density.

Ceramic packaging is typically used in state-of-the-art applications that require either maximum reliability or high-power with a hermetic seal. The packaging has two main methods, either a refractory ceramic processed separately from the chip assembly, or a ceramic DIP (CERDIP) technology. The refractory ceramic substrate is common for IC packaging. Alumina (Al_2O_3) powder is mixed with glass powder and an organic vehicle to form a slurry. The slurry is cast in to thin sheets, that are dried and patterned to build a multilayer substrate. Different layers are sintered together to from a monolithic body. Some challenges for these substrates are high shrinkage, high dielectric constant, increasing parasitic capacitance and the conductivity of alumina. The most common lead format for ceramic packages is brazed pins on a pitch to create the *pin grid array (PGA)*. A low-cost approach is to press two ceramic units together with a lead-frame positioned between them (CERDIP).

2.13.5 Final test

A final electrical test is done to ensure IC quality. The test is the same functional test performed at wafer sort, but the die is now tested as a final IC package. For advanced IC packages, special text fixtures are used, often called *sockets* or *contractors*.