3 Physical and Chemical Vapor Deposition

Thin films are used for electrical, optical, chemical, and mechanical modification of a substrate surface. This is a vast and important subject area and extends well beyond device fabrication applications. Common everyday examples of thin films include titanium nitride coatings used in cutting tools, polytetrafluoroethylene (PTFE) coatings on nonstick cookware, antimicrobial coatings to prevent growth of bacteria, reflective coating on mirrors, and antireflective coatings on eyeglasses. In device fabrication, we are mostly concerned with the deposition of metals and dielectrics to achieve electrical, chemical, and optical modifications. The design of optical thin films is discussed in Chapter 4, but in this chapter, we will discuss the techniques used for making thin films. The techniques can be broadly classified into physical vapor deposition (PVD) and chemical vapor deposition (CVD). Within each of these categories, there are several different techniques and subcategories. We will not attempt to describe all of these in detail, but only focus on the methods most commonly used in typical research cleanrooms.

3.1 PHYSICAL VAPOR DEPOSITION

In PVD, a solid source material is vaporized atom by atom, or molecule by molecule, and deposited on the substrate at a controlled rate. The energy source used for vaporization is the main distinction between the different PVD techniques. In thermal evaporation, thermal energy is used for vaporizing the source. How this thermal energy is delivered can also be very different between the different deposition methods. Heating with a resistive wire is the simplest approach, but it is also possible to deliver a more tightly focused energy using an electron beam. These are known as resistive evaporation and electron beam evaporation, respectively, as shown in Figure 3.1. One could also vaporize the source using energetic particles instead of thermal energy. This is similar to an abrasive blasting process except atomic projectiles are used instead of large particles. This is the sputter deposition process and can be thought of as a “cold” evaporation technique. Another approach is pulsed laser deposition (PLD), where energy is delivered via energetic laser pulses. All of these methods fall under the category of PVD. In CVD, the starting materials are in vapor or liquid form in a different chemical state than the final thin film. These starting materials (known as precursors) are delivered to the substrate surface at the appropriate pressure and temperature so that they may chemically react and form the desired film. In a thermal CVD, elevated temperatures are used to activate the reaction, and in plasma-enhanced CVD, a plasma is used to activate the reaction. Another variant of CVD is atomic layer deposition (ALD), where, instead of delivering the precursors simultaneously, they are delivered in alternating sequence to allow more control over the deposition properties.

3.1.1 THERMAL EVAPORATION

Thermal evaporation is conceptually very simple—the temperature of a material is raised until it evaporates at a sufficiently high rate to produce deposition on adjacent surfaces. The temperature could be just a few hundred degrees in the case of volatile metals like indium or zinc, or several thousand degrees in the case of metals like molybdenum or platinum. The impingement rate of the background gases on the substrate has to be smaller than the impingement rate of the evaporating species to prevent the background gases from reacting and producing oxides and nitrides. Obviously, this means that evaporation has to be done in a high vacuum environment. The mean free path also
Every material is constantly evaporating, whether it is a solid or liquid, hot or cold. The evaporation rate is a strong function of the material’s binding energy and its temperature. The number of atoms leaving the surface expressed in atoms/cm$^2$/s can also be written in the units of pressure if the atomic weight and the temperature are known, as discussed in Chapter 2. This is the vapor pressure of the material and can be written as

$$Z_A = 3.5 \times 10^{22} \frac{P}{\sqrt{M \times T}}$$

(3.1)

where
- $Z_A$ is the rate in atoms/cm$^2$/s
- $P$ is the vapor pressure in Torr
- $T$ is the vapor temperature in Kelvin
- $M$ is the molar mass

The evaporation rate can be converted to a thickness removal rate (or deposition rate in the case of condensation) using the following relationship:

$$r = \frac{Z_A}{\rho \times N_A}$$

(3.2)

where
- $\rho$ is the material density
- $M$ is the molar mass
- $N_A$ is the Avogadro number

**FIGURE 3.1** (a) Resistively heated evaporation and (b) electron-beam-heated evaporation.
Physical and Chemical Vapor Deposition

The vapor pressure curves of several common elements are shown in Figures 3.2 and 3.3. For example, the vapor pressure of gold is $2 \mu T$ at 1000°C increasing to $30 mT$ at 1500°C. The vapor pressure of platinum is much lower, reaching only $1 \mu T$ at 1500°C. Though it is not a fundamental requirement, there is a correlation between melting temperature and vapor pressures. At a given temperature, elements with a low melting temperature such as aluminum have higher vapor pressures than the metals with a high melting temperature such as tungsten. However, this is not universally true. For example, chromium with a melting temperature of 1857°C has a vapor pressure of 100 mT at 1500°C (which is substantially below its melting temperature), but platinum with a melting temperature of 1772°C needs to be raised to 2300°C to reach 100 mT (which is substantially above its melting temperature). Therefore, it is important to consult a vapor pressure chart of different materials when working with thermal evaporation.

Boiling temperature is not a meaningful concept in vacuum evaporation. The definition of boiling point is the temperature at which the vapor pressure becomes equal to the atmospheric pressure, or 760 Torr. In vacuum, everything could be thought of as boiling.

The goal in thermal evaporation is to elevate the temperature of the material to create a high enough vapor pressure that can then condense on the substrate as a thin film. Melting is not a requirement. Adequate vapor pressure is the primary requirement. However, in the vast majority of cases the materials do melt. As shown in Figure 3.1, this makes it necessary to place the substrate above the evaporating species in the vacuum chamber.

### 3.1.1.1 Resistance Heating Method

In resistance-heated evaporation, the desired material in the form of pellets is placed on a tungsten or molybdenum sheet shaped like a boat (or wire basket). Then, a large DC current is flowed through the tungsten boat to raise its temperature (see Figure 3.1). The reason for tungsten or molybdenum is because of their high melting temperature and low vapor pressure so that they would not deform or evaporate. However, due to their low resistance, a very large current has to be supplied to create the required heating power. For a 0.5 Ω filament (when hot), 100 A of current will be necessary to produce 5 kW of heating at a supply voltage of only 50 V.

In resistance heating, the temperature will be highest at the heating filament, then the bulk of the evaporating species, and the lowest temperature will be at the evaporating surface of the material. This is the main drawback of the resistance heating method. The temperature profile is such that everything else is at a higher temperature than the surface that really matters. As a result, the process often runs too “hot,” creating outgassing and evaporation of unwanted species in the chamber. Nevertheless, this method is widely used in the production of thin films because it is simpler than other forms of heating.

### 3.1.1.2 Electron Beam Evaporation

Thermal evaporation could be made more efficient with less collateral heating if we could heat the evaporating surface from the top rather than from the bottom. For example, consider heating the pellets with a lamp, laser, or electron beam. This would cause the evaporating surface, the surface that really matters, to have the highest temperature. The bulk of the evaporating species would have a lower temperature. The crucible could be water-cooled so nearly all but the evaporating surface would remain cool.

In e-beam heating, a hot tungsten filament generates a cloud of free electrons, like in an incandescent light bulb. The filament is held at a strong negative potential, of the order of $-10 kV$. This will result in the electrons being accelerated away from the filament (see Figures 3.1(b) and 3.4). Several beam forming plates near the filament redirect the electrons up and away from the filament, and a set of permanent magnets and electromagnets deflect the electron beam and focus them toward the center of the crucible. The electromagnets can be used to make small adjustments to the beam spot, as well as be programmed to perform sweeps to cover a larger surface area. The background pressure in the chamber needs to be low enough to produce a sufficiently long mean...
FIGURE 3.2 Vapor pressure curves of selected elements. (Phosphorous through Rhenium). (From Honig, R. E., RCA Journal 2, 195–204, 1957. Reproduced with permission from International Union for Vacuum Science, Technique and Applications.)
**FIGURE 3.3** Vapor pressure curves of selected elements. (Rubidium through Tungsten). (From Honig, R. E., *RCA Journal* 2, 195–204, 1957. Reproduced with permission from International Union for Vacuum Science, Technique and Applications.)
free path for these electrons to travel from the filament to the crucible without encountering any collisions. The energy is ultimately delivered to the evaporating surface by bombardment of high-energy electrons.

E-beam heating is more expensive to install and operate than resistive heating. It requires a high-voltage DC source, separate filament and beam current control circuits, and the associated safety issues due to the high voltage. Special precautions also need to be made inside the chamber to prevent charge build ups and arcing. However, for applications where material purity is important, e-beam is often the best choice.

What makes e-beam heating particularly useful is the high-power density at the evaporating surface. Electron beams can be focused down to a very small spot. In principle, the beam can be focused down to a few nanometers in size using elaborate focusing fields, as done in electron microscopes. However, for heating applications 1 mm is small enough and this can be easily accomplished with simple magnets and coils. In addition, the penetration depth of the electron after it strikes the material is less than a micrometer. Therefore, the beam power is delivered to the source over a very small volume. The high-power density makes it possible to achieve very high melt temperatures at relatively low beam powers. If the spot size is 1 mm² and the penetration depth is 1 μm, for a 10 kV 50 mA beam, the power density at the evaporating site will be 500 MW/cm³, for a total delivered power of only 500 W. Power is delivered to the material due to rapid deceleration of the high-energy electrons as they enter the material. Ohmic heating of the current after the initial deceleration is insignificantly small and does not contribute to much heating.

The mean free path of electrons in a solid is a few tens to a few hundred nanometers depending on temperature. This is the distance the electrons travel between collisions. The electrons in a solid are normally at equilibrium with the lattice. At 25°C, the electron temperature is also 25°C, and the average energy is $kT/q = 26$ meV. Even when the bulk temperature rises to 1500°C, the electron energy is still only 153 meV. Therefore, the energy of a 10 keV incident electron is several orders of magnitude larger. Upon entering the material, these high-energy electrons will undergo random collisions with the lattice with a net loss of energy after each collision. After a number of collisions, all of their excess energy will be transferred to the lattice, and the electrons will reach equilibrium with the lattice. After this, the electrons will be transported through the material as ordinary ohmic
Physical and Chemical Vapor Deposition

conduction. The penetration depth of high-energy electrons can be numerically modeled, but not really necessary for this application (some aspects of this is discussed in Chapter 6 under electron beam lithography). For 10 keV, this depth is of the order of a few hundred nanometers.

The crucibles are usually made of copper and are water-cooled. The evaporation species in the form of pellets are placed in the crucible pockets and heated with the electron beam. However, the thermal conductivity between the water-cooled copper pockets and the evaporation pellets will be high, which will require high beam powers to achieve the high temperatures necessary for evaporation. The thermal conductivity can be reduced by placing the pellets in a crucible liner made from a high-temperature material such as graphite, as shown in Figure 3.5. The liner allows high temperatures to be reached at lower beam powers. The liner material has to be electrically conductive to allow the electrons to flow through, but also act as an adequate thermal barrier.

In some cases, graphite can interact with the evaporating species and form compounds. Gold for example is prone to creating hotspots and spitting in a graphite crucible. A tungsten or molybdenum crucible liner can alleviate this problem, although its thermal conductivity will be greater than graphite and will require higher beam powers. One could also use a combination—a tungsten crucible with an outer graphite sleeve or spacer.

3.1.1.3 Thermal Evaporation Rate from the Source

Consider aluminum as the evaporation source. At 500°C, its vapor pressure will be about 1 pTorr, which is a very small value. Using the atomic weight of aluminum of 27 g/mol, this vapor pressure can be converted to atomic removal rate using Equation 3.1 resulting in

\[
Z_A = 3.5 \times 10^{22} \times \frac{10^{-12}}{\sqrt{27 \times (500 + 273)}} = 2.4 \times 10^8 \text{ atoms/cm}^2/\text{s}. \tag{3.3}
\]

The density of aluminum at 500°C is around 2.6 g/cm³. Therefore, the removal rate in thickness/time using Equation 3.2 is

\[
r = \frac{2.4 \times 10^8}{\frac{2.6}{27} \times 6.02 \times 10^{23}} = 4.2 \times 10^{-15} \text{ cm/s}. \tag{3.4}
\]
This is an extremely small number, so it is more meaningful to describe it in terms of monolayers/time rather than thickness/time. The monolayer thickness of aluminum is

$$a = \left( \frac{M}{\rho N_A} \right)^{\frac{1}{3}} = \left( \frac{27}{2.4 \times 6.02 \times 10^{23}} \right)^{\frac{1}{3}} = 2.6 \times 10^{-8} \text{ cm} = 2.6 \text{ Å}. \quad (3.5)$$

Therefore, the evaporation rate of aluminum at 500°C can be calculated as 1 monolayer every 71 days!

Now raise the temperature of aluminum to 1200°C. The vapor pressure will rise to 10 mT. The melting temperature is 660°C; therefore, it will be a liquid at 1200°C. Repeating the earlier calculation, using a density of 2.4 g/cm³, we can find

$$Z_A = 3.5 \times 10^{22} \frac{0.01}{\sqrt{27 \times (1200 + 273)}} = 1.75 \times 10^{18} \text{ atoms/cm}^2/s \quad (3.6)$$

$$r = \frac{2.4 \times 10^{18}}{27 \times 6.02 \times 10^{23}} = 3.5 \times 10^{-5} \text{ cm/s}. \quad (3.7)$$

This is 350 nm/s, which is a significant removal rate. In terms of monolayers, this will be

$$\frac{3.5 \times 10^{-5}}{27} \left( \frac{2.2 \times 6.02 \times 10^{23}}{3.5 \times 10^{18}} \right)^{\frac{1}{3}} = 1290 \text{ monolayers/s}. \quad (3.8)$$

### 3.1.1.4 Deposition Rate and Distribution

The evaporation rate from the source will obviously not be the same as the deposition rate because of the angular distribution of the flux and the distance between the source and substrate. The distribution of the evaporating flux density (atoms/cm²/s) can be fitted to a $\cos^n(\theta)$ relationship. The simplest model is $\cos(\theta)$, known as the Knudsen distribution, and is the kinetic gas analog of the Lambert’s law of optical radiation [1]. This makes the evaporating flux density

$$R(\theta) = R(0) \cos(\theta) \quad (3.9)$$

where $R(0)$ is the flux along $\theta = 0$. If we also assume that the source-to-substrate distance $D$ is much larger than the source diameter such that the source can be treated as a point source, the total flux $R_T$ (atoms/cm²) can be written as

$$R_T = \int_0^{\pi/2} (2\pi D \sin(\theta))(Dd\theta)(R(0)\cos(\theta)) = \pi D^2 R(0). \quad (3.10)$$

From this, we can get

$$R(0) = \frac{R_T}{\pi D^2} \quad (3.11)$$

$$R_T = \pi D^2 R(0). \quad (3.12)$$
This assumes evaporation from a point source, which is not always the case, especially with resistive heating method. Even in e-beam heating, the high vapor pressure near the evaporating source can lead to short mean free paths and increased scattering and make the evaporation source appear as if it has a large effective area [2]. Nevertheless, for most practical applications, the cosine law produces reasonably accurate results.

Consider gold at a temperature of 1750°C (melting temperature is 1064°C). The vapor pressure at 1750°C is 1 Torr. Therefore, using the atomic weight of 197 g/mol,

\[
Z_A = 3.5 \times 10^{22} \frac{1}{\sqrt{197 \times (1750 + 273)}} = 5.5 \times 10^{19} \text{ atoms/cm}^2/\text{s.} \tag{3.13}
\]

Furthermore, if we assume that the evaporating surface has an area of 10 mm²,

\[
R_T = Z_A A = 5.5 \times 10^{18} \text{ atoms/s.} \tag{3.14}
\]

If the source-to-substrate distance is 30 cm, the flux density at the substrate will be

\[
\frac{5.6 \times 10^{18}}{\pi \times 30^2} = 2 \times 10^{15} \text{ atoms/cm}^2/\text{s.} \tag{3.15}
\]

The standard density of gold is 19.3 g/cm³. Therefore, the film growth rate will be

\[
r = \frac{2 \times 10^{15}}{19.3 \times 6.02 \times 10^{23}} = 3.3 \times 10^{-8} \text{ cm/s or } 3.3 \text{ Å/s.} \tag{3.16}
\]

This is the rate at \( \theta = 0 \). At other angles, the rate will drop as \( \cos(\theta) \). If the substrate is a 6 inch wafer, based on the cosine distribution, we can expect a variation of 3% from center to the edge.

An interesting point to note is that, even though the vapor pressure at the evaporating surface is 1 Torr, the vapor pressure at the substrate will be significantly lower:

\[
P = \frac{Z_A \sqrt{M \times T}}{3.5 \times 10^{22}} = \frac{2 \times 10^{15} \sqrt{197 \times (1750 + 273)}}{3.5 \times 10^{22}} = 35 \mu \text{T.} \tag{3.17}
\]

This difference in pressure arises due to the geometry of the vapor flux and the molecular flow of the species. The expanding area of the vapor results in a decrease in the flux density and vapor pressure.

Gold has to be heated well above its melting temperature to reach the 1 Torr vapor pressure and achieve 3.3 Å/s at the substrate for the aforementioned geometries. Similarly, aluminum has to be heated to 1400°C (which is also well above its melting temperature) to reach 1 Torr vapor pressure and a corresponding growth rate of 9.6 Å/s. With chromium, the temperature required to get the same 1 Torr vapor pressure is 1700°C, which is significantly lower than its melting temperature of 1857°C. Chromium typically remains a solid during thermal evaporation. This is known as sublimation. Many dielectrics, including MgF₂ and SiO₂, as well as some metals such as chromium evaporate by sublimation.

3.1.1.5 E-Beam Evaporation of Dielectrics

A material has to conduct electrons in order to be evaporated by an electron beam. Therefore, it may seem odd that dielectrics such as SiO₂, TiO₂, and Al₂O₃ can be evaporated just as easily as metals. This is possible due to internal dielectric breakdown and leakage currents that develop through grain boundaries of the insulator at elevated temperatures. For this to happen, the e-beam power and
temperature has to be raised gradually. A rapid increase can cause the dielectric to charge up and explosively spit the material. Also, the current discharge will not be uniform across the entire surface. There will be some areas where the charge will build up to high levels and deflect the electron beam, and other areas where the current will reach high densities and become too hot. All of these can result in instabilities in the deposition rate. One technique to avoid these problems is to sweep the electron beam over a large surface area. This reduces the current density and allows each site to discharge and recover after being bombarded by electrons.

3.1.1.6 Reactive Thermal Evaporation

The background pressure during normal thermal evaporation is in the range of \( \mu \text{T} \), made up mainly of outgassing species from heated surfaces in the chamber. In reactive evaporation, a background gas such as oxygen or nitrogen is intentionally bled into the chamber to cause a chemical reaction and produce a compound film on the substrate. With a small gas flow, this pressure can be increased to the mT range. The level of pressure needed for reactive evaporation depends on the deposition rate, desired stoichiometry, and reaction rate. Examples of reactive evaporation include VO\(_2\) films from metallic vanadium, SiO\(_2\) from silicon, and MgO from magnesium. Reactive evaporation can also be performed with a plasma that increases the reactivity due to the free radicals generated in the plasma [3].

3.1.1.7 Thermal Evaporation of Alloys and Compounds

Whether or not a compound can be evaporated depends on its thermal stability at elevated temperatures. This will be dictated by the bond strength of the compound. Stable compounds like SiO\(_2\) and Al\(_2\)O\(_3\) can be easily evaporated. Since these bonds do not dissociate during evaporation, the resulting film will be at nearly the same stoichiometry as the evaporating source. However, a small loss of the volatile component is usually inevitable, such as the loss of oxygen when evaporating SiO\(_2\). TiO\(_2\) is particularly known for this because it exhibits many stable oxidation states [4]. This can be prevented, or controlled to some extent, by bleeding a small flow of background oxygen when evaporating these oxides, just like a reactive evaporation. However, many compounds will completely dissociate at high temperatures and evaporate as separate species. For example, if we attempt to evaporate GaP, which is a semiconductor, it will dissociate into gallium and phosphorous. Since the vapor pressure of phosphorous is significantly greater than gallium, the source will release phosphorous and retain gallium. As temperature is increased further, gallium will evaporate and deposit on the substrate. Therefore, only the most thermally stable compounds can be deposited by evaporation.

3.1.1.8 Ion-Assisted Deposition

Since evaporated atoms condense on the substrate gently and at very low energies, they tend to have higher porosity and poor adhesion and experience faster atmospheric degradation compared to more energetic depositions like sputtering. The porous nature also leads to lower refractive indices and the absorption of environmental moisture. Moisture can cause shifts in refractive index as a function of temperature. The porous nature can also induce a significant amount of intrinsic stress in the film. Since optical coating designs tend to utilize a large number of stacked thin films, film stress can become a major consideration.

One way to mitigate these problems is to increase the substrate temperature during deposition. However, many substrates cannot withstand high temperatures, and high temperatures can also result in outgassing, making the deposition process more difficult. Ion assistance in conjunction with evaporation can significantly improve the film characteristics without substrate heating, and in many cases, it can be used in combination with substrate heating. Ion bombardment can densify the films and reduce porosity, resulting in films with increased mechanical and environmental durability, lower scatter, lower optical absorption, and lower stress. As a result, ion-assisted deposition (IAD) is widely used in the production of optical films [5].

In an IAD setup, a separate ion source is installed inside the vacuum chamber. This ionizes a gas and directs a stream of energetic ions at the substrate, as illustrated in Figure 3.6. Ion sources
come in many different configurations, but they all contain a cathode, anode, and a neutralizer. One common variant, known as the end-Hall ion source, uses a hot tungsten filament that acts as the cathode (Figure 3.7). The electrons emitted from the cathode accelerate toward the anode in a spiraling path due to the permanent magnets installed between the anode and the cathode, and this results in an increased interaction length between the gas atoms and the electrons. These energetic
electrons collide with and ionize the neutral gas molecules. The ions are then accelerated away from the anode and an additional number of electrons are released from the filament to neutralize the ions being directed at the substrate. Although argon is the most commonly used feed gas for the ion source, oxygen, nitrogen, or many such gases can also be used, as reactive IAD processes [6].

A photo of an electron beam evaporator used in research with an IAD source is shown in Figure 3.8.

3.1.2 Sputter Removal and Deposition

Sputter removal can be considered a sublimation process because the solid is turned directly to a vapor without melting. It can also be thought of as a “cold” evaporation process because energy is delivered via external projectiles instead of through internal phonons (heat).

In a DC or RF plasma, the cathode plate is bombarded by ions to sustain the plasma through secondary electron emission. As the cathode voltage is increased, in addition to releasing secondary electrons, the ions will also eject neutral species from the cathode. These ejected species will land everywhere in the vicinity of the cathode including the substrate. This is the sputtering process. Since sputtering is a physical removal of material from the target, it is a PVD method just like evaporation. The material to be sputtered is referred to as the target. The material being deposited on is the substrate. The target surface undergoes sputter removal, and the substrate surface undergoes sputter deposition.

A typical sputtering cathode removed from a vacuum chamber is shown in Figure 3.9. The cathode plate is typically a water-cooled copper plate with an electrical connection on the water side and vacuum on the other side. The target material is thermally and electrically bonded to this
copper plate. The ions in the plasma strike the target instead of the copper cathode plate, ejecting the target atoms. Water cooling is necessary because a large portion of the power (60%–80% or more) supplied to the plasma is dissipated as heat on the cathode due to the ion bombardment [7,8]. Kinetic energies of the sputtered atoms, photon emission, and electron heating constitute only a smaller portion of the supplied power. A photo of a vacuum chamber with two magnetron sputtering cathodes and a rotating substrate stage is shown in Figure 3.10.

3.1.2.1 Sputter Removal Mechanism

Upon collision, the incident ions dissipate their kinetic energy to the target’s surface. If the ion energy is larger than the binding energy of the target atoms, it can dislodge and eject a target atom through a primary or secondary collision. Momentum and energy are always conserved, and kinetic
energy can also be conserved if the collisions are elastic. At higher ion energies, the primary collision can set off a cascading series of collisions that can propagate deeper into the target. This can still cause surface atoms to be sputtered if a reflecting cascade reaches the target’s surface with sufficient energy.

For primary collisions, the amount of energy delivered by the ions will be a function of the masses of the ions and the target atoms. The nuclear stopping power of the target atoms determines how far the ions will penetrate into the target. The angle of incidence determines the momentum transfer from the ion to the target. The energy required for the removal of an atom from the target will be related to the binding energy of the atoms in the target. Most of these effects are well understood and can be computationally modeled and experimentally verified.

Although any gas can be used in the plasma, argon is most commonly used gas for sputtering. The first reason is that it is an inert gas. Sputtering must be done with an inert gas plasma unless we intentionally want the target atoms to chemically react with the gas to form compounds during deposition. Second, sputter yield increases when the ion mass is closely matched to the target mass, which makes argon a better candidate than helium and neon for a large number of target elements. The heavier xenon and krypton ions are more effective for sputtering heavier atoms. Finally, argon is an inexpensive gas since it is the third most abundant gas in the atmosphere after nitrogen and oxygen. It is easy to extract, purify, and compress. Therefore, argon is a good compromise in overall performance for nearly all inert sputtering applications.

### 3.1.2.2 Sputter Yield

The sputter yield $S$ is the number of ejected target atoms for each incident ion. It can be calculated using ion–matter interaction models, and tables of experimental data have been compiled for various sputter gases, target materials, and ion energies. Sputter yield is generally found to vary as a function of ion mass, ion energy, angle of incidence, target atomic mass, and target surface quality. Sputter yield has been found to increase from left to right on the periodic table, which roughly correlates inversely with binding energy (heat of sublimation). The sputter yield is also related to the atomic masses of the ion and the target atom. Maximum energy transfer occurs when the ion mass is close to the atomic mass of the target. Table 3.1 shows tabulated values for sputter yield for a few elemental targets as a function of argon ion energies. A more extensive compilation of sputter yields for elements using different ions can be found in Reference 9.

There is a threshold energy below which no sputtering takes place. Beyond this threshold, yield increases with increasing energy and then reaches a plateau. For larger energies, sputter yield starts to decline because ions penetrate too far below the target’s surface to cause sputtering. This is the ion implantation regime, and is used for doping intentional impurities into semiconductors. The ion energy in this case is used for controlling the landing depth of the dopant species.

<table>
<thead>
<tr>
<th>Target</th>
<th>100 eV</th>
<th>200 eV</th>
<th>300 eV</th>
<th>600 eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>0.63</td>
<td>1.58</td>
<td>2.20</td>
<td>3.40</td>
</tr>
<tr>
<td>Al</td>
<td>0.11</td>
<td>0.35</td>
<td>0.65</td>
<td>1.24</td>
</tr>
<tr>
<td>Au</td>
<td>0.32</td>
<td>1.07</td>
<td>1.65</td>
<td>2.43</td>
</tr>
<tr>
<td>Cu</td>
<td>0.48</td>
<td>1.10</td>
<td>1.59</td>
<td>2.30</td>
</tr>
<tr>
<td>Ti</td>
<td>0.08</td>
<td>0.22</td>
<td>0.33</td>
<td>0.58</td>
</tr>
<tr>
<td>W</td>
<td>0.07</td>
<td>0.29</td>
<td>0.40</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Sputter yields can also be calculated from fairly simple empirical models. Figure 3.11 shows the calculated sputter yields of a few common metals using Ar\(^+\) ions as a function of the ion energy using the empirical model described in Reference 10. We can see that the sputter yield of gold with Ar\(^+\) ions shows a threshold energy of about 50 eV and reaches a peak of 12 at 90 keV. Titanium has a threshold around 100 eV and reaches a peak of 2 around 35 kV. Figure 3.12 shows the same calculations done with He\(^+\) ions, where it can be seen that the yields are about an order of magnitude smaller than with Ar\(^+\). Figure 3.13 shows the sputter yields for the same elements using Xe\(^+\) as the ions, where an increased sputter yield can be seen for all elements, and a higher implantation threshold.

**FIGURE 3.11** Calculated Ar\(^+\) sputter yield curves of a few common metals using the empirical model in Reference 10.

**FIGURE 3.12** Calculated He\(^+\) sputter yield curves of a few common metals using the empirical model in Reference 10.
There are also Monte Carlo approaches to modeling the interaction of the ions with atoms. A widely used freely distributed numerical simulation package is SRIM (Stopping and Range of Ions in Matter) developed by James Ziegler [11,12].

Sputter yield is also a function of the incident angle of the ions [13]. When sputter yield is plotted against angle of incidence, the peak yield occurs at around 60°–70° angle of incidence, as illustrated in Figure 3.14. At higher ion energies, the peak sputter yield moves toward the normal incidence angle [14]. In addition, surface roughness also affects the angular distribution of sputter yields [15,16]. However, in a glow discharge plasma the user does not have direct control over the angle of incidence of the ions. The exception is in ion beam sputtering, where a beam of ions is directed at a target’s surface from a separate ion source to induce sputtering.
3.1.2.3 Magnetron Sputtering

In a conventional DC plasma, electrons are released from the cathode through secondary electron emission and are accelerated toward the anode. These electrons create more electrons and ions through impact ionization, and the ions return to the cathode to release secondary electrons to sustain the plasma. Since the secondary electron emission coefficient $\gamma$ is typically much smaller than 1, many impact ionization events will be needed to sustain the plasma. Therefore, the ions will be spread across many mean free paths of distance from the cathode. The ions will have to undergo as many collisions as electrons, and they will be less energetic when they reach the cathode, having lost some of their energy to nonionizing collisions with the gas. This results in a low sputter rate at the cathode and heating of the plasma gas and the substrate. A magnetron plasma overcomes some of these problems by placing cylindrical permanent magnets behind the cathode to alter the trajectory of the electrons so that they traverse a circuitous route instead of moving parallel to the electric field lines. Ions, on the other hand, will move parallel to the electric field and will not respond to the magnetic fields because their velocities are much smaller than electrons. This can be viewed as electrons having a shorter effective mean free path than ions in a direction normal to the cathode. As a result, more ions will be generated closer to the cathode, resulting in a greater ion density, and they will reach the cathode with fewer or no collisions, while electrons continue to collide and ionize a large number of atoms. The higher energy and higher density of the arriving ions results in a higher sputter rate. The greater ion density also allows the magnetron cathode to sustain a plasma at much lower pressures than a simple cathode plate. Magnetrons can be operated at 1–10 mT, whereas simple plasmas may require 100–500 mT. Today, magnetrons are widely used in sputtering applications, and they come in a large number of different configurations for DC as well as RF excitations [7,8,17].

The placement of magnets on a circular magnetron cathode is shown in Figure 3.15. Two ring magnets are concentrically placed to create a radially symmetric magnetic field on the target surface. Between the two magnets where the magnetic field is parallel to the cathode plate, the plasma density will be significantly greater than elsewhere. This also results in a significantly higher sputter rate in those areas, and the target will show a characteristic racetrack-shaped erosion pattern, as shown in Figure 3.16. This uneven erosion can sometimes be a problem in production environments where maximum target utilization is an important consideration.

Ferromagnetic targets such as iron and nickel do not work well with magnetron cathodes because they will shield the magnetic fields from reaching the plasma. These target materials could still be used with a magnetron if they are sliced into very thin sheets to allow the magnetic fields to penetrate the material.

3.1.2.4 Sputter Removal Rate

The removal rate (atoms/s) is the product of the ion current multiplied by the sputter yield. This can be written as

\[
\text{Sputter Removal Rate} = \text{Ion Current} \times \text{Sputter Yield}
\]
For example, consider a tungsten target in an argon plasma, with a DC power of 200 W and a DC voltage of 400 V. In a magnetron plasma, we can assume that the cathode voltage is nearly equal to the ion energy. The sputter yield for 400 eV Ar\(^+\) on tungsten can be looked up from Figure 3.11 to be 0.35. The ion current is 200/400 = 0.5 A. Therefore, the sputter removal rate is

\[
R_T = \frac{I}{q} S.
\]

This is the total removal rate from the entire target surface, so in order to find the deposition rate on a substrate a certain distance away, we would need to know the geometry of the system.

### 3.1.2.5 Sputter Deposition Rate

For a normally incident ion, the ejected atoms will have an angular distribution that can be fitted to a \(\cos^n(\theta)\) relationship [18]. The value of \(n\) depends on the target material, orientation (if it is a crystal), and ion energy. The simplest model is \(\cos(\theta)\), which can give a rough order of magnitude result.

For magnetron cathodes, where the erosion occurs along a racetrack-shaped groove, the angular distribution can be calculated by integrating the \(\cos(\theta)\) function along the ring. This makes it more difficult to write down analytical expressions, but in most practical situations one simply performs a few calibration runs to determine the thickness distribution rather than rely entirely on analytical predictions.

At large distances from the target, however, the sputter deposition rate can still be approximated as a \(\cos(\theta)\) function. As with evaporation, if the sputter removal rate is written as

\[
R(\theta) = R(0)\cos(\theta)
\]

assuming that the target-to-substrate distance \(D\) is much larger than the target diameter, the total removal rate \(R_T\) can be written as
$R_T = \int_0^{\pi/2} \left( 2\pi D \sin(\theta) \right) \left( Dd \theta \right) \left( R(0) \cos(\theta) \right) = \pi D^2 R(0)$ (3.21)

where $R(0)$ is the deposition rate at $\theta = 0$. From this, we can get

$$R(0) = \frac{R_T}{\pi D^2}.$$ (3.22)

Continuing the previous example with tungsten, the deposition rate at a 30 cm distance from the target will be

$$R(0) = \frac{R_T}{\pi 1.1 \times 10^{18}} = 3.9 \times 10^{14} \text{ atoms/cm}^2/\text{s.}$$ (3.23)

If the density and atomic number of tungsten are known, this rate can be converted to a film growth rate. The standard density of tungsten is 19.25 g/cm$^3$, and the atomic weight is 183.8 g/mol. Therefore, the film growth rate will be

$$r = \frac{Z_A}{\rho N_A} = \frac{3.9 \times 10^{14}}{19.25 \times 6.02 \times 10^{23}} = 0.6 \times 10^{-8} \text{ cm/s or 0.6 Å/s.}$$ (3.24)

This is the rate normal to the target surface at $\theta = 0$. At other angles, the rate will drop as $\cos(\theta)$.

### 3.1.2.6 Dependence of Sputter Deposition Rate on Pressure

If the plasma power is held constant and the gas pressure is changed, both the ion current and voltage will change such that the product remains a constant. An increase in pressure will produce an increase in the ion density in the plasma. As a result, the ion current will increase and the voltage will decline. Even though this reduction in voltage will result in a lower sputter yield, due to the increase in ion current and the nonlinear relationship between sputter yield and voltage, there will be an overall increase in the sputter removal rate.

As pressure continues to increase, the voltage will continue to drop and the decline in sputter yield will start to dominate the sputter removal rate. Increased scattering from the gas will also enlarge the angular distribution of the sputtered atoms. Both of these effects will result in a decline in the sputter deposition rate. Therefore, if power is held constant and pressure varied, the sputter removal rate will initially increase and then decrease as shown in Figure 3.17. The optimum operating point is the pressure at which the deposition rate reaches a maximum.

### 3.1.2.7 Energy of the Sputtered Atoms

The energy of the ejected species will be a function of the incident ion’s energy and its interaction with the target. Due to the numerous interaction pathways in the target, the ejected atoms will contain a continuous energy spectrum. For most sputtered surfaces, the atoms have energies in the range of 1–10 eV. For example, the energy spectrum of copper atoms sputtered from 500 eV and 5 kV Ar$^+$ ions is shown in Figure 3.18. This was calculated using the SRIM software that uses a Monte Carlo model [11]. It is interesting to note that the energy distribution remains nearly the same between 500 eV and 5 kV, with the only difference between the plots is the area under the curve (which corresponds to sputter yield). Nevertheless, the important point is that, sputtered atoms are ejected with energies far greater than the energies encountered in thermal evaporation. It is insightful to calculate the equivalent temperature of a sputtered atom. For the 10 eV atom, this temperature is equivalent to
3.1.2.8 Sputter Up versus Sputter Down

The target can be mounted facing down toward the substrate, or facing up. Gravity does not influence the ions or sputtered atoms because their energies are too high. Sputter down is more flexible because it allows differently shaped substrates to be placed on a flat surface rather than having to mount them upside down with retaining clips. Sputter up is beneficial for avoiding large particulates and contaminants (which are influenced by gravity) from falling down on the substrate. The casings

\[ T = \frac{2E}{3k} = \frac{2 \times 10}{3 \times 8.62 \times 10^{-3}} = 7.7 \times 10^4 \text{ K}. \]  

(3.25)
and other surfaces can become heavily coated and may peel off at some point during the deposition. The target can also once in a while eject large chunks of atoms due to thermal stress or defects on the target surface. However, sputter-up can also cause debris to fall on the target face which can then get sputtered, leading to contamination of the deposition film.

3.1.2.9 Compound Sputtering

When compounds are sputtered, ion bombardment will cause the compounds to be ejected as atoms instead of as whole molecules. The atoms will also have different sputter yields due to their different masses and diameters. If one type of atom has a higher yield than the other, it will be preferentially ejected over the other types. This may seem like a serious problem when sputtering compound materials, but it has been found that a certain steady-state composition and depth profile are quickly reached where the sputter removal of both elements occurs at the original stoichiometric composition [19]. This should not be surprising because the only way to consume a target to its entirety is by depleting its constituent elements in their original proportions.

Consider a target with a binary compound $AB$. If $A$ has a higher sputter yield than $B$, the initial sputtering process will remove more of $A$ than $B$. This will make the target surface become enriched in $B$ and deficient in $A$. As a function of depth, the $B/A$ ratio will be highest near the surface, falling to the original stoichiometric ratio after roughly the range of the incident bombarding ions. The majority of $B$ will get sputtered from the surface where its concentration is greatest and the incident ions are also more energetic. The majority of $A$ will be sputtered from slightly below the surface, where the ions will be less energetic due to secondary collisions. This will result in a stoichiometric removal of the atoms. More importantly, the same depth profile will be maintained as the target erodes to completion. This is depicted in Figure 3.19. As an example, consider SiO$_2$. This target often develops a brown tint due to the enrichment of silicon on the surface. However, the net erosion of the target will occur as stoichiometrically correct SiO$_2$.

**FIGURE 3.19** Target depth profile of a binary compound where the sputter yield of $A$ is greater than $B$. 
Nevertheless, stoichiometric removal does not always imply stoichiometric deposition. In the case of SiO$_2$, not all oxygen atoms will recombine with the silicon atoms on the substrate to reform SiO$_2$. Some oxygen will be lost to the pumping system due to its higher volatility compared to silicon. A common technique used to compensate for this effect is to flow a small amount of oxygen with the argon feed gas to the plasma. This allows the depositing film to become replenished with new oxygen atoms to make up for the loss of oxygen.

Compared to metals, the bond strength and energy of vaporization of dielectrics are high, which makes their overall sputter rates very low. Models for the calculation of the sputter yield of ionic compounds such as oxide and nitrides are significantly less mature than that of elements.

### 3.1.2.10 Co-Sputtering

Co-sputtering is the process of sputtering different target materials from different cathodes such that they deposit on the substrate at the same time. It forms a simple mixture of the sputtered species. Although the same mixtures can be sputtered from a single alloy target, co-sputtering allows one to change the composition of the film without having to purchase a new target every time a different composition is required. Examples include SiGe films made by co-sputtering silicon and germanium and conductive optical films such as Zn-In-Sn-O produced by co-sputtering ZnO and ITO (indium-tin-oxide). Reactive co-sputtering can also be done by introducing a reactive gas. Silicon and titanium can be co-sputtered with oxygen gas to produce a mixture of SiO$_2$ and TiO$_2$. SiO$_2$ has a refractive index of 1.48 and TiO$_2$ has a refractive index of 2.5. By co-sputtering, one could realize a range of refractive index value between these extremes, as well as create graded refractive index profiles for optical thin-film applications.

### 3.1.2.11 Reactive Sputtering

An elemental target can be sputtered with a mixture of argon and another reactive gas (such as oxygen or nitrogen) to create compounds. Sputtering a metallic target to produce a dielectric thin film has a number of attractive qualities—dielectrics generally have very low sputter yields but metals have a high sputter yield; dielectrics have poor thermal conductivity and the target can easily crack due to thermal stress during sputtering; sputtering of metallic targets can be done with a DC plasma, whereas dielectrics require an RF plasma; the production of metallic targets is much easier and can be obtained in greater purity and at lower cost compared to dielectric targets [17]. The reactive sputtering of aluminum to form Al$_2$O$_3$ is a good example because the sputtering rate of Al$_2$O$_3$ is anemically low, but the sputter rate of metallic aluminum is very fast.

However, a major difficulty with reactive sputtering of metals is the narrow process window for achieving optimum conditions. This is due to the competition between the reaction rate that produces a dielectric film on the target’s surface, and the sputter removal rate of that dielectric. This balance is greatly influenced by the vastly different cathode voltages and sputter rates when the target is fully metallic vs. when it is fully dielectric.

Considering the reactive sputtering of aluminum, referring to Figure 3.20, as the oxygen flow is increased from zero, initially the metal atoms will be sputtered, and some of these metal atoms will become oxidized at the substrate. The cathode voltage will remain high and any oxides forming on the target will be quickly sputtered off, leaving the target surface mostly metallic. In other words, the sputter removal rate of the oxide will be equal to the oxidation rate of the target. Despite the increase in oxygen flow, the oxygen partial pressure in the chamber will not show a noticeable increase because all the oxygen will be consumed (gettered) by the sputtered metal atoms.

As oxygen flow is increased further, a point will be reached where all the sputtered metal will be fully oxidized. This is the optimum point of operation for reactive sputtering. However, if the oxygen is increased further, even briefly, the target surface will build up a continuous layer of oxide. The cathode voltage will also drop rapidly (and the current will increase) due to the higher secondary electron emission coefficient of Al$_2$O$_3$ compared to aluminum, and the sputter rate will also drop precipitously. The sputtered species will be Al$_2$O$_3$ rather than aluminum. At this point,
the oxygen flowing into the chamber will show a rapid rise in the oxygen partial pressure having depleted all the sources of pure metal, and the target is said to have been “poisoned.” In the case of DC plasma, this can produce arcing and instabilities due to the dielectric breakdown of the thin layer of Al$_2$O$_3$ on the target face. In the case of an RF plasma, a steady plasma will be sustained but with a very low sputter rate and a low DC bias.

Reducing the oxygen flow will reverse this condition, except the process will not follow the same trajectory back to the original condition. This is a well-known hysteresis effect in reactive sputtering [20,21], and it makes automation difficult for maintaining operation at the optimum point. Due to the low sputter removal rate of the oxide, the oxidation rate of the metallic target exceeds the removal rate of the oxide, and the oxygen level has to be reduced significantly below the previous transition point where it became “poisoned.” A typical hysteresis curve is shown in Figure 3.20.

3.1.2.12 Thermal Evaporation versus Sputtering

Typical background pressures encountered during magnetron sputtering is about 1–10 mT. The mean free path at this pressure is about 1 cm. During thermal evaporation, the pressure is much lower, of the order of 1 μT, except in reactive evaporation or IAD. At 1 μT, the mean free path is 100 m. As a result, thermal evaporation can be considered to be almost a line-of-sight deposition. The evaporating species arrive at the substrate without encountering any collisions. In sputtering, depending on the distance between the target and substrate, there could be some collisions. The line-of-sight deposition in thermal evaporation has important benefits when films need to be patterned by lift-off lithography.

In evaporation, the arriving species will have energies close to the thermal energies of the source. The highest temperature encountered during thermal evaporation is of the order of 2500°C. The average thermal energy at 2500°C is 234 meV. In sputtering, the energies of the arriving species are of the order of 10 eV, which is several orders of magnitude higher. Therefore, thermal evaporation is a low-energy deposition compared to sputtering. The atoms land on the substrate much more gently in thermal evaporation, whereas in sputtering they collide with the substrate at high energies causing damage and heating. On the other hand, the high deposition energy in sputtering can be advantageous for compacting and densifying the film. In thermal evaporation, the film will have a lower density than in sputtering, and voids and gaps in the film can form more easily. The porous nature of the evaporated films can be also exploited to make unique thin films, such as nanostructured thin film for applications in sensors and for ultralow refractive index applications [22,23]. These are some of the inherent differences between sputtering and thermal evaporation.

Another important difference between sputtering and thermal evaporation is their ability to deposit compounds. Neither method can maintain a perfect stoichiometry, but sputter deposition
has many advantages over thermal evaporation, especially for alloys and mixtures. The difference in sputter yield between two elements is typically smaller than the difference in their vapor pressures. Furthermore, as discussed earlier, the target’s surface composition becomes conditioned to sputter the constituent elements stoichiometrically, although some loss of more volatile species is always inevitable. On the other hand, stable molecules like SiO$_2$ may not dissociate under thermal evaporation, so these can be evaporated or sputtered equally well.

### 3.1.3 Pulsed Laser Deposition

Pulsed laser deposition is a PVD method where a series of high-energy pulses from a ultraviolet (UV) laser beam are focused onto a target to ablate the atoms off the surface. The ablated atoms are then collected on a substrate surface located a certain distance away. No background gas is necessary for this process, although a gas can be used to perform a reactive deposition or to compensate for loss of species from a compound target. In that sense, PLD can be thought of as a sputtering process with a greater flexibility to use with any background gas, or without any gas at all. The laser and the focusing optics are all located outside the vacuum chamber, so it also has the advantage of needing minimal installation inside the chamber [24]. A typical layout of a PLD system is shown in Figure 3.21.

By far the greatest advantage of PLD is the stoichiometric removal of the target atoms, even more so than sputtering. This arises due to the short duration of the ablation process and the high laser fluence. The target surface temperature reaches an extremely high value, on the order of 5000 K, for a few nanoseconds. Because this is substantially higher than the ablation threshold of all elements, and because it lasts for a much shorter duration than the time necessary to establish a steady-state, all constituent elements are removed equally regardless of their volatility. As a result, PLD is used mostly for producing thin films of complex compounds that are too unstable to evaporate, or have widely different sputter yields and volatility to use with sputter deposition methods. Examples of films produced with PLD include yttrium barium copper oxide (YBCO) in superconductors and copper zinc tin sulfur (CZTS) in solar cells. The biggest drawback of PLD is the poor film uniformity and slow deposition rates. Since ablation occurs from a single small point on the target and the ablated plume has a narrow angular distribution, the deposition typically covers only a small area on the substrate. Moving the substrate to a greater distance will improve uniformity, but the rate will also drop very fast, and this distance is ultimately limited by the size of the vacuum chamber.

UV lasers are the most commonly used excitation sources in PLD systems, such as KrF (248 nm), ArF (193 nm), and F$_2$ (157 nm). The UV wavelength allows most of the energy to be absorbed within a short penetration depth of the target, which is necessary for the atomic removal of the target. Longer wavelengths will deposit the laser energy deeper inside the target and could result in clusters

![Figure 3.21 Layout of a pulsed laser deposition system.](image-url)
of atoms being ejected. Even with UV lasers, ejection of micron-sized particles and molten liquid droplets can still occur, and it continues to be a problem with PLD.

As mentioned before, PLD allows a wide range of gases and pressures to be used. It is possible to use the gas to decelerate the ablated species, or filter out certain energies. Since pressure controls the scattering rate of the ablated atoms with the background gas atoms, the speed of the depositing atoms on the substrate can be controlled by adjusting the gas pressure. Similar techniques have also been used to create nanoparticles [25]. PLD is also more suitable for reactive deposition of certain metals that require a high level of background pressure. One example is the synthesis of VO$_2$, which when produced from pure metallic vanadium requires a narrow process window with a high level of oxygen pressure and high substrate temperatures [26]. The pressure can be too high for magnetron sputtering or e-beam evaporation, but PLD is more flexible to allow a large range of pressures to be used during deposition.

### 3.2 CHEMICAL VAPOR DEPOSITION

One of the characteristics of PVD is that the source material is in the same chemical state as the final thin film. Reactive sputtering and reactive evaporation can be considered minor exceptions to this definition, where a gas is used to modify the chemical state of the film. In CVD, all of the starting materials are volatile and are in a different chemical state than the intended film material. The film is produced from gaseous precursors as a result of a chemical reaction that occur at the surface of the substrate. Besides device fabrication, CVD is widely used in many industrial processes, such as titanium nitride and tungsten carbide coatings on cutting tools, silicon carbide diffusion barrier coatings in turbine engines, and other thermal barrier and wear-resistant coatings as well as in the production of ceramics [27,28].

The deposition rate in CVD will be directly related to the gas pressure. Higher pressures will provide more reactive species and result in a greater reaction rate. This is another major difference between CVD and PVD. The pressures used in CVD are generally in the range of hundreds of mT to several Torrs in the viscous flow regime, whereas in PVD the pressures are rarely higher than a few mT.

CVD can be used for making dielectric or metal films. Crystalline films can also be made under specific conditions on lattice-matched substrates, and is known as epitaxy. A characteristic of CVD is that the films produced are generally conformal (good step coverage)—that is, they exhibit nearly equal deposition rates on horizontal and vertical surfaces. This is a desirable property in applications such as passivation, planarization, and insulation, as well as in filling holes (known as vias) with metals in integrated circuit processes. The conformal property arises due to the isotropic diffusion of the reactive gases, whereas in PVD the deposition species follow a line-of-sight path. The stoichiometry of a compound can also be easily controlled in CVD over a very large range. This allows one to tune the material properties such as density, refractive index, and stress to meet different application requirements. The major drawbacks of CVD are the requirement for high substrate temperatures and the use of potentially hazardous gases. With PVD, substrate heating or the use of gases are not fundamental requirements, except when necessary to condition the film during the deposition.

SiO$_2$, Si$_3$N$_4$, SiC, amorphous silicon, refractory metals, and diamond are examples of films that can be made with CVD. The gas precursor for silicon is most often SiH$_4$ (silane) or SiH$_2$Cl$_2$ (dichlorosilane). Examples of metal precursors are WF$_6$ (tungsten hexafluoride), (CH$_3$)$_3$Al (trimethyl aluminum), CuCl$_2$ (copper chloride), etc. A few commonly used CVD reactions are as follows:

- SiH$_4$ + 2N$_2$O → SiO$_2$ + 2N$_2$ + 2H$_2$
- SiH$_4$ + O$_2$ → SiO$_2$ + 2H$_2$
- 3SiH$_4$ + 4NH$_3$ → Si$_3$N$_4$ + 12H$_2$
- 3SiH$_2$Cl$_2$ + 4NH$_3$ → Si$_3$N$_4$ + 9H$_2$ + 3Cl$_2$
- WF$_6$ + 3H$_2$ → W + 6HF
The precursors used in CVD can be generally categorized into halides, hydrides, and metal organics. In the aforementioned list, WF$_6$ and SiH$_2$Cl$_2$ are halides, SiH$_4$ and CH$_4$ are hydrides, and Al(CH$_3$)$_3$ is a metal organic.

One of the goals in CVD is to allow the primary reaction that produces the film to occur on the substrate and not in the gas phase. Gas phase reactions are undesirable because the by-products can precipitate on the substrate and produce a film similar to PVD. Depending on the size of the precipitates, this can result in large grains, pin holes, and film roughness. Gas phase reactions can be reduced by choosing the pressure and temperature such that the reactions only occur on the heated substrates.

For the reaction to be thermodynamically favorable, the parameter known as the Gibb's free energy has to be negative. For example, the reaction SiH$_4$ + 2N$_2$O → SiO$_2$ + 2N$_2$ + 2H$_2$ has a Gibb's free energy $\Delta G = -1121$ kJ/mol at room temperature, becoming $\Delta G = -1184$ kJ/mol at 1000°C. These values were obtained from the FactSage thermodynamic software [29]. Similarly, SiH$_4$ + O$_2$ → SiO$_2$ + 2H$_2$ has $\Delta G = -913$ kJ/mol at room temperature and has $\Delta G = -832$ kJ/mol at 1000°C. However, the reaction kinetics is much more difficult to predict, and would depend on the exact reaction pathway. The activation energy $E_a$, which dictates the reaction rate, is often measured rather than calculated. Most of these reactions require an elevated temperature, in the range of 500°C–1000°C in order to proceed at a reasonable rate. For the nonchemist, it may seem odd that silane (SiH4) and oxygen require high temperatures to react, given that silane is a highly pyrophoric gas that reacts with oxygen in the air spontaneously and violently. This has to do with the pressure. At atmospheric pressures, even if the probability of reaction is very low at room temperature, there are a large number of oxygen molecules that some of them will react. The energy released during the reaction will be sufficient to cause a local increase in temperature to accelerate the next reaction, eventually leading to a rapid increase in reaction rate.

There are two main types of CVD reactors: cold wall reactors and hot wall reactors. In a cold wall reactor, only the substrates inside the reactor are heated, either inductively or radiatively, while the walls are kept cool. This prevents deposition from taking place on the walls because such depositions could eventually accumulate and flake off leading to debris and contaminations on the substrates. However, the cold walls can also lead to undesirable temperature gradients in the chamber. As a result, cold walls are most suitable for applications requiring thick-film depositions. Hot wall reactors are more common in semiconductor applications. Here, the walls and the substrate are heated to the same temperature. Depositions will take place on all surfaces, including the chamber walls, but uniformity will be significantly better than in cold wall reactors.

The design of a CVD system is driven mostly by the rate-determining step for the reaction. In the simplest model, there are two rate-determining steps: (1) diffusion rate of the reactants from the free flowing gas to the substrate surface through the stagnant boundary layer and (2) surface reaction rate. The smaller of the two will limit the overall reaction rate. This can be thought of as two serial processes each with a certain conductance (or resistance). The smallest conductor (largest resistor) will limit the overall flow through the system, as shown in Figure 3.22.

For uniform deposition, it is desirable to have a laminar gas flow instead of a turbulent flow. Turbulent flows can cause differences in gas concentration inside the chamber and lead to non-uniformities in the deposition. The flow characteristics of a viscous fluid are determined by its Reynolds number, which is a dimensionless parameter specified as

$$Re = \frac{\rho ud}{\eta}$$

(3.26)
where
\( \rho \) is the density (which is directly related to pressure)
\( \nu \) is the flow velocity
\( d \) is the diameter of the reaction tube (or chamber)
\( \eta \) is the viscosity

Reynolds number is a representation of the ratio between the viscous force and the inertial force in a flow. When the flow is dominated by viscous forces, laminar flow results. When it is dominated by inertial forces, turbulent flow results. Generally, \( Re < 1000 \) is desired to ensure laminar flow. The gas velocity \( \nu \) will have to be maintained at a value necessary to supply the feed gas for the reaction, and the dimensions of the reactor will be mostly dictated by the substrate sizes. As a result, \( \rho \) (or equivalently pressure) is the only quantity that can be changed to ensure laminar flow. Most CVD reactors operate at Reynolds numbers of around 100.

With laminar flow, there will be a thin stagnant boundary layer on all surfaces, and transport through this layer occurs by diffusion. A difference in partial pressure will develop between the free flowing gas stream and the stationary substrate surface. This flow rate can be written as

\[
J_{12} = h_g (P_1 - P_2)
\]

(3.27)

where
\( P_1 \) is the partial pressure of the reactant species in the free flowing gas
\( P_2 \) is the partial pressure near the substrate surface

The adsorption process is

\[
J_{23} = k_a (P_2 - P_3)
\]

(3.28)

with the pre-exponential factor following the well-known Arrhenius equation

\[
k_a = A_a e^{(-E_a / RT)}
\]

(3.29)

where \( E_a \) is the activation energy for the adsorption process. Simultaneously, there will be a desorption process:

\[
J_{32} = k_d (P_3 - P_2)
\]

(3.30)

with

\[
k_d = A_d e^{(-E_d / RT)}
\]

(3.31)

where \( E_d \) is the activation energy for the desorption process.
Finally, the surface reaction rate is

\[ J_{30} = k_s P_3 \]  

(3.32)

with

\[ k_s = A_s e^{(-E_{as}/kT)} \]  

(3.33)

where \( E_{as} \) is the activation energy of the reaction.

The combination of all three factors can be visualized using a circuit model where \( 1/h_g \) is the diffusion resistance, \( \frac{1}{k_a - k_d} \) is the resistance due to the net adsorption, and \( 1/k_s \) is the reaction resistance. The final current flow \( J_{30} \) represents the film growth rate (see Figure 3.23).

We can consider two limits:

1. If \( \frac{1}{h_g} \gg \frac{1}{k_a - k_d} + \frac{1}{k_s} \) (diffusion resistance is large), then \( J_{30} \approx h_g P_1 \). In this case, the reaction rate is determined only by the diffusion resistance. This is the diffusion-limited case.

2. If \( \frac{1}{h_g} \ll \frac{1}{k_a - k_d} + \frac{1}{k_s} \), then \( J_{30} \approx \frac{P_1}{k_a - k_d} + \frac{1}{k_s} \). In this case, the adsorption and reaction resistances are greater than the diffusion resistance, and the reaction rate is determined mostly by the surface properties. This is the surface-limited case.

The boundary layer thickness \( \delta \) increases as the Reynolds number gets smaller. This is expressed as [30]

\[ \delta = \frac{5x}{\sqrt{Re}} \]  

(3.34)

\[ \text{FIGURE 3.23} \quad \text{Circuit description of the rate-determining steps of the chemical vapor deposition reaction.} \]
where \( x \) is the distance from the leading edge of the flow container. The diffusion coefficient varies inversely with pressure as

\[
D = \frac{2}{3pd^2} \sqrt{\frac{k^3T^3}{\pi^4m}}
\]

where
- \( p \) is the pressure
- \( d \) is the diameter of the gas molecules
- \( m \) is the molecular mass

The viscosity is independent of pressure and is expressed as

\[
\mu = \frac{2}{3d^2} \sqrt{\frac{kT}{\pi}}.
\]

Combining Equations 3.26 and 3.34 through 3.36, the diffusion parameter \( h_g \) that varies as \( D/\delta \) can be found to be

\[
h_g \propto \frac{D}{\delta} \propto \frac{T^3}{P^2}.
\]

Although this equation is not quantitative, it still provides some guidance on how pressure and temperature play a role in the gas transport through the boundary layer. We can see that the boundary layer diffusion parameter \( h_g \) will get larger at lower pressures and higher temperatures. Therefore, pressure and temperature will determine which of the two rate-limiting factors that we discussed earlier will dominate the process.

Most commonly encountered CVD systems are

- Atmospheric pressure CVD (APCVD)
- Low-pressure CVD (LPCVD)
- Plasma-enhanced CVD (PECVD)

### 3.2.1 Atmospheric Pressure Chemical Vapor Deposition

This is typically used as a purge-and-flow system at near atmospheric pressures. Due to the high pressure, from Equation 3.37, \( h_g \) will be small, and the primary rate-determining step will be the diffusion rate through the boundary layer (case 1 of the two limiting cases discussed earlier). Typically, no vacuum systems are used in APCVD. The chamber is initially purged with argon or nitrogen to remove residual gases. Then highly diluted reactive gases are flowed over the heated substrates at atmospheric pressures. The dilution is necessary to reduce the gas phase reactions that can occur before the gases reach the substrates. Deposition rates are high, generally greater than 1000 Å/min. Since the reaction is limited by the diffusion rate, the substrates have to be placed adequately far apart to prevent bottlenecks in the gas transport. Obstructions can create turbulent flows and this will result in variations in deposition rates. Also, some reactions in the gas phase are unavoidable due to the higher pressures and slower diffusion (which results in longer gas phase residency times). This results in a large number of particles being produced, which have to be continuously removed with an exhaust, but some of them will inevitably become embedded in the film.
In a tube reactor configuration the substrates have to be placed horizontally and parallel to the gas flow to ensure a uniform flow free of turbulence. An alternative configuration to achieve higher throughputs is a linear nozzle injector. The gases are injected at high speeds from separate ports within the injector with nitrogen or argon curtains to reduce mixing in the gas phase. The high injection speed reduces the residency time and therefore gas phase reactions. The heated substrates typically move horizontally under the injectors.

Due to the challenges in film uniformity, one rarely finds APCVD systems in typical device fabrication cleanrooms. However, they are attractive for some applications such as the manufacture of solar cells where the devices are made on large flat panels and low cost and high throughputs are more important than high quality and uniformity [31].

### 3.2.2 Low-Pressure Chemical Vapor Deposition

In LPCVD, the pressure is in the range of 1 Torr or lower. From Equation 3.37, we can verify that \( h_g \) will get larger, so the primary rate-determining step will be the surface rates (adsorption, desorption, and surface reaction rates). This is case 2 of the two limiting cases we discussed. The primary factors that affect deposition rate will be temperature and pressure. Due to the low pressures, the deposition rates will be lower, typically in the range of 10–100 Å/min. Most importantly, maintaining uniform temperature and pressure is easier than maintaining uniform flow characteristics. This is the main advantage of LPCVD over APCVD. Since the deposition rate is independent of the gas flow characteristics, it is possible to stack the wafers vertically very close to each other in a tube furnace to improve manufacturing throughput. However, microscopic features on the substrate may still encounter diffusion barriers that will limit the step coverage over those structures. For this reason, a high desorption rate \( k_d \) is beneficial, as this would increase the overall surface resistance and minimize any effects from \( h_g \). This is referred to as a “high exchange flux” condition, where only a small fraction of the species arriving at the substrate participate in the reaction, while a large portion is desorbed from the surface [32].

Since the gas flow pattern is less critical in LPCVD, the reactor can be configured as a horizontal or vertical tube, depending on how the wafers are stacked. Figure 3.24 shows a horizontal tube setup. Due to the high exchange flux condition, a large portion of the reaction species go unused and will end up in the exhaust stream. Due to the hazardous nature of most LPCVD precursors, an exhaust scrubber is required before the gases can be vented out. It is also possible to recycle the gases, but such a system is rarely used due to concerns of contamination.

The high temperature and conformal nature of the deposition makes LPCVD ideal for making dense pinhole-free thin films. As a result, LPCVD is widely used in electronics, photonics, and MEMS. The most common films are Si₃N₄ (silicon nitride) and SiO₂ at temperatures ranging from 500°C to 800°C. Although some LPCVD depositions, such as tungsten, can be done at lower temperatures, the main
drawback of this method is the high temperature. This eliminates certain substrates and temperature-sensitive devices, including CMOS electronics in the later stages of the fabrication process.

### 3.2.3 Plasma-Enhanced Chemical Vapor Deposition

For CMOS electronics, there was a need to develop films similar in quality to LPCVD but at lower substrate temperatures. For example, when making metal interconnects on an integrated circuit, each metal layer has to be interleaved between insulating dielectric layers, and exposing the nearly complete chips to LPCVD temperatures would have a detrimental impact on their thermal budget. In PECVD, the pressure is maintained at similar values as LPCVD (<1 Torr), but the substrate temperature is lowered to around 300°C. At this temperature, the LPCVD reaction rate would be nearly zero. Therefore, RF plasma excitation is used to reduce the activation barrier ($E_{as}$ in Equation 3.29). The plasma decomposes the gases into free radicals, lowers the activation energy, and significantly accelerates the reaction rate.

Since this is still a surface reaction rate-limited process, the gas flow characteristics will not significantly affect the uniformity, just as it was with LPCVD. However, the film growth rate will be a strong function of the plasma density and the proximity of the wafers to the plasma. This places restrictions on the chamber configurations and substrates placement. The simplest RF plasma is a parallel plate configuration, which will require the substrates to be placed horizontally similar to the APCVD system. Figure 3.25 shows an example of a parallel plate PECVD system. The wafers are placed horizontally on the heated plate in close proximity to the cathode. The cathode plate also contains a gas shower head to uniformly distribute the gas, and the gas is drawn symmetrically around the hot plate by four pump inlets. Horizontal tube configurations similar to LPCVD can also be used, but special electrode configurations will be needed to ensure a uniform plasma density in the entire tube volume.

The RF discharge powers are generally kept low, at the minimum power necessary to overcome the activation energy $E_{as}$. Excessive discharge powers can increase the probability of reaction in the gas phase and can produce particulates and PVD-like properties. Also, at high RF powers, the reaction point will move from being surface reaction-limited to transport-limited, and nonuniformities due to gas flow patterns can also start to emerge.

A drawback of PECVD films compared to LPCVD is the inclusion of high levels of byproducts and other gases. Hydride-based PECVD reactions generally contain excess hydrogen in the films. This arises due to the low temperature that prevents the gases from fully desorbing during the
reactions. As a result, achieving stoichiometrically correct films becomes difficult, which can affect their physical properties and resistance to chemical etching.

The ion bombardment of the plasma can densify the film, which can be beneficial, but it can also lead to a compressively stressed film. Ion damage to other components on the substrate can also be a concern in PECVD. Despite these disadvantages, PECVD is widely used for making the dielectric layers for metal interconnects in CMOS manufacturing [33].

For more comprehensive reviews of CVD processes, the reader is directed to the references listed at the end of this chapter [29,31,34–36].

3.2.4 ATOMIC LAYER DEPOSITION

ALD is relatively new and is quickly growing to become a mainstream deposition method. It is a variant of the conventional CVD process, except the precursors are flowed sequentially rather than simultaneously. Assuming two precursors containing the species $A$ and $B$ to produce a compound $AB$, the precursor containing $A$ is flowed first, and the molecules are allowed to attach to the substrate by adsorption. The chamber is then purged with an inert gas, and the precursor containing $B$ is flowed next. The $B$ precursor molecules will react with the adsorbed $A$ precursor molecules and produce the film $AB$ (see Figure 3.26). Since there is only a finite number of $A$ precursor molecules adsorbed on the surface, the reaction is self-limiting, that is, it will come to a stop once all the adsorbed $A$ precursor molecules have reacted. The film produced during one cycle will typically be one monolayer thick, or less, depending on the extent of coverage and the sticking coefficient of the precursors on the substrate. This cycle is repeated many times to produce the desired film thickness. Because adsorption is a nondirectional process driven by diffusion, all surfaces will be coated equally with excellent step coverage, as shown by the example in Figure 3.27. ALD is best suited for applications that need very thin, very-high-quality, pinhole-free films. The gate oxide of MOS transistors is the most coveted application for ALD. While the supremacy of silicon originally relied on its ability to form high quality thermal oxides as the gate dielectric of MOS transistors, ALD has made it possible to make gate dielectrics from other materials with even higher dielectric constants, such as $\text{Al}_2\text{O}_3$ and $\text{HfO}_2$.

$\text{Al}_2\text{O}_3$ and $\text{TiO}_2$ are some of the most developed processes used for ALD. $\text{Al}_2\text{O}_3$ is created using trimethylaluminum ($\text{Al(CH}_3)_3$) and $\text{H}_2\text{O}$ as the two precursor gases. $\text{TiO}_2$ is created by $\text{TiCl}_4$ and $\text{H}_2\text{O}$. The reactions proceed as follows [38]:

![Figure 3.26 Pulse and purge description of an atomic layer deposition process using two gases precursors A and B. (a) Precursor gas containing A is flowed and allowed to attach to the substrate; (b) purge gas removes all traces of precursor A from the chamber; (c) precursor gas containing B is flowed and allowed to react with the adsorbed A precursor molecules; and (d) purge gas removes all traces of precursor B from the chamber.](image-url)
The process temperature can be anywhere from room temperature to 300°C. The temperature not only affects the reaction rate, but also the adsorption and desorption rates. Therefore, higher temperatures do not necessarily lead to higher deposition rates.

While binary compounds are the most developed for ALD, single-element metals can also be deposited by reducing a metal-containing precursor. These processes are generally more difficult than binaries and require plasma activation of hydrogen. Nevertheless, processes for a number of metals such as tantalum, silicon, germanium, and tungsten have been developed, and the list continues to grow. Although ALD is a very promising technique, its major disadvantage is the slow rate of growth due to the pulse and purge cycles.

### 3.3 THIN-FILM MEASUREMENTS

#### 3.3.1 THICKNESS MEASUREMENT WITH A QUARTZ CRYSTAL MICROBALANCE

Measuring the deposition rate of incident atoms in real time is an important requirement in thin-film deposition. The quartz crystal microbalance (QCM) is the most widely used method for in situ measurement of thin films. This is based on measuring the mass loading effect on a piezoelectric crystal. One may also be familiar with quartz crystal oscillators as precise timekeeping components in electronic circuits and watches. QCMs are also used in a large number of chemical and biological applications to measure small changes in mass [39,40].

Quartz is the crystalline form of SiO₂. It is also a piezoelectric material. This means that when a voltage is applied across the crystal faces, it will induce a strain. The crystal orientation of the quartz substrate is selected such that a transverse strain is induced when a voltage is applied across the two parallel faces of the substrate. If the voltage is oscillating, a transverse acoustic wave will be induced in the crystal. The acoustic wave will reflect off both substrate boundaries and setup a standing wave. This resonance frequency will be a function of the crystal thickness and the
acoustic velocity in quartz. The faces will be displacement anti-nodes at this resonance. Most commercially available QCMs come in round discs metallized on both sides for electrical connections, as shown in Figure 3.28. The nominal thickness of the crystal is 275 μm, which produces a fundamental resonance frequency of 6 MHz, as well as a number of other harmonics.

The resonance frequency is measured by using the crystal as a feedback element of an amplifier circuit (see Figure 3.29). The electrical impedance of the crystal reaches its lowest value when the frequency of the electrical signal is exactly equal to the resonance frequency of the acoustic signal. When used as a feedback element, the amplifier will enter a self-sustained oscillation at this frequency, which can then be measured with a frequency counter.

Since the face of the crystal is an anti-node of the standing wave, a thin film that becomes attached to the face will act as an additional inertia that will reduce the vibration frequency. The relationship is described in the Sauerbrey equation:

$$\frac{\Delta f}{f} = \frac{2f}{Z_q} m$$  \hspace{1cm} (3.38)$$

where

- $f$ is the resonance frequency
- $\Delta f$ is the frequency shift

FIGURE 3.28 Quartz crystal microbalance.

![Figure 3.28 Quartz crystal microbalance.](image)

FIGURE 3.29 Active oscillator circuit with quartz crystal feedback.

![Figure 3.29 Active oscillator circuit with quartz crystal feedback.](image)
\[ Z_q \] is the acoustic impedance of quartz
\[ m \] is the mass per unit area of the film

If the density of the film is known, the mass per unit area can be converted to a thickness value. This equation treats the film as an infinitesimally thin mass attached to the end of the quartz crystal face. As a result, it only works for very thin films that produce frequency changes of 2% or less. For thicker films, the acoustic impedance of the film must also be taken into consideration. This is done in the so-called Z-match method. It is an extension of the Sauerbrey equation and includes the acoustic impedance of the film material [39]. The acoustic impedance of the film is specified as a ratio compared to that of quartz, and is known as the z-ratio. This is the most widely used method in thin-film deposition systems. In order to translate the change in oscillation frequency \( \Delta f \) to a film thickness, the model requires two parameters for each film: its density and z-ratio.

In a typical application, a fresh crystal starts with a frequency of 6 MHz. As films are deposited, the frequency will drop. Generally, the sensitivity of the crystal starts to degrade when its frequency falls significantly below it nominal value, such as 5 MHz. When this happens, the crystal should be discarded and a new crystal installed. Quartz crystals are inexpensive and can be replaced often.

To ensure correct readings, the QCM is installed inside the vacuum chamber in the path of the incident deposition flux, as close as possible to the substrate location. A typical example of a QCM installation in an evaporation chamber is shown in Figure 3.30.

As with any measurement system, there are a number of factors that produce inaccurate results from a QCM that are worth discussing.

### 3.3.1 Temperature Sensitivity

The crystal angle (known as the AT-cut) is selected to minimize the sensitivity of QCM frequency to temperature changes. However, this sensitivity is designed to be a minimum only at room temperature. The temperature sensitivity is worse at elevated temperatures that may be encountered in thin-film deposition systems. In thermal evaporation systems, the crystal is exposed to a significant amount of radiative heating. Therefore, the crystals are normally installed in a water-cooled fixture to maintain a constant temperature. Nevertheless, the front face of the crystal can still experience temperature swings due to the low thermal conductivity of quartz. A rise in

![Figure 3.30](image-url)
temperature will increase the resonance frequency and register as a negative rate of film growth, and a drop in temperature will reduce the frequency and register as a positive rate of film growth. In a typical deposition cycle, the crystal will under-indicate during the ramp up and over-indicate during the ramp down. This aspect may not make a huge difference in thick-film applications where the ramp-up and ramp-down portions are relatively short compared to the total deposition time, but it will significantly affect the results when trying to grow very thin films in the range of 10–50 Å.

3.3.1.2 Tooling Factor
To get accurate indications, the crystal should be located as close to the substrate as possible. However, due to the finite substrate sizes, some difference in thickness between the QCM and the substrate is unavoidable. This difference should be a fixed number and can be calibrated out during trial runs. The ratio $t_{\text{substrate}}/t_{\text{QCM}}$ is known as the tooling factor (TF). Since the substrate is normally placed at the center of the deposition flux and the QCM at the edge, the TF is usually greater than 1. Typical numbers are 1.0–1.2 depending on the geometry of the system.

In thermal evaporation, the deposition rate follows a cosine distribution fairly accurately so the TF should also follow a cosine value and should remain constant over a wide range of operations. This is not the case with sputter deposition. The distribution of the deposition flux is a function of pressure and power, so the TF will also change with these parameters.

3.3.1.3 Film Stress
QCM relies on having stress-free films with good adhesion to the crystal face. A compressively stressed or tensile-stressed film will produce a lateral force on the crystal face and change its curvature, especially after a significant accumulation of film material. This will affect the crystal resonance and produce unreliable readings. Excessively stressed films can also create internal defects and voids. These areas will change the way the acoustic fields propagate through the film and affect the readings. It is also possible for the films to entirely delaminate from the QCM and show a large frequency jump, which will be indicated as a large negative deposition rate.

3.3.1.4 Deposition Energy
The energy of the incident atoms can also affect the QCM operation. In sputter deposition, the QCM will be bombarded with high-energy atoms that will transfer momentum to the crystal and produce random acoustic excitations. These excitations will appear as noise in the QCM resonance. As a result, it is normal to see large fluctuations in QCM indications during sputter deposition, especially at higher discharge powers and lower pressures. In thermal evaporation, the readings will contain a lot less noise. As a result, QCM is less reliable in sputter deposition systems than in evaporation systems.

3.3.1.5 Density and $z$-Ratio
Although the density of most materials is well known and documented, the film being grown in a particular chamber using a particular process will rarely conform to these values. Sputter deposition produces a denser film than evaporation. Pressure, power, and substrate temperatures can also affect the density. They can also vary greatly depending on if the film is polycrystalline or amorphous. The vast majority of films produced by PVD are amorphous or polycrystalline. For some materials, the density values quoted in standard databases are for their crystalline form. For example, the density of silicon, germanium, silicon dioxide, titanium dioxide, etc., found in databases are most likely for their crystalline form. Using these values in a QCM measurement in a PVD system can lead to errors. The $z$-ratio has even greater uncertainties. The acoustic impedances of many materials are simply not known. In this case, the default value of 1.0 is used, which essentially assumes the acoustic impedance of the film to be the same as that of quartz.
3.3.2 **Thickness Measurement with a Stylus Profiler**

While QCM is the most commonly used method for monitoring the thickness during deposition, it has a number of shortcomings as described in the previous section. One always needs a reliable technique to verify the film thickness after the deposition so that the QCM can be calibrated. Probably the most reliable and repeatable thickness measurement tool is the contact stylus profiler. It measures a step height variation by dragging a fine tip across the sample and measuring its deflection. If all sources of vibrations can be reduced, the accuracy and repeatability of this method can be as little as 1 nm.

In order to measure the thickness of a deposited film, a sharp step is required. This is most easily done by placing an ink dot with a fine-tipped marker on a companion substrate, performing the deposition and dissolving the ink away in a solvent such as acetone. This would expose the substrate and the sharp transition step at the film/substrate interface can be used to measure the film height with a profilometer (Figure 3.31).

3.3.3 **Measurement of Optical Properties**

One could measure the optical transmission and reflection of the film in real-time through transparent viewports in the deposition chamber. An ellipsometer can also be used. But these techniques require the film to be transparent at the interrogation wavelength and would also require the viewports to remain clean and free of deposition buildup.

Similar to the stylus profiler, one could also use a companion sample to measure the optical reflection and transmission as a function of wavelength after the deposition has been completed. This can then be fitted to known dispersion models of materials to extract the refractive index and film thickness. This method is particularly useful for optical thin-film applications where the optical thickness may be slightly different from the physical thickness measured from the stylus profiler.

3.3.4 **Thin-Film Stress**

Film stress is the tendency of a film to want to shrink or expand. If it wants to shrink but it is prevented to do so due to its attachment to the substrate, the molecules in the film will be held in a stretched position. This produces a film that is under tensile stress. If it wants to expand, but it is...
prevented from doing so, the molecules will be held in a compressed position. This is a film with a compressive stress.

When the compressive or tensile forces exceed the adhesion force between the film and substrate, the film can peel, crack, buckle, or delaminate. Even if the film does not separate, it will change the curvature of the underlying substrate. In applications with tight mechanical tolerances, this can become a problem. Therefore, understanding and controlling stress is an important aspect of thin-film engineering.

3.3.4.1 Origins of Film Stress

The deposition process itself introduces stresses in a film. For example, in sputter deposition the high energy of the incident species will compact the film into a compressed state. The distances between the atoms will be smaller than their equilibrium distances. After the deposition, the film will want to expand to reduce its internal energy, which is when it will warp the substrate or delaminate. In thermal evaporation, the opposite can happen. Due to the very low energies, the film can have a larger atomic distance than its equilibrium distances [41]. After the deposition, atoms may migrate closer together causing the film to become stressed in a tensile state. In CVD, the films can be tensile or compressive depending on the reaction pathways. Reactions with byproducts that desorb from the substrate often tend to form tensile films. Inclusion reactions produce compressive films. For example, Si$_3$N$_4$ from SiH$_4$ and NH$_3$ produces tensile films due to the desorption of hydrogen, but SiO$_2$ by direct oxidation of silicon produces compressive films due to the inclusion of oxygen. In PECVD, film stress can be controlled to some degree by utilizing the ion bombardment effect. All of these stresses are known as intrinsic stress or residual stress.

Extrinsic stress is caused by external means, such as changes in temperature. Any mismatch in the coefficient of thermal expansion between the film and the substrate will induce a stress when the temperature changes. Absorption of moisture or other vapors into the film can also induce extrinsic stress. Extrinsic stresses are usually smaller in magnitude than intrinsic stresses.

3.3.4.2 Measurement of Stress

Stress is defined in the same units as pressure. This is the amount of force that exists within a cross-sectional area of the film. Imagine a stretched wire. This will be under tensile stress. If we take the force that keeps the wire stretched and divide it by its cross-sectional area, we will get the stress value. The most commonly used unit for stress is Pa, where 1 Pa = 1 N/m$^2$. However, the magnitudes commonly encountered in thin films are in MPa or GPa. Tensile stresses are designated with positive values and compressive stresses with negative values.

Stress in a film can be measured using a number of different methods. Raman spectroscopy can often reveal stresses in a film. Acoustic methods and infrared spectroscopy can also be used. However, the most widely used method is a direct measurement of the substrate curvature induced by the film stress. By measuring the curvature before and after the thin-film deposition, the stress can be accurately evaluated. It may seem surprising that a thin film of the order of 1000 Å can bend a 0.5 mm thick substrate, but the curvature we are speaking of is extremely small and imperceptible to the eye. For example, the radius of curvature before deposition might be 100 m (yes, meters), and after deposition it might be 80 m. This small change in curvature indicates a tensile stressed film.

The relationship between curvature and stress is given by Stoney’s equation:

$$\sigma = \left( \frac{E}{1-\nu} \right) \left( \frac{t_s^2}{6t_f} \right) \left( \frac{1}{r_e} - \frac{1}{r_b} \right)$$  \hspace{1cm} (3.39)

where

- $E$ is the Young’s modulus of the substrate
- $\nu$ is the Poisson’s ratio of the substrate
- $t_s$ is the substrate thickness
\( t_f \) is the film thickness  
\( r_a \) is the radius of curvature after deposition  
\( r_b \) is the radius of curvature before deposition

This model assumes that the strain energy is stored entirely in the substrate. This is the reason for why the mechanical properties of the film are not contained in the equation. Since silicon is a very common substrate, it is useful to note its values. For (100) wafers along the (110) direction, \( E_{Si} = 171 \text{ GPa} \) and \( \nu_{Si} = 0.06 \), and along the (100) direction it is \( E_{Si} = 130 \text{ GPa} \) and \( \nu_{Si} = 0.28 \) [42]. For example, consider a 5000 Å film on a 500 \( \mu \text{m} \) silicon substrate. If the pre- and post-deposition radii of curvatures were \(-50 \text{ m}\) and \(-30 \text{ m}\), respectively, along the (110) direction, the film stress can be calculated to be \(-153 \text{ MPa} \) (compressive).

The radius of curvature can be measured very accurately using a number of different methods. One method is to use a stylus profiler over a very long scan length of at least 1 inch. It is important to scan at exactly the same location before and after the deposition because the intrinsic substrate curvature may not be symmetric. Figure 3.32 shows an example of a profile scan before and after the film deposition. Another approach is to use a noncontact optical interference method. This method can produce a 2D map of the surface topography that can then be compared and subtracted after the deposition. This method can produce a stress map of an entire wafer.

Another useful number is the stress value multiplied by the film thickness. This quantity will be in the units of force per length and represents the shear force imparted by the film on the substrate per unit length. As the film thickness increases, the stress value may remain the same, but the shear force will increase. At a certain shear force, the film may separate from the substrate. Therefore, delaminations are more likely to occur when the films grow thicker, even if the stress values are the same.

FIGURE 3.32 Measurement of the radius of curvature with a stylus profilometer.
3.3.4.3 Compressive Stress

Compressive stress is noted with a negative stress value. It causes the substrate to become more convex-shaped (as viewed from the film side). Too much compressive stress will tend to lift the film off the substrate in order to make it more convex. This will show up as blistering and buckling. To the naked eye, these buckling delaminations may only appear as a discoloration or as a surface texturing, but under a microscope the buckles should be easy to identify. An example of buckling delaminations of a sputtered silicon film is shown in Figure 3.33.

3.3.4.4 Tensile Stress

Tensile stress has a positive stress value. It causes the substrate to become more concave as viewed from the film side. Excessive tensile stress can cause cracks in the film or it can separate from the substrate at the edges and curl inward to make a more concave film. Very thick films produced by thermal evaporation can display this character. Figure 3.34 shows an example of curling from the substrate edges due to excessive tensile stress from a thick germanium film made by electron beam evaporation.

![FIGURE 3.33](image1) Example of bucking delamination from an excessive compressive stress viewed under a microscope.

![FIGURE 3.34](image2) Example of peeling due to excessive tensile stress.
3.3.4.5 Stress Reduction
It is possible to reduce intrinsic stress by increasing the energy of the atoms during deposition so that they relax to a low-stress configuration. The most common method to accomplish this is by raising the temperature of the substrate during deposition, in the range of 300°C–500°C. However, in some cases the extrinsic stress could become large when the substrate is cooled down after the deposition. Furthermore, not all substrates can withstand high substrate temperatures. An alternative method is ion bombardment of the thin film during the deposition, as with IAD. The ions carry significant energy (several hundred eV), so they are equivalent to a significantly higher temperature than any substrate heater can accomplish. However, IAD is generally only effective with tensile films. With increasing ion energy, the tensile stress can be reduced, or even be turned into a compressively stressed film.

3.4 THIN-FILM MATERIALS
In this section, a brief summary of the thin-film materials most commonly used in many laboratories is described.

3.4.1 Titanium
Titanium is most commonly deposited by evaporation or sputtering. It is one of the most reactive metals, and therefore requires a very low base pressure and extremely low levels of desorption during deposition. Otherwise, one will end up with an oxidized form of titanium. During evaporation, it is common to notice a decline in the background pressure as titanium starts to evaporate (which is known as the gettering effect). With most other materials, the pressure will actually increase during deposition due to an overall increase in the chamber temperature and the associated outgassing. Titanium is commonly used as an adhesion layer for metals such as gold, silver, and copper. Evaporation works slightly better than sputtering because it is easier to monitor the desorption rates during deposition.

3.4.2 Chromium
Chromium is also reactive like titanium, and it requires a very low base pressure before deposition. It also exhibits some gettering effect during deposition. It can be sputtered or evaporated. Under evaporation, it sublimes rather than melt. With electron beam evaporation, the beam spot should be continuously scanned to prevent evaporating from a single spot because fresh material will not flow in like with other materials that melt. Chromium is also commonly used as an adhesion metal like titanium.

3.4.3 Aluminum
Aluminum is also reactive and requires a low base pressure. It can be sputtered or evaporated. Although it melts at a very low temperature (660°C), it needs to be raised significantly above its melting temperature to create a reasonably high vapor pressure. As a result, it tends to wick into many surfaces and profusely wet the crucible liners and create contaminations.

3.4.4 Copper
The reactivity of copper is not very high, so it can be easily sputtered or evaporated. Copper has a very high electrical conductivity and is widely used as an interconnect metal in electronics.
3.4.5 Gold

Gold is very inert and does not react with background gases under normal conditions. It is easy to evaporate or sputter. However, due to its high cost, evaporation is more economical than sputtering because it can be bought in small pellets as needed for each project. Also due to its high cost, the protective liners used in vacuum chambers (typically aluminum foil) can be sent to reclaim and extract the gold. In most cases, an adhesion layer, such as titanium or chromium is necessary before depositing gold. It is often used as the final metallization in devices because it will not oxidize or corrode.

3.4.6 Silver

Silver evaporates and sputters well, and also has fairly low reactivity. It is a soft metal like gold and requires an adhesion layer. When deposited at very small thicknesses (<10 nm) directly on silicon or silicon dioxide surfaces, it forms distinct droplets due to its poor wetting properties. Thin silver films have reasonable transparency in the visible spectrum, so they are used as infrared blocking films on architectural windows.

3.4.7 Platinum

Like gold, platinum is an expensive metal, so although it can be sputtered, it is more economical to evaporate it from small pellets. It is also not reactive. Platinum has a very large work function so it can be used as an ohmic contact in p-type semiconductor materials, as well as for making Schottky diodes. It has poor adhesion so it must be used with a metal like titanium. Platinum requires high powers to evaporate because it needs to be raised to a very high temperature in order to produce a reasonable vapor pressure. It melts at 1768°C, but needs to be raised to about 2500°C to produce about 100 mT of vapor pressure.

3.4.8 Nickel

Being a ferromagnetic metal, nickel cannot be sputtered with a magnetron cathode because it would shield the magnetic fields and prevent it from confining the plasma. It can, however, be sputtered if it is sliced into a very thin sheet and then bonded to a copper target. It can be easily evaporated, although with an e-beam evaporator, one needs to be aware of its effects on the focusing magnets. Initially, the beam may not focus properly, but once the nickel reaches its Curie temperature of 355°C, its magnetism will disappear and it will behave normally.

3.4.9 Tungsten

Tungsten has one of the highest melting temperatures, therefore, it cannot be easily evaporated. Fortunately, it has a reasonably high sputter yield, so it is most often deposited by sputtering. Tungsten can also be deposited with CVD using WF₆ as the precursor gas. This is used for via-filling applications in integrated circuits.

3.4.10 Molybdenum

Molybdenum is very similar to tungsten and is difficult to evaporate due to its low vapor pressure and high melting temperature, but it can be easily sputtered. It has an interesting property that it etches fast in hydrogen peroxide, forming an oxide that is readily water soluble.

3.4.11 Vanadium

This metal is very similar to titanium, and exhibits a pronounced gettering effect. It can be sputtered or evaporated easily, but the background pressure has to be low to prevent oxidation.
3.4.12  Silicon

Silicon can be sputtered or evaporated, but it is highly reactive so it requires a very low base pressure and low desorption rates. PVD silicon films are not very stable, because the unfilled silicon bonds can oxidize over time. It can also be produced by CVD reactions. One of the common variant is the hydrogenated amorphous silicon that is produced by the pyrolysis of silane gas in a CVD reactor. The hydrogen terminated silicon is very stable, and it is widely used in applications requiring amorphous silicon (such as solar cells).

3.4.13  Germanium

Germanium is similar to silicon, but it is less reactive and can be evaporated or sputtered, or deposited by CVD.

3.4.14  Aluminum Oxide

Al₂O₃ has a very low sputter rate, and for this reason it is used as a shielding material in many plasma fixtures. Sputtering from an Al₂O₃ target is very slow, but it can be deposited by reactive sputtering from an aluminum target. It can also be evaporated fairly easily, directly from a dielectric source or reactively from aluminum. ALD and CVD processes can also be used to produce Al₂O₃.

3.4.15  Magnesium Fluoride

MgF₂ finds many applications in optical coatings because it is one of the few materials with a refractive index lower than SiO₂. It can be sputtered or evaporated easily. It produces very high vapor pressures at low temperatures, so it is really easy to evaporate. However, the evaporated film tends to be highly stressed, so it needs to be used in conjunction with IAD or substrate heating when depositing large thicknesses (which is needed in optical applications due to its low refractive index). The loss of fluorine during deposition is difficult to compensate because one cannot easily flow fluorine as a background gas, so a common technique is to use a small oxygen background to convert any metallic Mg to MgO to keep the optical film losses low (albeit at a slightly increased refractive index).

3.4.16  Silicon Dioxide

SiO₂ is the ubiquitous “oxide” and can be sputtered, evaporated, or deposited by CVD, among many other techniques. Sputter rates tend to be very slow, but evaporation can be fast. SiO₂ is so stable that oxygen loss during deposition is usually not a problem, but a small oxygen background can help, and does not seem to cause any adverse effects to the film.

3.4.17  Titanium Dioxide

TiO₂ is a popular optical film because it has one of the highest refractive indices in the visible spectrum (about 2.5). It sputters very slowly, and the targets tend to crack very easily due to thermal gradients during sputtering. Evaporation of TiO₂ is also not straightforward because it easily decomposes to form one of many other stable oxide states, such as Ti₂O₃ and Ti₂O₅. Normally, one evaporates Ti₂O₃ pellets under an oxygen ambient and substrate heating to create TiO₂ on the substrate.

3.4.18  Niobium Oxide

Nb₂O₅ is also a dielectric of interest in optical thin-film applications because of its high refractive index of 2.3 in the visible spectrum. It is an attractive alternative to TiO₂ albeit at a slightly lower refractive index. It can be easily sputtered with a small oxygen background gas.
3.4.19 **Zinc Sulfide**

ZnS is also a material of significant interest in optical applications. It has a reasonably high refractive index of 2.3 in the visible spectrum. It is also used as an infrared thin film because of its extended transparency in the mid-wave and long-wave infrared spectrum. ZnS is most easily deposited by sputtering. It can also be evaporated but the zinc contamination of the vacuum chamber should be considered before evaporating ZnS.

3.4.20 **Vanadium Oxide**

VO₂ is an interesting material because it exhibits an insulator-to-metal phase transition at 68°C. It is usually deposited from a metallic vanadium source by sputtering, evaporation, or PLD in an oxygen ambient at elevated substrate temperatures. Like TiO₂, this material exhibits mixed valence states. VOₓ is a mixed oxide of vanadium that is much easier to deposit than VO₂. It has a large temperature coefficient of resistance, and is used as the sensor element in thermal imaging cameras.

**PROBLEMS**

3.1Consider thermal evaporation of platinum on a 4 in. diameter substrate in a vacuum chamber with a throw distance of 45 cm. If the desired deposition rate is 3 Å/s, assuming the melted source area is 5 mm², calculate the approximate temperature of the metal. Determine if the metal will be melting or sublimating. Calculate the expected center-to-edge variation in film thickness on the 4 in. substrate.

3.2 If aluminum is sputtered with argon at 200 W discharge power and 200 V cathode voltage, calculate the average target removal rate in Å/s.

3.3 Referring to Figure 3.15, determine what would happen to the electron trajectories if the inner magnet is flipped so that its N pole is facing up.

3.4 Using the SRIM software, calculate the average sputter yield and the energy distribution of the sputtered atoms for copper and tungsten for normally incident Ar⁺ ions with an energy of 500 eV.

**LABORATORY EXERCISES**

3.1 Perform a thermal evaporation on the largest possible substrate that can be accommodated, and take thickness measurements at various radial positions. Examine if the uniformity fits the expected cos(θ) distribution.

3.2 Install two small (approx. 1 inch × 1 inch) silicon substrates in a thermal evaporation chamber. The first one should be mounted perpendicular to the deposition flux (which is the normal configuration), and the second one should be at a very steep angle of incidence, such as 85° or greater. An angle block may have to be utilized for the latter. Then deposit approximately 1000 Å of titanium. After the deposition, the film on the first sample should visually appear as a normal metallic film, and the second sample should appear dark in color. Examine the cross-sections of the films under a scanning electron microscope. The first sample should not show any structure but the second sample should contain tilted nano-columns. This is known as glancing angle deposition [22]. Investigate how the angle of incidence of the deposition flux plays a role in this growth mechanism.

3.3 For the above two samples, measure the substrate curvature with a stylus profilometer before and after the deposition. The profile scans must be taken at exactly the same locations. Apply Stoney’s equation to calculate the film stress for both cases, and determine if it is compressive or tensile.
REFERENCES