

Nanofabrication

5

A. Sarangan

University of Dayton, Dayton, OH, USA

5.1 Nanofabrication

There is no single accepted definition of nanofabrication, nor a definition of what separates nanofabrication from microfabrication. To meet the continuing challenge of shrinking component size in microelectronics, new tools and techniques are continuously being developed. Component sizes that were in tens of micrometers became single-digit micrometers, and then hundreds of nanometers, and then went down to a few tens of nanometers where they stand today. As a result, what used to be called microfabrication was rebranded as nanofabrication, although the governing principles have remained essentially the same. The main driver of this technology has been the manufacture of integrated circuits, but there have been tremendous fallout benefits to other areas, including photonics.

Nanofabrication can be loosely divided into three major areas: thin films, lithography, and etching. Each of these are vast subject areas in and of themselves, but in this chapter we attempt to cover their essential concepts in a concise fashion for someone new to these areas. The goal is not to provide a working experience that allows one to walk into a laboratory and perform these tasks, but to provide an overall understanding of what these areas are as well as the pros and cons of the most commonly used techniques. Hopefully it will build a foundation for more specific training for anyone who wants to venture further into these areas.

5.2 Thin films

Thin film science is a vast and mature area that permeates nearly all disciplines. Thin films are found in almost every manufactured product, from eyeglasses to display screens to automobiles and aerospace components. They are used to achieve various purposes, such as optical modification (antireflection), chemical modification (corrosion inhibition), mechanical modification (scratch resistance), electrical modifications and thermal modifications.

In this section, we focus primarily on films on the order of a few tens of nanometers thick for electronic and photonic applications. One example is the gate dielectric in metal-oxide-semiconductor (MOS) transistors. This dielectric is often less than 10 nm thick, and in many cases it approaches 2 nm. Its properties such as dielectric constant, thickness, and defect density are critical parameters that affect the transconductance and switching characteristics of the devices. Another area in which nanoscale films are widely used is in optical filters. In the visible and infrared range, film

thicknesses are in the range of 10–1000 nm. In the extreme-ultraviolet (EUV) range, the multilayer films used to make photomasks are in the range of 1–2 nm. Hence, these films span a wide range of thicknesses depending on their applications.

In the following sections, we will discuss some of the commonly used methods for producing thin films of interest in the field of nanophotonics. We only cover the most salient features of each technique and their pros and cons. The reference list at the end of the chapter provides the interested reader more sources to learn about these topics in greater depth [1–8].

The properties of thin films vary greatly depending on the method used to make them. The typical properties of interest are

- *Conformal nature of the film:* Conformal films grow on all topographic features, whereas nonconformal films are line-of-sight and grow only on certain planes.
- *Film density:* Although materials have standard densities, the films made from different techniques can have a lower or higher density.
- *Dielectric constant:* Higher densities generally result in a higher refractive index, although other factors such as contamination and compound formations can alter the refractive index from their nominal bulk values.
- *Stress:* Films under high stress can introduce curvatures to the substrate, or in extreme cases they can separate from the substrate by peeling or cracking. Stress will also affect the refractive index of the film.
- *Chemical composition:* In single elements, this is not usually a concern, but in compounds, the film stoichiometry can vary greatly depending on the growth conditions.
- *Electrical conductivity:* Grain size, packing density, and impurities can significantly alter the conductivity of thin films.

The methods used to make thin films can be broadly classified as physical methods and chemical methods. Physical methods involve transferring the material from the source to the substrate without changing its chemical state. In chemical methods, the film is created as a by-product of a chemical reaction.

5.2.1 Physical methods

By far the most commonly used physical methods are evaporation and sputtering. Both are vapor coating methods in which the source material is transferred atom-by-atom to the substrate. These are generally referred to as physical vapor deposition (PVD). PVD is usually done in a high-vacuum chamber to reduce the interaction with gas species in the environment.

5.2.1.1 Evaporation

Evaporation is easy to understand—the source material is heated to a high temperature until it starts to evaporate. The important parameter here is vapor pressure. Every material has a characteristic vapor pressure as a function of its temperature, as shown in [Figure 5.1](#). In other words, every material is always evaporating. For most solids at room temperature, this vapor pressure is extremely low. For example, consider gold. Its vapor pressure at room temperature is well below 10^{-15} torr and cannot be detected.

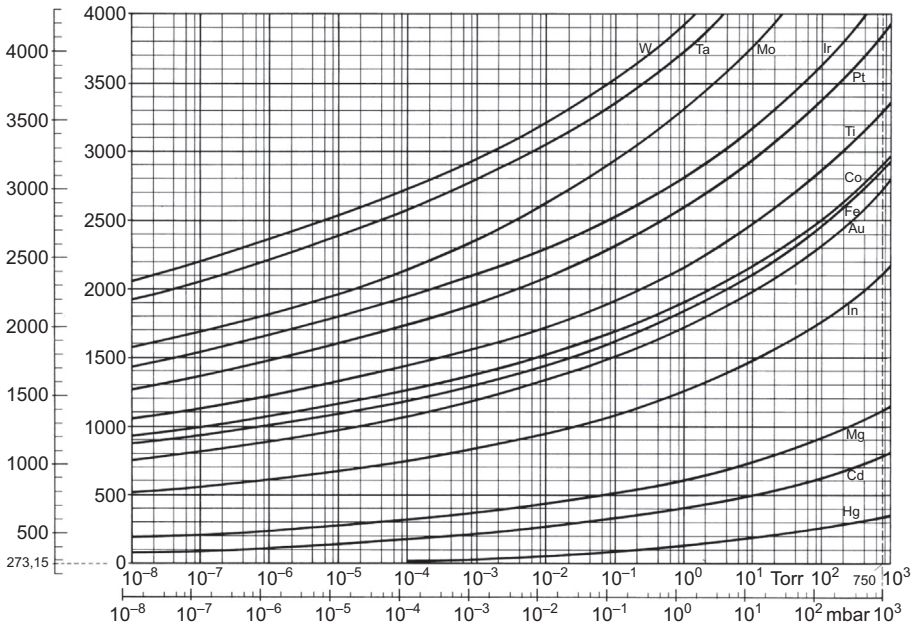


Figure 5.1 Vapor pressure versus temperature curve for metals.

Source: Fundamentals of Vacuum Technology, Oerlikon Leybold Vacuum, Figure 9.13, 2007. Reproduced with permission.

At 1000 °C, its vapor pressure rises to 20 μ Torr, and at 1500 °C it becomes 100 mTorr. In addition, the vapor pressure will also be a function of position inside of the vapor flux. The pressure will drop as the vapor diffuses away from the evaporating source. The vapor pressure at the substrate surface can be estimated from simple geometrical considerations of the vapor flux. From ideal gas laws, we can then relate pressure to the impingement rate of atoms. If we further assume that all impinging atoms condense on the substrate surface, then we can calculate the growth rate of the film on the substrate. Although this is the physical description of the evaporation–condensation phenomena, such calculations are rarely done in practice because the source temperature and vapor pressures are typically not measured. Instead, one monitors the deposition rate on the substrate by controlling the power delivered to the evaporating source, leaving source temperature and vapor pressure as intermediate parameters.

The unit of measure in thin film science is Angstroms rather than nanometers, but we will adopt the units of nanometers here for consistency with other chapters. Typical deposition rates are in the range of 0.1–5 nm/s. To reach these levels, most materials have to be heated to near or above their melting temperature. Some materials evaporate by sublimation well below their melting temperature. Examples of sublimating materials are chromium and silicon dioxide. Boiling temperature is not a relevant concept here—this is the temperature at which the vapor pressure is equal to the ambient pressure. Because the ambient pressure in a vacuum chamber is extremely low, every material can be considered to be boiling all of the time in vacuum.

There are two major types of evaporation. They differ by the method of power delivered to the source. In one method, the source in the form of pellets is placed in a metallic boat and is heated by passing a very high current through it, in the range of 100 A. This is known as resistively heated evaporation. An advantage of this technique is its simplicity because only a high current (low-voltage) DC source is needed to power the source. The disadvantage is the lack of heating efficiency and contamination. Because all parts of the heating circuit will be hotter than the source pellet, they will also evaporate to some extent, which can lead to contamination and outgassing.

The second method is electron-beam heating. In this technique, electrons emitted from a heated filament are accelerated by a high voltage in the range of 10 kV and focused to a small spot on the source pellet, as shown in Figure 5.2. Because the process is done in a vacuum, electrons can be easily accelerated and manipulated by magnetic fields without colliding with gas molecules. This method is more efficient because energy is delivered accurately to the source pellet with minimal heating of the other fixtures. As a result, contamination is kept to a minimum. A disadvantage of this method is the complexity of the power source. A high-voltage source is more complex in its design than a low-voltage, high-current source. Safety features are also a major consideration because a 10-kV DC can be lethal. Nevertheless, electron-beam evaporation continues to be one of the most commonly used evaporation methods in thin film research and development.

Because of the physical trajectories of the evaporating species and the high-vacuum environment, the films produced by evaporation tend to be directional and nonconformal. This can be an advantage in many applications, such as lift-off lithography. The highly directional nature of the evaporating flux is also utilized in a special class of thin film known as nanostructured thin films. Instead of containing a randomly packed isotropic structure, these films contain nanocolumns and nanochiral structures, which can be used to engineer the electrical and optical properties of these films. A class of patterning known as shadow-masked lithography also relies on a line-of-sight deposition for which evaporation is ideally suited.

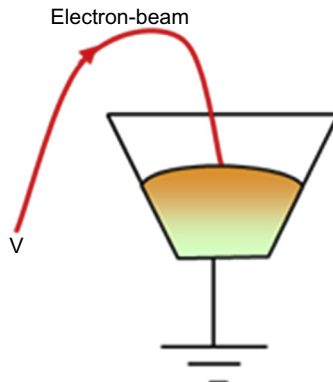


Figure 5.2 Electron-beam evaporation.

5.2.1.2 Sputtering

Sputter deposition utilizes an electrically excited gas plasma in a vacuum system. The ions in the plasma are accelerated toward the cathode, which upon bombardment eject neutral atoms from the cathode surface. The ejected atoms collect on all surfaces including the substrate surface. This is illustrated in [Figure 5.3](#). Therefore, the cathode has to be constructed from the same source as the material being deposited. Unlike evaporation, the source material here is kept at a low temperature by flowing cooling water behind it. Atoms are ejected from the source (which in this context is referred to as the target) by momentum transfer rather than by heat. This is a fundamental difference between sputtering and evaporation. Because of the incident momentum, the resulting films will be more compacted and denser than in evaporation. The low target temperature also enables the deposition of certain compounds such as oxides and nitrides, which may otherwise decompose at elevated temperatures encountered in evaporation. Nevertheless, for complex compounds, pulsed laser deposition (see [Section 5.2.1.3](#)) is the preferred technique over sputtering.

The atomic weight of the plasma gas species and the gas pressure play a significant role in the sputter yield, which is defined as the number of target atoms that are ejected for each incident ion. Ordinarily, the gas species is selected to be chemically inert so that it does not interact with the target material and form compounds. For this reason, argon gas is typically used in most sputter systems. It has a reasonably high atomic mass (~ 40 u, where u is the unified atomic mass unit of 1 g/mol), is inert, and is relatively inexpensive. Sputter yield data for various gas species, target materials, and ion energies can be found in the literature. In general, sputter

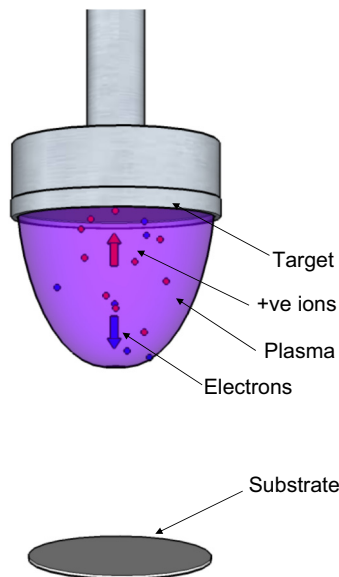


Figure 5.3 Sputter deposition configuration.

yield is related to the atomic number and masses of the plasma ion and the target material. Sputter yields also exhibit a threshold energy. These can be determined through empirical models as well as numerical models using Monte Carlo methods. For example, the sputter yield of Ag from Ar^+ ions is shown in Figure 5.4. This was calculated using the empirical model presented in Ref. [3]. We can see that it shows a threshold energy of 10 eV and rises from 0.7 at 100 eV to 2.7 at 500 eV. When He^+ ions are used, the sputter yield is about an order of magnitude smaller. At very high energies, the yield declines again because of a process known as ion implantation, in which the incident ions bury themselves deeper in the target instead of ejecting surface atoms.

Sputtering is more versatile than evaporation and is more widely used in industrial processes because the targets and plasma sources can be constructed in various shapes to accommodate different coating configurations. They can be circular, rectangular, cylindrical, or other exotic shapes to fit specific needs. By far the most common type of targets used in research laboratories are circular, 2 or 3 inch in diameter. Sputter deposition can take place upward or downward or even sideways, whereas evaporation is typically limited to only an upward configuration.

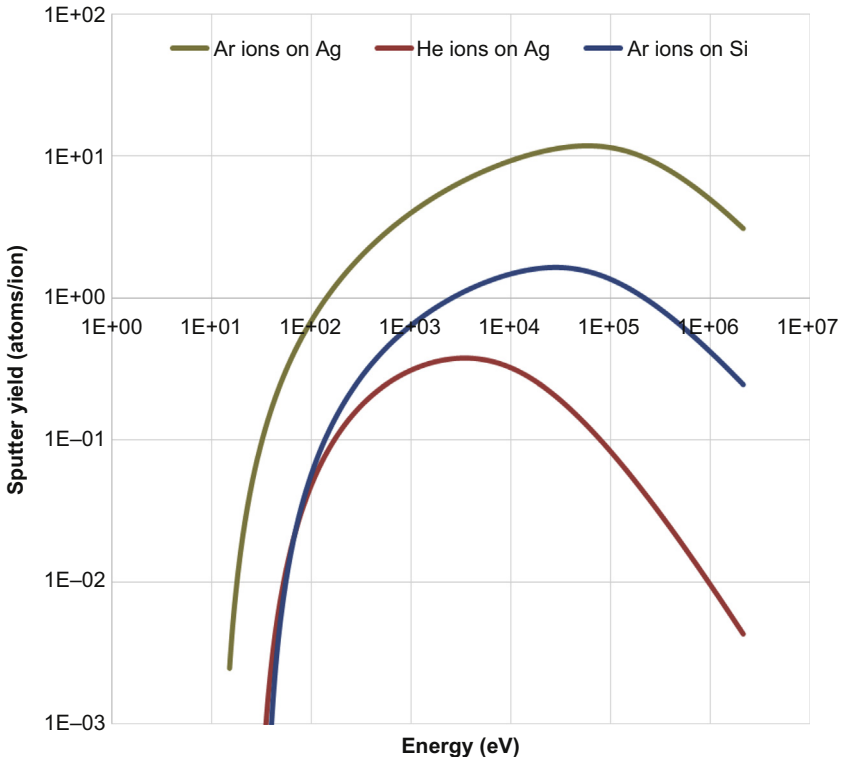


Figure 5.4 Calculated sputter yield versus energy [3].

Power is delivered to the plasma by a DC source or radiofrequency (RF) source. For metallic target materials, a DC source can be used. Insulating targets will require an RF source, where the target acts as a coupling capacitor to the plasma. RF excitation also requires an automatic impedance matching network because the plasma will not have a constant impedance but will vary with its operating conditions such as pressure, material type, target thickness, etc. Many sputter systems used in research tend to utilize RF excitation because it can be used on conductive and nonconductive targets, whereas a DC excitation can be used only on conductive targets.

Because of the presence of the plasma gas, depending on the operating pressure, the sputtered atoms may undergo a few collisions before reaching the substrate. Therefore, sputter deposition can be somewhat less directional than evaporation.

In most cathodes designed for sputtering, magnets are placed behind the target to alter the trajectories of the electrons in the plasma. This has the effect of focusing the plasma close to the target, resulting in a higher ion density and a higher sputter removal rate. These are referred to as magnetron cathodes.

Another variant of sputter deposition is reactive sputtering. In this technique, a small amount of reactive gas such as oxygen or nitrogen is mixed with the argon gas to cause the ejected target species to form compounds. For example, it is possible to create TiO_2 from a metallic Ti by flowing a small amount of oxygen with the argon. However, this is not as simple as it may first appear. Excessive oxygen can cause the target surface to become oxidized and will result in an extremely low sputter yield. This condition is known as target poisoning. The ideal condition is achieved when the oxidation rate of the target surface is exactly balanced by the removal rate of the oxide, and the ejected species are dominantly metallic, which are subsequently oxidized on the substrate surface to achieve the correct stoichiometry. Furthermore, these oxides and nitrides may have several different stable stoichiometric states; therefore, simply mixing a reactive gas will not yield the desired result. For example, vanadium oxide (VO_2) has interesting thermal properties and is used in MEMS actuators. However, during reactive sputtering of a metallic vanadium target with an oxygen gas, the resulting film will contain a mixture of VO_2 , V_2O_3 , V_3O_5 , etc. Deposition conditions such as oxygen partial pressure, substrate temperature, and plasma discharge power are all used in combination to direct the resulting species toward a dominantly VO_2 composition. The same is true with TiO_2 , which is used as the high-index film in multilayer optical filters because it has the highest refractive index in the visible spectrum.

5.2.1.3 Pulsed laser deposition

Pulsed laser deposition (PLD) uses high-energy laser pulses on the order of a few nanoseconds to ablate the target material. The laser radiation is focused on the target surface, is absorbed, and rapidly evaporates the material, resulting in the ablation of atoms, which are subsequently collected on the substrate. Because the energy source (laser) is outside of the vacuum chamber, this method has the advantage of being done in an ultra-high vacuum or under a wide range of ambient pressures and gas species.

The target is generally rotated so that the same spot is not repeatedly ablated. A schematic of a PLD setup is illustrated in [Figure 5.5](#).

The biggest advantage of PLD is the stoichiometric removal of the target material (i.e., the atoms are removed without discrimination). This is due to the fast, transient nature of the ablation process and the high laser fluence, which creates a surface temperature of approximately 5000 K within a few nanoseconds. Short pulses of fluence levels much higher than the ablation threshold allow all components of the target to be ablated equally, irrespective of their binding energies. In contrast, in evaporation and sputtering the volatility of the species affects the removal rate. In sputtering, the material with the higher sputter yield will be removed faster, and in evaporation the material with the higher vapor pressure will be removed faster. As a result, PLD is most often used for the deposition of complex ceramic films such as yttrium barium copper oxide (YBCO); lead zirconium titanate (PZT); and many other carbides, oxides, and nitrides that are difficult or impossible to deposit with other methods. However, stoichiometric removal does not always imply stoichiometric deposition because some of the more volatile species can sputter off of the substrates or simply evaporate. Some compensation techniques are necessary, such as background reactive gases and the use of multiple targets.

Most of the currently used PLD systems utilize excimer lasers emitting in the ultraviolet (UV) range, such as KrF (248 nm), ArF (193 nm), and F₂ (157 nm). UV light is desirable because of its short penetration depth in materials, resulting in the removal of atoms closest to the surface. Longer wavelengths would result in a deeper penetration, which would cause subsurface evaporation and eruption that can lead to larger clusters being ejected. However, in practice, the availability of UV lasers, vacuum windows, and other optical elements limit the lower wavelength to approximately 200 nm.

PLD also allows the user to tune the energy of the species arriving at the substrate by introducing an inert gas atmosphere such as argon. The gas atmosphere increases the scattering rate of the ejected species and slows them down. This technique can be used to produce changes in the film morphology and stress. It is also possible to create nanoparticle deposition by increasing the gas pressure until the ejected species collide with each other to form clusters of atoms. The size of the nanoparticles can also be controlled this way.

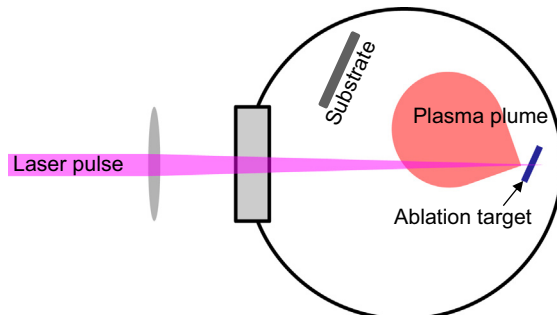


Figure 5.5 Pulsed laser deposition.

Although PLD has advanced significantly in recent years, there are still many challenges. Film uniformity is one of the biggest challenges. Because the ablation occurs at a single point and the plasma plume has a narrow angle, only a small substrate area will get coated during each pulse. The source-to-substrate distance also plays a role. Larger distances will produce a more uniform film, albeit at the expense of a reduced deposition rate. Although the laser beam and the substrate can be continuously translated to improve the film uniformity, it is still inferior to evaporation or sputtering. Another challenge with PLD is the deposition of larger particulates, sometimes up to 10 μm in size, which occur because of molten droplets being ejected during the ablation. These can also be mitigated to some extent by using physical filters.

5.2.2 Chemical methods

Chemical methods can produce films with excellent uniformity, coverage, and stoichiometry, but they require different gases and sometimes different chambers for each film type. The most common chemical method is chemical vapor deposition (CVD). In this technique, gas precursors are introduced into a chamber and the substrate is heated to a sufficiently high temperature to cause a reaction and produce the film of interest. There are different types of CVD, such as low-pressure CVD (LPCVD), atmospheric pressure CVD, plasma-enhanced CVD (PECVD), and atomic layer deposition (ALD).

In CVD, the pressure has to be low enough and the substrate temperature high enough that the reaction only takes place on the substrate surface and not in the gas phase. Gas-phase reactions would result in particles forming and depositing on the surface.

5.2.2.1 Low-pressure CVD

In LPCVD, the substrate is placed in a quartz tube and heated while flowing the precursor gases to maintain a constant pressure, typically on the order of a few Torr. In this regime, the deposition rate is primarily governed by the pressure and temperature and not by the gas flow characteristics. Because maintaining pressure and temperature is relatively easy, LPCVD films tend to be very uniform and very conformal. LPCVD is used for creating films such as silicon nitride, silicon dioxide, silicon carbide, and several germanium compounds. The required substrate temperature is governed by the reaction chemistry and is typically above 700 $^{\circ}\text{C}$.

5.2.2.2 Plasma-enhanced CVD

PECVD is a variant of LPCVD in which a plasma is used to reduce the substrate temperature to less than 300 $^{\circ}\text{C}$. This was developed to meet the needs of the complementary MOS (CMOS) manufacturing process in which high-quality dielectrics were required as insulation layers between the metal interconnect traces, but the LPCVD temperature was too high for integrated circuits in their later stages of manufacture. In PECVD reactors, the plasma is in close proximity to the substrate and is typically

at very low discharge power levels such that gas-phase reactions do not occur. The chemistries are very similar to LPCVD, except for the lower substrate temperatures.

5.2.2.3 Atomic layer deposition

ALD is a CVD technique that has recently become popular because of its highly conformal pinhole-free coverage and its ability to be run under a wide range of pressures and temperatures. An example of a conformal deposition is shown in Figure 5.6. The highly conformal nature allows one to achieve a uniform coverage around high-aspect-ratio structures and in deep holes. Unlike other CVD techniques, the gas precursors in ALD are flowed one at a time, not simultaneously. The gases are switched on and off with inert purge gas flows in between. Each gas flow duration is referred to as a pulse. During the first precursor pulse, the reactant molecules are allowed to adsorb onto the substrate surface. The extent of adsorption will be a function of temperature, pressure, and time. The chamber is then purged and the second precursor pulse is applied. The second precursor will react with the adsorbed gases from the first precursor to produce the desired film. Because the gases mix only at the surface, it is inherently a surface reaction. It is also highly conformal and can be used to deposit a film over high-aspect-ratio structures. Furthermore, the reaction is self-limiting because once all of the adsorbed molecules from the first precursor have been consumed, the reaction comes to a stop. The film growth proceeds by repeating the above cycle many times. During each cycle, the growth is typically on the order of a monolayer or even smaller. By counting the number of cycles, the layer thickness can be very accurately controlled. The average growth rate in ALD is only a few Angstroms per cycle, so the overall growth rate is slow. This is one of its biggest disadvantages.

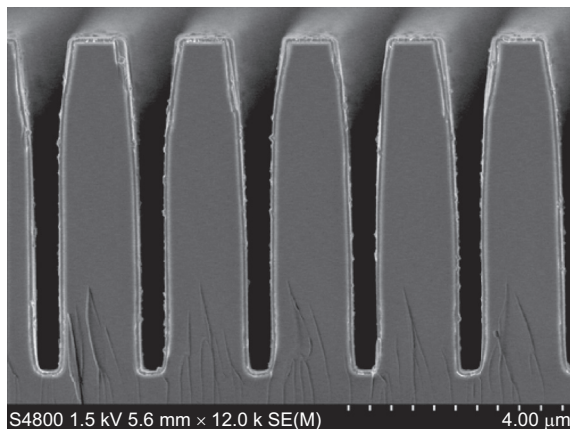


Figure 5.6 Conformal film by ALD deposition.

Reprinted with permission from J. Am. Chem. Soc., 2009, 131 (10), pp. 3478–3480. Copyright 2009 American Chemical Society.

ALD chemistries have been developed for several different oxides, nitrides, and metals, and the list of chemistries and applications continues to grow. Because both precursors are never flowed simultaneously, gas-phase reaction will not occur regardless of pressure or temperature. However, the pressure and temperature will affect the adsorption rate and hence the film growth rate during each cycle. This flexibility allows one to develop a deposition process for virtually any substrate, including polymer substrates. ALD is currently most actively being developed for CMOS gate dielectrics, such as aluminum oxide (Al_2O_3) and hafnium oxide (HfO_2), because of ALD's ability to produce defect-free ultrathin films.

5.2.3 Epitaxy

All of the aforementioned thin film growth techniques produce amorphous films. Amorphous means that the atoms in the film are randomly packed with no specific order. These are the simplest kind of films to produce. On the other hand, a crystalline film has a long range order and can be grown only under very specific conditions. Most materials exist in amorphous, crystalline, or polycrystalline states. Their mechanical, optical, thermal, and electrical properties will also change depending on their state. For example, carbon in its amorphous state is a black powder, whereas in its crystalline state it is diamond—an optically clear solid with a high refractive index. Silica in its amorphous form is a white powder and in its crystalline form is quartz.

Because crystalline films are more difficult to produce, amorphous films are used whenever their properties are sufficient for the application. Such is the case with optical coatings—the refractive index of most films can be made with excellent reliability and repeatability even in their random amorphous state. Metal films are also used in amorphous form because their electrical conductivity and optical reflectivity can be reliably reproduced. The major difference between amorphous and crystalline materials appears in their electronic band structure. Amorphous materials do not have a distinct band structure because of their random orientations; crystal structures do. Certain applications require well-defined electronic band structures for their operation. Examples are semiconductor electronics, optoelectronics, and components that require birefringence or piezoelectric properties. In some limited cases, it is possible to crystallize an amorphous film by annealing it at high temperatures. However, epitaxy is the formal process of creating high-quality crystalline films. Epitaxy can be chemical or physical, just like their amorphous counterparts, but we discuss epitaxy as a separate topic because of its distinct differences compared with other types of film growth.

A crystalline film can only be grown on a substrate whose lattice structure is closely matched to the film being grown. If the substrate is identical to the film, then it is known as homoepitaxy. If the substrate is slightly different, but is still compatible, then it is called heteroepitaxy.

Let us examine how silicon crystalline wafers are grown, which is the cornerstone of all of today's electronics. One starts with amorphous or polysilicon (which can be a powder or chunks) reduced from silicon dioxide mined from sand. This silicon is melted into a liquid, and a seed crystal is inserted into the melt. This starts a layer-by-layer solidification process on the seed crystal that eventually grows in size.

This is a homoepitaxial liquid-phase growth known as the Czochralski process. There is another method called the float zone process that can produce higher purity silicon crystals. In this method, the polysilicon is made into a rod and is heated from the outside with a moving coil without contacting the silicon. Certain II–V compounds use another technique known as the Bridgeman method. In this method, the melt is slowly cooled within the container from one end with a seed crystal. All of these methods produce a cylindrical-shaped crystalline rod known as the boule. The crystal is then shaped, sliced into thin wafers, and individually polished.

For some applications, the purity and defect density of silicon grown from melts is not adequate because of the contamination from the crucibles and the mechanical finishing process. Therefore, a thin layer of crystalline silicon film is grown on top of these wafers using a vapor-phase epitaxy. This is a homoepitaxial vapor-phase growth. Vapor-phase epitaxy can produce higher purity films with a lower defect density. This requires appropriate precursor gases that contain silicon, such as silane. These silicon wafers are known as epitaxial silicon, which is widely used in high-performance electronics.

In addition to silicon, epitaxy is most extensively used in III–V semiconductors such as GaAs, InP, InAs, etc. Many III–V semiconductors exhibit an interesting property in which their band structures can be adjusted by including other elements, such as $\text{Ga}_x\text{Al}_{1-x}\text{As}$, without significantly changing their original crystal structure. This allows one to stack epitaxial layers with differing electronic band structures on top of each other. This is heteroepitaxy and has become commonplace in optoelectronics devices such as laser diodes, light-emitting diodes (LEDs), and quantum well devices.

5.2.3.1 Metal organic CVD

Metal organic CVD (MOCVD) is a CVD process for growing epitaxial films, very similar to LPCVD, and is done by flowing precursor gases over the substrate. In III–V semiconductors, the metallic element is carried by an organic gas such as trimethylgallium ($\text{Ga}(\text{CH}_3)_3$) and trimethylindium ($\text{In}(\text{CH}_3)_3$) along with arsine (AsH_3) or phosphine (PH_3). The gases are allowed to decompose due to pyrolysis on the heated substrate surfaces to produce the desired film. The process pressures are typically in the range of 10–100 torr, resulting in relatively fast growth rates. One drawback of MOCVD is the toxic and explosive nature of the precursor gases, which makes them difficult to use in small research laboratories. Nevertheless, MOCVD is a scalable process amenable to volume manufacturing because many substrates can be simultaneously placed in the chamber. As a result, it is widely used in the manufacture of quantum well lasers, LEDs, and other components.

5.2.3.2 Molecular beam epitaxy

Whereas MOCVD is similar to LPCVD, molecular beam epitaxy (MBE) can be considered similar to PVD evaporation and is performed in ultra-high vacuum. This makes MBE better suited for applications that require very high purity levels. Solid sources such as gallium or indium from different effusion cells are typically allowed

to sublime and condense on the substrate. The cells are shuttered to allow rapid and precise transition from one material to another. The high-vacuum environment also allows one to use a wide range of diagnostic tools during growth. Many MBE systems use reflection high-energy electron diffraction (RHEED) to monitor the progression of growth, with the ability to count monolayers as they grow. The source configurations also make this system a lot less hazardous than MOCVD. Chemical beam epitaxy is a variant of MBE in which gas sources are used instead of solid sources but the principles are very similar. The biggest drawback of MBE compared with MOCVD is the slow growth rate and the inability to grow many wafers at once. Nevertheless, it is much more widely used than MOCVD in research facilities to study the fundamental properties of epitaxial film growth and in some limited production environments.

Table 5.1 is a brief summary of each thin film growth technique discussed in this chapter.

5.3 Lithography

Lithography literally means printing artwork or text on a surface. In the context of nano- and microfabrication technology, lithography is used to apply a pattern on a substrate surface so that it can be subsequently transferred to the underlying substrate. The lithography material itself is used as a sacrificial film that is discarded after the pattern is transferred.

Optical lithography uses light to replicate a pattern from a master (photomask) onto the substrate. The process is very similar to traditional photographic reproduction from negative films. A photosensitive polymer (photoresist) is applied to the substrate and then exposed to light through the photomask. The exposed photoresist undergoes a chemical reaction that results in a change in solubility. This is subsequently used to dissolve parts of the photoresist, leaving a patterned photoresist. There are several different photolithography techniques, such as contact, projection, immersion, and interference as well as UV, deep-UV, EUV, and X-ray based methods.

Nonoptical lithography involves electron beams, ion beams, or mechanical forces to create the pattern on the resist film. These are e-beam lithography (EBL), focused-ion beam (FIB) lithography, and nanoimprint lithography (NIL). A brief review of these techniques and their limitations are described in the following sections.

5.3.1 Photolithography

5.3.1.1 Light sources

UV light sources are commonly used in photolithography. This is not only because short wavelengths lead to better image resolution, but also because of the widespread availability of photochemicals sensitive to UV light. The mercury vapor lamp still continues to be the dominant UV light source in photolithography with emission lines at 405 nm (h-line), 365 nm (i-line), and 254 nm. Of these, 365 nm (i-line) is used the most, and many photoresists have been developed for this spectral range. The demand

Table 5.1 Summary of thin film deposition techniques

	Substrate temperature	Deposition energy	Pressure	Step coverage	Defect density	Uniformity	Deposition rate	Commonly used materials	Common application
Evaporation	Wide range	Low	Vacuum or reactive gas	Highly directional	High	High	Fast	Most metals, single elements, and stable dielectrics, such as Au, Ag, Cu, Si, SiO ₂ , MgF ₂ , etc.	Optical and electrical films, other generic applications
Sputtering	Wide range	High	Moderate. Mostly argon, but it can also include reactive gases	Directional	Moderate	High	Fast	Same materials as evaporation, plus additional metals and dielectrics such as W, VO ₂ , etc.	Optical and electrical films, other generic applications
PLD	Wide range	High	Wide range	Directional	Moderate	Poor	Slow	Complex compounds such as YBCO, PZT, and ferroelectric materials	Currently mostly used for exploration
LPCVD	High	Surface reaction	Moderate	Conformal	Very low	High	Fast	Si ₃ N ₄ , SiO ₂	Masking and MEMS
PECVD	Moderate	Surface reaction	Moderate	Somewhat directional	Low	High	Fast	Si ₃ N ₄ , SiO ₂ , polySi	Electrical insulation, passivation, masking
ALD	Wide range	Surface reaction	Wide range	Highly conformal	Very low	High	Slow	Al ₂ O ₃ , HfO ₂ , SiO ₂ , and certain metals	Gate dielectrics, passivation
MOCVD	High	Surface reaction	Moderate	Epitaxial	Low	High	Moderate	Compound semiconductors—GaAs, InP, AlGaAs	Manufacture of optoelectronic devices
MBE	Wide range	Surface reaction	Vacuum	Epitaxial	Very low	High	Slow	Compound semiconductors—GaAs, InP, AlGaAs	Research and development in epitaxy and optoelectronics

for higher and higher resolution has led to the utilization of deep-UV light sources, such as excimer lasers at 248 and 193 nm. EUV sources are currently being developed at 13.5 nm. In lithography applications, the illumination intensity has to be uniform across the entire substrate surface. Gaussian beams are not acceptable and speckle patterns from laser sources need to be eliminated. Therefore, significant effort is spent in shaping the beam into a flat, uniform profile.

5.3.1.2 Photoresists

Photoresists are light-sensitive, organic polymers in a solvent and are most commonly applied to the substrate by spin coating. Photoresists consist of a photoactive compound, a resin, and a solvent. The purpose of the solvent is simply to allow the photoresist to be spin coated. After the spin coating step, the solvents are removed by heating the photoresist. The resin is the structural component of the photoresist that is used for subsequent pattern transfer. In i-line (365 nm) photoresists, the most commonly used resin is the novolac resin (belongs to the phenolic family). The photoactive compound is diazonaphthoquinone (DNQ). Upon exposure to UV light, the DNQ releases a photoacid that increases the solubility of the resin. This type of photoresist is also referred to as a positive-tone photoresist because the exposed areas will eventually be removed and the unexposed areas will remain. The dynamic range of the solubility can be greater than three orders of magnitude and highly nonlinear, which is what gives photolithography its excellent contrast.

The exposure is commonly measured in the units of millijoules per square centimeter (mJ/cm^2). This is the illumination intensity in mW/cm^2 multiplied by the exposure time. Typical dose values range from 50 to $500 \text{ mJ}/\text{cm}^2$.

Although the above description is for the most common type of photoresists, there are other photoresists based on different chemistries and mechanisms. One common variant is the negative-acting photoresist. These photoresists have the opposite behavior, in which the solubility of the photoresist is reduced after exposure to light. This occurs because of polymerization reactions that increase the molecular weight of the resin.

In the deep-UV wavelengths, the photosensitivity becomes greatly reduced because of absorption by the resin. As a result, a mechanism known as chemical amplification is utilized, in which a single photoacid molecule can act as a catalyst to create many reactions. Photoresists in the EUV range are far less common and are still under development.

5.3.1.3 Photomasks

Photomasks are transparent glass substrates on which metal patterns are created to block the transmission of light. This metal is typically chromium because it has excellent adhesion to glass and is very opaque to UV wavelengths. Standard photolithography processes are used to create the photomask, including the application of photoresists, exposure, and pattern transfer to the underlying chromium layer. However, for the exposure, instead of a photomask, scanning tools such as a laser

scanner or an electron-beam scanner are used. A UV laser source such as HeCd or ArF is used to raster scan the laser spot across the entire surface while turning the laser beam on or off with a shutter driven by software that contains the photomask design. Assuming a HeCd laser at a wavelength of 325 nm, the smallest spot size that can be achieved is on the order of 300 nm. A binary feature such as a line may require multiple widths of this spot. As a result, the generally advertised limit for laser-written photomasks is approximately 1 μm . Electron-beam writing is used when smaller features are required.

5.3.1.4 Contact photolithography

Contact photolithography is the simplest to describe and the most widely used method in research laboratories. The photomask is placed in physical contact with the photoresist-coated substrate and exposed to UV light, typically with a 365-nm mercury lamp, as illustrated in Figure 5.7. The extent of the contact is important because it determines the resolution of the features that can be imaged. The term critical dimension (CD) is often used in photolithography to refer to the width of the smallest line that can be printed. An unintentional small gap can cause the exposed patterns to diffract and become enlarged. Even in the ideal case of a perfect contact with a zero gap, there will still be diffraction through the thickness of the photoresist. The aerial image can be calculated, but because this is a near-field phenomenon it would require a numerical evaluation of the Fresnel integrals. However, if we can approximate the light exiting a photomask opening as a Gaussian shape, it becomes possible to derive a rough analytical formula. Using this approximation, the following formula shows the width of the Gaussian aerial image as a function of the photoresist thickness z , air gap s , and UV wavelength λ :

$$W_{\min} \approx \frac{3}{2} \sqrt{\lambda \left(s + \frac{z}{2} \right)}. \quad (5.1)$$

At a wavelength of 365 nm, with a zero gap and a photoresist thickness of 500 nm, the smallest feature we can print with contact lithography is approximately 450 nm. This is the best-case scenario, and any contamination on the substrate or photoresist will introduce additional gaps between the mask and photoresist and will degrade

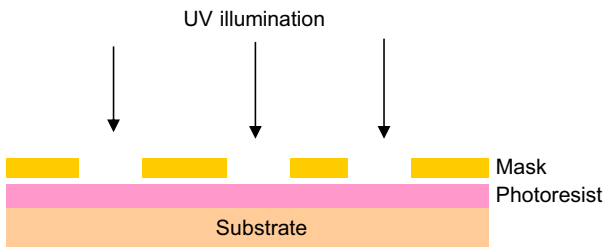


Figure 5.7 Contact photolithography.

the resolution. A dust particle 1 μm in size will create a 1- μm air gap. Using the above formula, we can verify that W_{\min} will enlarge to 1 μm . This is one reason why contact photolithography must be performed in an ultra-clean particle-free environment.

One can reduce the photoresist thickness to improve resolution, but this results in a compromise. A photoresist that is too thin will not withstand the subsequent pattern transfer steps. In addition, the scaling of resolution with photoresist thickness is not linear—a 50% reduction in thickness will only produce a 30% improvement in resolution. The same is also true with the source wavelength. Therefore, the often claimed resolution limit of contact photolithography is about 500 nm.

The biggest advantage of contact photolithography is the simplicity of the equipment. The optical components are fairly simple and are used primarily for creating a uniform illumination profile. As a result, most research and development facilities rely primarily on contact photolithography. If the required resolution is beyond what can be achieved with this method, then EBL or FIB lithography can be used.

5.3.1.5 Projection photolithography

In projection systems, the photomask is held at a distance from the substrate and an optical imaging system is used to project the image of the photomask onto the substrate surface. The system has many similarities to an optical microscope, and even the resolution conditions are the same. A simplified projection system is shown in [Figure 5.8](#).

The resolution limit using Abbe's criteria is

$$R = \frac{\lambda}{2\text{NA}}, \quad (5.2)$$

where λ is the UV wavelength and NA is the numerical aperture. This equation can be derived by considering a cone of incident waves up to a maximum angle of $\sin \theta$. All of these waves combine together to produce the aerial image. The features with the highest resolution will arise from the waves that have the largest wave vectors in the plane of the image. If k is the free space wave vector, then the largest wave vector parallel to the image plane will be $nk \sin \theta$, as illustrated in [Figure 5.9](#).

Two of these counterpropagating waves will produce a standing wave with a wave vector of $2nk \sin \theta$. This will contain the largest spatial frequency of the image. If we convert this to spatial period, then we can get $R = \frac{\lambda}{2n \sin \theta}$. If we substitute $n \sin \theta = \text{NA}$, we can easily obtain Abbe's equation. R is the smallest distance between two bright fringes in the image and is defined as the resolution of the image. In lithography, this quantity is called the pitch. If we assume the lines and spaces have equal widths, then the width of a single line will be half of this value. The half-pitch is typically defined as the CD of the lithography system.

Therefore, the half-pitch, HP, can be written as

$$\text{HP} = \frac{\lambda}{4\text{NA}}. \quad (5.3)$$

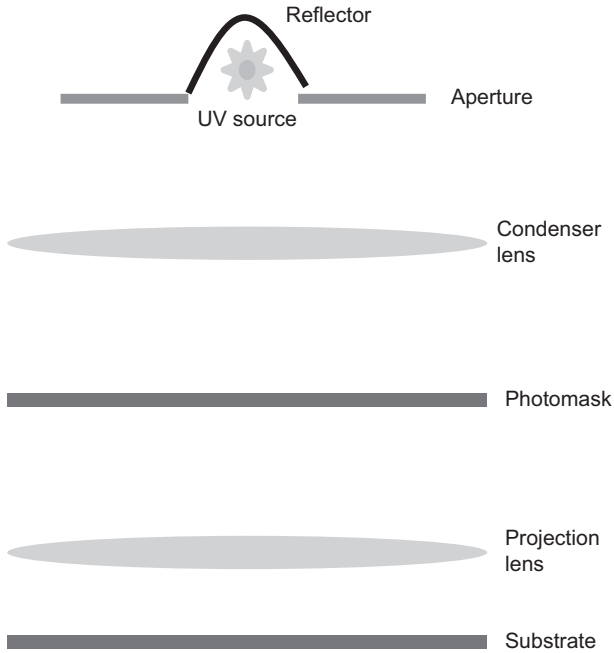


Figure 5.8 Elements of a projection photolithography system.

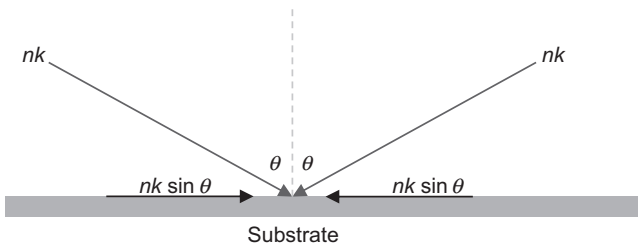


Figure 5.9 Numerical aperture and maximum spatial resolution.

In general, we lump the $1/4$ and several other optical system and photoresist factors into a single parameter k_1 and write the equation as

$$\text{HP} = k_1 \frac{\lambda}{\text{NA}}. \quad (5.4)$$

In older projection lithography systems, NA was lower than 1.0 and k_1 was higher than 0.25. For example, using a mercury lamp at 365 nm, if $\text{NA} = 0.5$ and $k_1 = 0.5$, we get a half-pitch of 365 nm. This was a fairly typical result, and the wavelength was often loosely treated as being equal to the half-pitch. Although this was not significantly better than contact lithography, most manufacturing systems used projection lithography because of their increased throughput and reliability. Because the mask

is never brought in contact with the substrate, the mask can be kept clean and the substrates can be easily switched. Furthermore, projection systems also allow one to demagnify the image. For example, using 5:1 reduction optics, 500-nm features can be 2.5 μm on the mask. This relaxes the feature sizes required on the photomask and the cost of producing photomasks.

The depth of focus is the vertical distance above and below the image plane where the smallest features remain in focus. This is

$$\text{DOF} = k_2 \frac{\lambda/n}{\sin^2 \theta}. \quad (5.5)$$

For a 365-nm illumination with $\text{NA} = 0.5$ (i.e., $\sin \theta = 0.5$) and $k_2 = 0.5$, we can get the $\text{DOF} = 730$ nm. This means the substrate and photomask have to be held perfectly parallel within 730 nm, and the photoresist film has to be perfectly flat with no topographic features exceeding this DOF value. However, even in the best wafers, minor substrate curvatures can easily exceed this DOF. This is one reason why an entire wafer is not exposed with a single image in a projection system. It becomes nearly impossible to satisfy the DOF when the image size is larger than approximately 1 in. Instead, projection lithography systems use a step-and-repeat configuration. The reduced image of the photomask is projected and exposed, then the wafer is translated and the same pattern is exposed over and over again to fill the entire wafer. Hence, these systems are also known as steppers. This is illustrated in [Figure 5.10](#). The maximum size of a single exposure is defined as the field size and is determined by the substrate and mask flatness as well as the quality of the optics and spherical aberrations in the system compared with its DOF. For example, if a system can only maintain the DOF specifications within a 1- by 1-in area, and if the reduction optics are 5:1, then a 5- by 5-in photomask will be printed as a 1- by 1-in field on the substrate, which can be repeated several times to cover the entire substrate.

Although device dimensions have been shrinking, the demand for larger and larger chips have been driving the field sizes upward. This requires larger imaging

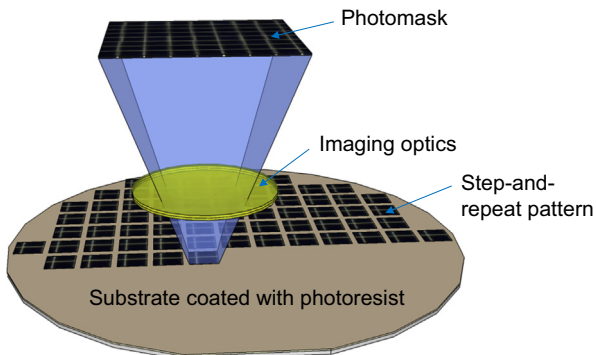


Figure 5.10 Step-and-repeat configuration.

lenses; however, this leads to increased aberrations. One way to overcome this problem is by projecting only a small portion of the image through a slit aperture using a smaller lens. The mask and the wafer are then linearly scanned synchronously to cover the entire aerial image. This is known as the step-and-scan system, and it is currently the most widely used commercial system.

To improve resolution, projection systems have moved toward shorter and shorter wavelengths, from 254 nm (KrF) and then to 193 nm (ArF) excimer lasers in the deep-UV range. In addition, the k_1 parameter has also significantly improved because of better optics with lower aberrations and photoresist performance and is currently very close to the theoretical limit of 0.25. The numerical aperture has also increased over the years and is now at 1.35. If it seems odd that the numerical aperture is higher than 1.0, it's because this is the effective numerical aperture by replacing air with water as the imaging medium. This is known as immersion lithography. In microscopy, this is a commonly used method in which high-refractive-index oils are used to increase the numerical aperture. However, many oils are incompatible with photoresists, and many are not transparent in the deep-UV wavelengths; therefore, water has become the preferred medium for immersion lithography. The refractive index of water at 193 nm is 1.44. Putting all of these together, the HP becomes 35 nm and the DOF becomes 38 nm. This is a very small number, and it is difficult to maintain the parallelism between the substrate and photomask over more than 1 in. Even then, elaborate methods such as laser interferometric methods are used to measure the substrate-to-photomask distance and make fine corrections in real time.

Thirty-five nanometers was once considered the limit of the 193-nm immersion lithographic systems (abbreviated as "193i"). In recent years, techniques such as phase-shifted masks, optical proximity corrections, off-axis illumination, double patterning, dual tone resists, self-aligned double-patterning, etc. have been successfully developed to push the limit of this resolution even further. Currently, features as small as 20 nm are being produced by major integrated circuit manufacturers using the 193-nm laser source. This is an impressive $\frac{\lambda}{10}$ resolution that is being mass produced in consumer electronics without circumventing the effects of diffraction. Some of these resolution enhancement techniques are briefly described here.

Double patterning

Double patterning is based on the concept that exposed lines and spaces do not have to be of equal widths. Assuming a positive-tone photoresist, a higher dose will result in narrower photoresist lines and wider spaces. Because the aerial image formed on the photoresist surface will always contain diffused edges due to diffraction, increasing the dose will enlarge the area that will receive the threshold dose required for dissolution. As a result, it is possible to print lines that are smaller than half-pitch. If the line width is 25% of the pitch, then that would leave 75% of the empty space to be used for printing another line, effectively doubling the density of lines. The half-pitch then becomes 17.5 nm.

In practice, this is done with two layers of hard masks (typically dielectrics) under the photoresist. The first set of lines are exposed, developed, and etched into the top hard mask layer. A new coating of photoresist is then applied, and the second mask

is carefully aligned to the previously created features and exposed. The second photoresist pattern combined with the previously patterned hard mask are then used to etch into the bottom hard mask layer. The top hard mask can then be removed leaving the bottom hard mask with the desired high-density pattern. This process is referred to as litho-etch-litho-etch. To simplify this process and to increase the reliability, another process known as litho-freeze-litho-etch was developed. In this process, only one hard mask layer is used. After the first photoresist pattern is developed, it is chemically treated to prevent it from dissolving during the second application of the photoresist. The second photoresist is then applied, aligned, and patterned. Finally, a single etch step is performed to transfer the high-density pattern into the underlying hard mask.

Dual tone photoresists

Positive-tone photoresists have low solubility when unexposed and become highly soluble when exposed beyond their threshold dose. Dual tone photoresists are designed to have low solubility when the dose is too low and also when the dose is too high. This allows the photoresist to dissolve only when the dose falls within a certain band in the middle. When exposed with an array of diffused lines and spaces, the resulting photoresist pattern will produce two lines for each line in the aerial image, effectively doubling the density of lines.

Self-aligned double-patterning

Self-aligned double-patterning (SADP) is widely used in today's manufacturing systems. The process uses the edges of each line to create a new line. Because each line contains two edges, it effectively doubles the line density. Furthermore, this process does not require a second lithography step or alignment. Hence, it is referred to as a self-alignment process.

First, the photoresist is exposed and developed with the initial features. These features are then isotropically coated with a film. The film is then directionally etched so that it is removed from all horizontal surfaces. This leaves the films on the sidewalls of each feature. The density of these features will be twice that of the original line density. This is illustrated in [Figure 5.11](#). The isotropic coating and directional etch can be repeated for a second time, doubling the line density again. This will result in four times the original density. This is referred to as self-aligned quadruple patterning. Yet another iteration will result in self-aligned octuplet patterning.

Because of its simplicity and self-alignment capability, SADP has allowed the manufacturers to use their existing 193 immersion systems to push the limits further without significant retooling.

5.3.1.6 EUV lithography

F₂ lasers emitting at 157 nm were expected to supersede the ArF 193-nm lasers in projection lithography. That did not happen because of several technical hurdles, one of which was the requirement for CaF₂ optics because silica is not transparent at this wavelength. In addition to the high cost of CaF₂ optics, special design considerations were required to overcome its birefringence. As a result, the 157-nm method

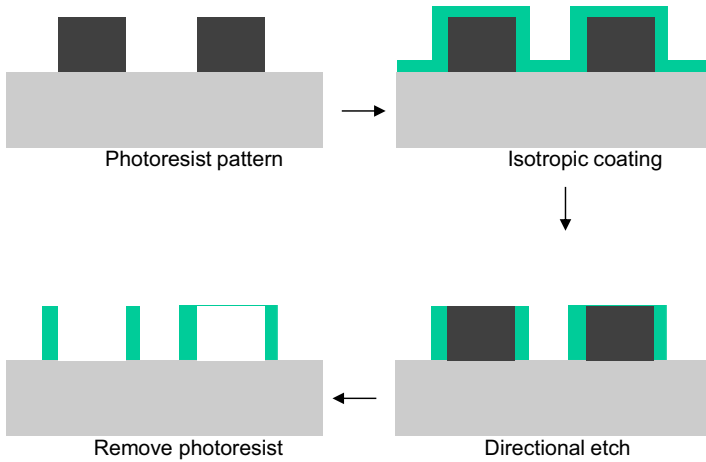


Figure 5.11 Illustration of the SADP process.

never became a mainstream lithography method, and the industry continued to look for sources in the EUV range. One of these is the 13.5-nm EUV method. This is not just a shorter wavelength, but it is also a dramatic shift in all aspects of lithography because the wavelength is almost in the regime of X rays. The whole optical train has to be in vacuum. High-power light sources with efficiency and uniformity are required. New photoresists and new approaches for making photomasks had to be developed. Because of the difficulty in making refractory optics in EUV, an all-reflective approach was developed using multilayer Bragg mirrors. The photomasks are also reflective instead of transparent. This is currently an active area of research, and many tool manufacturers are heavily investing in the EUV method. For more information on EUV lithography, the reader is referred to the special sections on EUV in the SPIE JM3 journal [9,10].

5.3.1.7 Laser interference lithography

Laser interference lithography is a maskless projection method to create periodic lines of exposure with line widths as small as $\frac{\lambda}{4n}$. A laser beam is split in two and recombined at the substrate surface to create a periodic array of bright and dark lines (see Figure 5.12). If two coherent beams of wavelength λ are incident at an angle θ from opposite sides of the surface normal, then Figure 5.9 can be used to show that

$$\Lambda = \frac{\lambda}{2n \sin \theta}, \quad (5.6)$$

where Λ is the pitch and n is the refractive index of the incident medium. The smallest pitch will occur when the incident angle is 90° . Using a 266-nm laser, the smallest pitch will be 133 nm with a line width of 66 nm (assuming a 50% duty cycle). In practice, it is not possible to illuminate the substrate at 90° incidence. The maximum

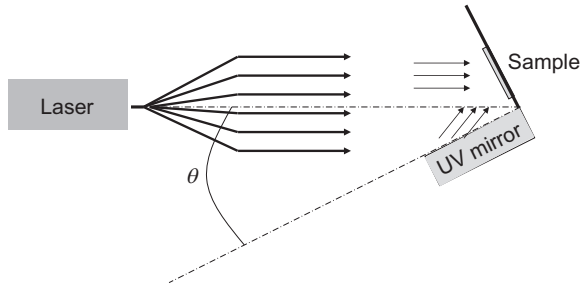


Figure 5.12 Laser interference lithography using a Lloyd's mirror setup.

angle is likely to be close to 60° . This would result in a pitch of 153 nm and a line width of 76 nm. It is also possible to perform two consecutive exposures after rotating the substrate in plane by 90° to achieve a two-dimensional grid pattern.

One commonly used configuration for interference lithography is the Lloyd's mirror setup, which contains two planes at right angles to each other. A mirror is installed on one plane, the substrate is placed on the other plane, and the laser beam is expanded and collimated to cover both planes. From simple geometry, we can show that the angles of incidence of each beam on the substrate will be equal, which creates the interference pattern. Furthermore, the entire fixture can be rotated to change the angle of incidence, which gives it the flexibility to easily change the pitch of the periodic fringes.

Interference lithography is used to make diffraction gratings and photonic crystals. Its advantage is that it is a maskless method and requires a fairly simple setup to achieve line widths that rival even the most advanced projection lithography system. It can cover a large substrate area, much larger than what is currently possible with mask projection systems. It also does not suffer from any depth-of-field issues. Compared with the tens of millions of dollars it costs to acquire and set up a 193-nm projection lithography tool, an interference lithography setup can be built for far less. Its biggest disadvantage is that it can only make periodic structures. Nevertheless, interference lithography is still a powerful tool that is used anytime large-area periodic nanostructures are required.

Figure 5.13 shows two typical examples of one- and two-dimensional periodically patterned photoresist films made by interference lithography.

5.3.2 Nonoptical lithography

Although optical lithography is the dominant technique, there are other techniques that do not use light to print a pattern on a resist layer. We will use the term resist rather than photoresist to indicate the lack of photosensitivity in this application.

5.3.2.1 Electron-beam lithography

Of all of the nonoptical lithographic methods, this is the most commonly used method. It uses a beam of electrons rather than photons to expose the resist to induce a chemical

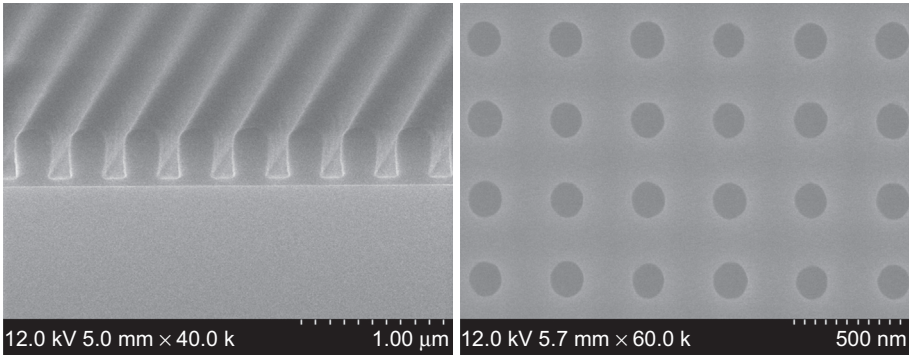


Figure 5.13 Scanning electron microscope images of photoresist features made using laser interference lithography.

change, which subsequently results in a change in solubility. However, unlike photolithography, there is no mask and the beam does not illuminate the entire substrate. In that sense, EBL does not have any similarities with contact lithography or projection lithography. Instead, the electron beam is generated, accelerated, and focused down to a small spot and scanned across the substrate to create the image. The scanning mechanism is done by a combination of mechanical translation of the substrate stage as well as tilts in the electron beam. The pattern is generated by modulating the beam current on and off as the beam is scanned. The equipment is similar to a scanning electron microscope, where the e-beam, steering coils, and the substrate are all housed in a high-vacuum chamber. The resolution limit follows the same relation as with photons:

$$HP = k_1 \frac{\lambda}{NA}. \quad (5.7)$$

The wavelength of electrons can be computed from de Broglie's principle, which states $\lambda = \frac{h}{p}$, where h is Plank's constant and p is the momentum. If the kinetic energy of the electron is E , then we can use $E = \frac{1}{2}mv^2$ and $p = mv$ to obtain $p = \sqrt{2mE}$ where m is the mass of the electron. This allows us to express the wavelength in terms of kinetic energy as

$$\lambda = \frac{h}{\sqrt{2mE}}. \quad (5.8)$$

If the electron is accelerated by 10 kV, then it will acquire a total energy of 10 keV and the wavelength can be calculated to be approximately 12 pm. This is many orders of magnitude smaller than UV wavelengths and is the main attraction of electron beams. It has the potential for an extremely small resolution limit, although much of it is still unrealized as of today.

The numerical apertures of electron-beam systems are typically very low, on the order of 0.01. This has the benefit of an extremely large depth of focus. Despite the

low numerical aperture, the expected resolution should still be on the order of 1 nm. However, in practice, this is not the case. The k_1 parameter in electron-beam systems is much larger than in optical systems. As we saw earlier, the smallest theoretically possible value for k_1 is 0.25, which with 10-keV electrons would lead to a subnanometer resolution. However, in current EBL systems, the value of k_1 is on the order of 5. This is due to the poor spherical and chromatic aberrations in the magnetic focusing system as well as the interaction of the electron beam with the substrate that emits secondary electrons that cause blurring. As a result, the practical resolution of EBL is only in the range of 5 nm.

The biggest advantage of EBL is its high resolution. The depth of focus is also very large because of the small numerical aperture. This makes it a very forgiving system for performing lithography on substrates with topographic features and small curvatures. Currently EBL is widely used in the manufacture of photolithography masks, especially when the required resolution on the mask is beyond what can be achieved with a laser scanner. This is by far the largest commercial application of EBL today. In research and development, EBL is used whenever the required features are smaller than approximately 500 nm, by directly writing the pattern on the resist layer, bypassing a mask. Although deep-UV projection with immersion can easily reach below 100 nm, such tools are generally not available outside of large production environments. Hence, for research and development, EBL becomes the tool of choice where contact photolithography leaves off. The biggest disadvantage of EBL is its slow speed. Because it is a scanning system, it is inherently slow. This speed is inversely related to the resolution—as the beam spot size is reduced to achieve a higher resolution, its speed will also decline. To write a 1- by 1-in area could take several hours depending on the density of the pattern and the required dose. Another drawback is the requirement for the substrate to be conductive. As in the scanning electron microscope, the beam current needs a path to ground through the substrate to maintain charge neutrality. Any localized charging effects can significantly diminish the resolution. For photomask writing, although the substrate is glass, a metal film is deposited first before coating the resist film, and this metal film is electrically grounded. Writing on a resist film on a silicon substrate is also possible because silicon is partially conductive. Other cases of purely insulating systems will require careful consideration of how to dissipate the charge accumulation.

The dose in EBL is the charge per unit area, commonly expressed in microcoulombs per square centimeter ($\mu\text{C}/\text{cm}^2$). The beam current in combination with the scan speed affects the dose that is deposited on the substrate. Because long write times are the main disadvantage of EBL, optimization algorithms are used to minimize the write times. The current and beam size are dynamically adjusted based on the resolution of the local feature being written. They are increased when writing large features and reduced for finer features.

The resist mechanism is different from UV photoresists. In positive-tone resists, a process known as chain scission decreases the molecular weight of the exposed areas and increases the solubility. In negative-tone resists, a cross-linking process increases the molecular weight and reduces the solubility.

5.3.2.2 FIB milling

This is a maskless, resistless scanning lithography technique. It is very similar to EBL except it uses an ion beam instead of an electron beam to write the pattern. In addition, the beam is not used for exposing a resist; instead, the beam directly sputters and removes the substrate (or thin film), hence bypassing the need for a sacrificial resist layer.

Liquid metals are used for the ion source because of their heavier atomic weights, which will result in a higher sputter yield. Of these, gallium is the most common because of its low melting temperature. The gallium atoms are heated, ionized, accelerated, and focused to a small spot on the order of 5 nm on the substrate stage. The acceleration energy of the incident ions and substrate type will determine the sputter yield and hence the removal rate of atoms.

Redeposition of the sputtered atoms is a major consideration in FIB. This is when the sputtered atoms land back in the vicinity of the milled site, causing surface roughness. For this reason, a gas injector is used to volatilize the sputtered species and accelerate the removal rate of atoms. With suitable precursor gases, it is also possible to induce a CVD at the ion bombardment site. These different operational modes—sputter removal, reactive etch, and deposition—make FIB a very versatile tool. However, similar to EBL, its main disadvantage is its low speed. FIB is currently used to make repairs to photomasks by milling away unwanted metal traces and depositing missing metal traces as well as surgical repairs to integrated circuit chips.

Another very useful aspect of FIB is that it can be easily combined with a scanning electron microscope in the same vacuum chamber in a dual beam configuration. An electron gun and an ion gun can both be directed at the substrate, which allows one to direct the ion beam and monitor the milling process in real time. As in EBL, the substrate must be electrically grounded.

5.3.2.3 Nanoimprint lithography

NIL is very simple to describe. A mold (stamp) is first constructed with surface relief features on it and is then pressed against a polymer material to transfer the pattern. This is illustrated in Figure 5.14. The polymer material is typically a sacrificial resist film that is used as a mask to etch the underlying substrate or film. Interest in NIL has grown significantly because of its potential for nanoscale manufacturing.

There are several different NIL approaches. Thermocompression NIL uses elevated temperatures to soften the polymer film to allow it to flow around the surface reliefs of

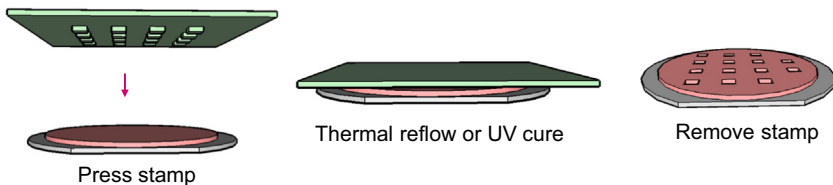


Figure 5.14 Nanoimprint lithography.

the stamp. The stamp is removed after allowing it to cool back to room temperature. This is also known as hot embossing lithography. Another variant uses UV light to cross-link and harden the polymer film. A soft liquid film is used, and during the imprint the film is exposed to UV light. This can also be used in combination with elevated temperature. UV-NIL requires the substrate and/or stamp to be transparent to UV wavelengths such as quartz.

The main advantage of NIL is the extreme simplicity of the process. There are no expensive optics, ultra-high-vacuum chambers, electron beams, or excimer lasers. Feature sizes also do not have a fundamental resolution limit. However, in practice, there are several challenges that have prevented NIL from being adopted on a larger scale. NIL requires physical contact between the stamp and the substrate. After several cycles, the stamp will have to be cleaned. In that sense, this is similar to contact photolithography, except NIL can transfer much finer features than contact photolithography. The flatness of the substrate and stamp are also critical factors. Any protrusions or topographic features will compromise uniform contact between the stamp and the film. However, this is not a significant disadvantage, at least when compared with deep-UV projection systems, which also have limited depth-of-focus problems.

NIL is currently used in many research and development laboratories, but it is still under development for large-scale manufacturing applications. The proposed NIL manufacturing systems use a step-and-repeat configuration just like in the UV steppers. Some of the manufacturing challenges are air entrapment between the stamp and the substrate, particulates, stamp contamination, and throughput.

[Table 5.2](#) summarizes some of the main advantages and disadvantages of the different lithography methods discussed in this section.

5.4 Pattern transfer

Once the resist film is patterned by lithography, chemical processes are used to transfer that pattern into the underlying substrate or film. This is the pattern transfer process. There are exceptions, such as FIB lithography, in which the pattern is directly etched without a resist, and the lift-off process, in which a patterned resist is used to lift an overlying film simply by dissolving the resist and not by etching the film. The etching processes are generally divided into wet-chemical etching and plasma etching.

5.4.1 *Wet-chemical etching*

This process requires only liquid chemicals; therefore, it is very simple and inexpensive to implement. For example, for etching a copper film, as is routinely done in the manufacture of printed circuit boards, the resist is applied and patterned over the copper film and then a chemical such as ammonium persulfate or ferric chloride is used to selectively etch the copper, leaving only the areas of copper that are protected under the resist film. The resist is then stripped off with a solvent.

Table 5.2 Summary of lithographic methods

Method	Wavelength	Approximate half-pitch ^a	Depth of focus	Common applications	
Contact lithography	365 nm (Hg)	500 nm	Infinite	R&D	
Projection lithography	365 nm (Hg)	350 nm		R&D and small production	
Projection lithography	193 nm (ArF)	75 nm		Production systems	
Projection immersion lithography	193 nm (ArF)	35 nm		Production systems	
Projection lithography with immersion and resolution enhancement	193 nm (ArF)	20 nm		Production systems	
EUV lithography	13.5 nm	5 nm		Still under development for production systems	
Laser interference lithography	325 nm (HeCd), 266 nm (YAG), 248 nm (KrF), 193 nm (ArF), etc.	100–500 nm		Periodic structures such as gratings	
EBL	0.01 nm (e-beam)	5 nm		Large	R&D laboratories and for making masks for production systems
FIB lithography	Gallium ions	10 nm		Large	Mask repair, chip repair, R&D
NIL	N/A	N/A		N/A	R&D with potential for commercial use

R&D, research and development.

^aHalf-pitch values given here are realistic values for these systems and not their theoretical limit.

Although simple, there are several limitations to this type of etch. First, the etch chemistry is directionally isotropic. This means the film is etched vertically and horizontally at nearly equal rates. The consequence of this is a curved sidewall that results in a larger opening at the top than the original opening in the resist. The excess lateral etch under the resist is known as the undercut.

The etch process can be described as a sequence of three steps: (1) the etchant diffuses from the bulk liquid to the surface being etched, (2) the surface reaction takes place, and (3) the by-products diffuse out to the bulk liquid. The rate of each step determines the overall etch rate. This is known as the rate-limiting step. The actual model is complex because it involves the diffusion of different species, all of which depend on their respective concentration gradients, but we can consider a simplified model as follows:

$$\frac{1}{R} = \frac{1}{R_i} + \frac{1}{R_r} + \frac{1}{R_o}, \quad (5.9)$$

where R_i is the diffusion rate of the fresh reactants to the reaction site, R_r is the etching reaction rate, and R_o is the diffusion rate of the by-products away from the site. The smallest of these factors determines the overall etch rate R . The surface reaction rate R_r is a function of the concentration of the reacting species and is exponentially dependent on temperature. This can be written as

$$R_r = k_0[A]^n[B]^m e^{-\frac{E_a}{kT}}, \quad (5.10)$$

where $[A]$ and $[B]$ are the concentrations of the reacting species (assuming there are only two species), E_a is the activation energy for the reaction, k is the Boltzmann constant, T is the temperature, n and m are constants, and k_0 is the rate constant. The diffusion rates R_i and R_o are functions of the geometry, such as mask opening and aspect ratio of the etch profile. The latter will be a dynamically varying parameter as the etch progresses.

Wet-chemical etching is mostly used for etching large geometries where the diffusion rates are large; therefore, the rate-limiting step is primarily the surface reaction rate. Therefore, the overall rate depends primarily on the concentration of the fluid and the temperature. In this regime, the etch profile will be ideally isotropic. The diagram in [Figure 5.15](#) illustrates this, where a thin film of thickness t is wet-etched through a mask opening of width w .

Using geometry, when the width of the etched trace is w at the bottom of the trench, the width at the top of the trench will be:

$$w' = w + 2t, \quad (5.11)$$

or expressed as a fraction of the original space width,

$$\frac{w'}{w} = 1 + 2\frac{t}{w}. \quad (5.12)$$

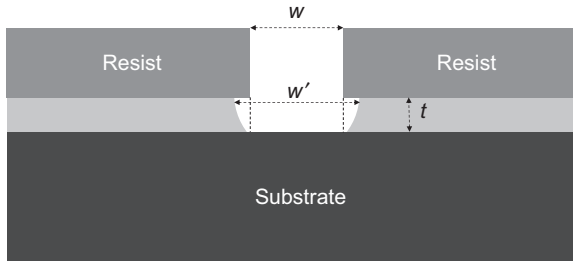


Figure 5.15 Illustration of wet etching characteristics.

Here are some numerical results to illustrate the line width:

- $t = 100 \text{ nm}, w = 500 \text{ nm} \rightarrow \frac{w'}{w} = 1.4$
- $t = 100 \text{ nm}, w = 250 \text{ nm} \rightarrow \frac{w'}{w} = 1.8$
- $t = 100 \text{ nm}, w = 100 \text{ nm} \rightarrow \frac{w'}{w} = 3.0$

Clearly, as the mask dimensions shrink toward and below the thickness of the film, we will not be able to etch the film without a significant loss of resolution. The film thickness is often limited by other design criteria such as the optical dielectric constant or electrical resistivity; therefore, it is not a parameter that can be easily changed. This is the biggest drawback of wet-chemical etching—the isotropic nature results in a significant enlargement of the traces as the dimensions get smaller. The advantage of wet-chemical etching is its low cost. Another significant advantage is the high selectivity. Selectivity is defined as the etch rate of the film (desired etch rate) divided by the etch rate of the resist (undesired etch rate):

$$S = \frac{R_{\text{film}}}{R_{\text{resist}}}. \quad (5.13)$$

Because the etch rate is entirely driven by chemical activity, it is possible to find chemicals that will etch the film and leave the resist virtually untouched. Most materials commonly used in electronics have well-established etchants with high selectivity with the resist. A comprehensive list of wet etchants for various materials can be found in the references listed at the end of this section. However, there are some cases in which the resist simply will not stand up to the etchant. One example is the etching of silicon with HNA solution (hydrofluoric, nitric and acetic acid mixture). This is an aggressive etchant, and most polymer-based resists will be removed very quickly. Therefore, the commonly used technique is to use the resist to pattern a more chemically resistant hard mask such as silicon nitride. Then, the patterned hard mask is used to etch the underlying silicon with the HNA solution.

It should also be noted that not all wet chemicals produce isotropic etches. In crystalline materials, the etch rate can be strongly dependent on the crystal orientation. The etch rate R_r will then have a strong dependence on the angle θ . Such etches are known as anisotropic wet-chemical etches. One example of this is the etching of

silicon with KOH (potassium hydroxide). When the surface orientation of silicon is $\langle 100 \rangle$, the etch rate drops to zero at 54° to the surface along certain planes. The end result will be an etch profile that has sidewalls sloping at 54° where the etch rates come to a stop. An example of this etch profile is shown in Figure 5.16. This type of etch is used for making components in which the etch needs to come to a stop at a sharp corner (such as in atomic force microscopy tips) and in micro electro mechanical systems (MEMS).

5.4.2 Plasma etching

Plasma etching is also referred to as dry etching because it is performed in a gas phase without the use of liquids. It is sometimes referred to as reactive ion etching (RIE), although the correct term should be ion-assisted chemical vapor etching. In plasma etching systems, the substrate is placed in a vacuum chamber on the cathode of the plasma generator and gases are introduced to produce the reaction. This is illustrated in Figure 5.17. One advantage of this system is that fairly safe gases can be fed into the chamber, which in a plasma state become dissociated into highly reactive species. An example of this is CF_4 . This gas is fairly inert under normal conditions, but in a plasma it can generate many F atoms (free radicals), which are highly reactive and spontaneously attack silicon to produce SiF_4 . Because SiF_4 is a gas, silicon will be readily turned to a gas in such a plasma reaction. In addition, ions in the plasma will bombard the cathode similar to sputtering. This action creates an additional source of energy that can accelerate the etch rate parallel to the ion trajectories. Because ions are directed at normal incidence to the cathode, this has the effect of accelerating the etch rate normal to the substrate. This will result in an etch profile that has minimal undercut and is strongly anisotropic, as illustrated in Figure 5.18.

An important requirement in plasma etching is for the by-products to be volatile; that is, they need to be able to evaporate away into the pumping system. Any nonvolatile by-products will remain on the substrate as a deposited thin film and may impede

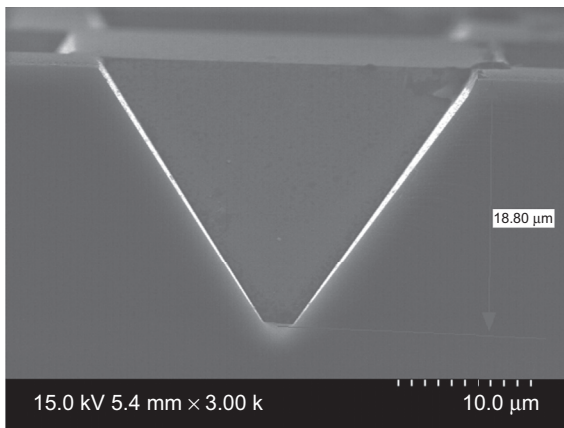


Figure 5.16 Profile of $\langle 100 \rangle$ silicon etched in KOH solution.

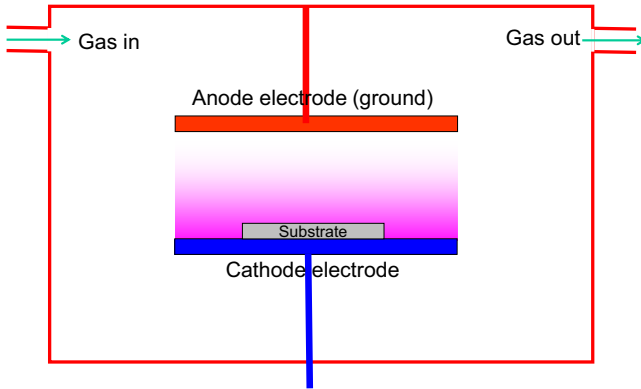


Figure 5.17 Parallel plate plasma etching configuration.

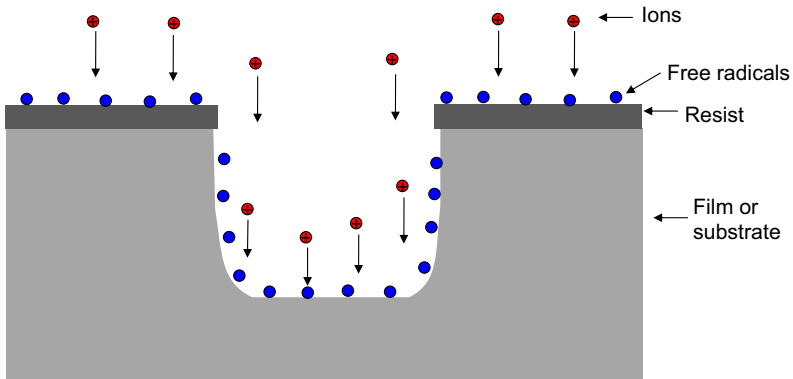


Figure 5.18 Interaction of the plasma with the substrate and mask.

the etching. Therefore, the gas chemistry has to be matched to the substrate being etched to produce the most volatile by-products. In addition to selecting the correct gas species, one can also control the substrate temperature, pressure, and plasma discharge powers to control the reaction and the by-products. The most common gases used in etching are fluorine-based or chlorine-based. Common fluorine-based gases are CF_4 , SF_6 , CHF_3 , C_4F_8 , etc. and chlorine-based gases are Cl_2 , BCl_3 , CCl_4 , etc. Most semiconductors and metals can be etched with these gases. Unlike wet-chemical etching, it is not a trivial task to switch gases to etch different materials because several systems have to be specifically installed for each gas type, including piping, flow controllers, and exhausts.

Plasma etching is widely used for etching nanoscale features because it can produce vertical sidewall profiles with little or no undercut. An example is shown in [Figure 5.19](#). This anisotropic feature can be further enhanced by allowing the deposition of certain passivating films to occur along vertical sidewalls during the plasma reaction. Along horizontal surfaces, these films will be removed by ion bombardment,

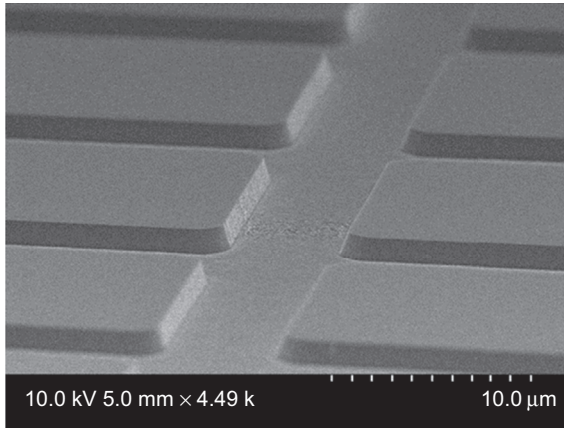


Figure 5.19 Example of a plasma etched substrate.

but they will remain on the vertical walls. By coating the sidewalls with these films, it is possible to produce very deep high-aspect-ratio features. Such etch processes are known as deep-RIE. Another aspect that makes plasma etching suitable for small features is the ability of gas species to diffuse into smaller features better than liquids. Plasma etching is also highly repeatable compared to a liquid bath.

The disadvantage of plasma etching, in addition to the higher cost of the system compared with wet-chemical etching, is the reduced selectivity. This arises because of the ion bombardment that can lead to sputter removal of the resist. Therefore, the resist gets consumed faster than in wet-chemical etching. As a result, ion-driven etches will have a poorer selectivity than free-radical-driven etches. Many dielectrics such as SiO_2 need ion bombardment to break bonds and make the molecules reactive. Without ion bombardment, these films will not etch. This will naturally lead to a lower selectivity. On the other hand, materials such as Si and certain metals can be etched by free-radicals alone, without the assistance of ions. These will have a higher selectivity. The resist thickness and selectivity determines the maximum etch depth that can be achieved. For example, if the resist thickness is 100 nm and the selectivity is 2, then the maximum etch depth that can be achieved will be 200 nm.

In a capacitively coupled plasma chamber, as shown in [Figure 5.17](#), it is not possible to independently control the ion density and the free radical density. Both of these quantities are coupled. The only variable that can be controlled is the plasma discharge power, which will simultaneously increase both the ion density and the free radical density. When etching different materials, some only require ion bombardment, others only require free radicals, and some require both. Therefore, it is useful to have independent control of these parameters. This is accomplished through a capacitively coupled and inductively coupled plasma. In this system, the capacitively coupled part controls the ion density and free radical density, and the inductively coupled part controls only the free radicals. This is a very common configuration used in advanced plasma etching systems today.

Problems

- Compare the following metals in terms of the temperature required to reach a vapor pressure of 1 torr and whether it will be a liquid or solid at this temperature: (a) titanium, (b) tungsten, (c) gold, (d) indium (use reference [2]).
- Consider the sputter deposition of gold using an argon plasma.
 - If the ion acceleration is 500 V and the ion current is 500 mA uniformly distributed over a 1-in square target area, then calculate the rate of atoms sputtered from the target, S . You need to look up the sputter yield of gold.
($S = 6.2 \times 10^{18}$ atoms/s)
 - The substrate is held at a distance of $D = 6$ in from the target. Approximating the target as a point source, and assuming the sputtered atoms have an angular distribution of $\cos \theta$ where θ is the angle from the normal to the target surface, show that the deposition rate of the substrate along $\theta = 0$ is $R = \frac{S}{\pi D^2}$. Calculate this value.
($R = 8.6 \times 10^{15}$ atoms/cm²/s)
 - Using the above value, calculate the film growth rate on the substrate at $\theta = 0$. You need to use the atomic weight, density, and Avogadro's number.
(14.5 Å/s)
- In this problem, consider a PLD system using a 157-nm F₂ laser.
 - If F_0 is the laser fluence (energy per pulse per unit area), α is the absorption coefficient, and R is the reflection coefficient, then show that the absorbed energy density is $U_a(z) = F_0(1 - R)\alpha e^{-\alpha z}$.
 - The ablation depth is defined as the depth at which the absorbed energy density becomes equal to the energy density required to vaporize the material U_{vap} . Using this definition, show that the ablation depth can be written as $d_{\text{vap}} = \frac{1}{\alpha} \ln\left(\frac{F_0}{F_{\text{th}}}\right)$, where $F_{\text{th}} = \frac{U_{\text{vap}}}{(1 - R)\alpha}$.
 - The following experimental values were measured for the ablation of fused silica:

F_0 (J/cm ² /pulse)	d_{vap} (nm/pulse)
2	40
10	135

Calculate the threshold fluence F_{th} and the absorption coefficient α .

($F_{\text{th}} = 1$ J/cm²/pulse; $\alpha = 1.7 \times 10^5$ /cm)

- Look up the complex refractive index of fused silica at 157 nm from Ref. [5].
 - Look up the explanation offered in Ref. [6] for the large difference between the two absorption constants.
- Consider a projection lithography system that uses an i-line mercury lamp source, NA = 0.6 and $k_1 = 0.5$ with 4× reduction optics. Determine the smallest line width allowed on the photomask and the corresponding line width of the exposed line.
(1.21 μm, 304 nm)
 - Consider a 193i lithography with SADP and NA = 1.3 capable of printing 22-nm-wide lines. Calculate the effective value of k_1 for this system and the depth of focus.
(0.15, 49 nm)

6. Consider the Lloyd's mirror configuration in [Figure 5.12](#). Assume the incident beam is collimated.
 - a. Show that the interfering beams are incident at angles of $+\theta$ and $-\theta$ from the substrate normal.
 - b. Derive [Eqn \(5.6\)](#).
 - c. Assume the source is a 325-nm HeCd laser. The required pattern is a two-dimensional photonic crystal consisting of oval posts that has a pitch in the x -axis of 300 nm and pitch in the y -axis of 250 nm. Describe how this structure can be created using this system.
7. Calculate the wavelength of an electron that has been accelerated by 5 kV. Explain why the exposure resolution using this beam is typically significantly larger than this wavelength in most EBL systems.
(17 pm)
8. Consider a 100-nm-thick metal film with a 250-nm-wide and 400-nm-thick photoresist line on top. If the metal is wet-etched until all of the metal is removed except the area under the resist, then calculate the width of the resulting metal line. Assume the etch selectivity is infinite.
(50 nm)
9. The same structure as above is etched with a plasma instead of wet chemistry. The etch is anisotropic with a selectivity $S = 0.8$. Calculate the width of the resulting metal trace and the thickness of the remaining photoresist when the etch is completed.
(250 nm, 275 nm)

References

- [1] R.E. Honig, Vapor pressure data for the more common elements, *RCA Rev.* 18 (1957) 195–204.
- [2] D.R. Lide (Ed.), *CRC Handbook of Chemistry and Physics*, 2005. Internet version.
- [3] Y. Yamamura, H. Tawara, Energy dependence of ion-induced sputtering yields of monatomic solids, *At. Data Nucl. Data Tables* 62 (2) (March 1996).
- [4] L. Martinu, D. Poitras, Plasma deposition of optical films and coatings: a review, *J. Vac. Sci. Technol. A* 18 (6) (November–December 2000).
- [5] Palik, *Handbook of the Optical Constants of Solids*, CRC Press, 1998.
- [6] P.R. Herman, et al., Processing applications with the 157-nm fluorine excimer laser, *Proc. SPIE* 2992, Excimer Lasers, Optics, Appl. 86 (March 31, 1997).
- [7] S.M. George, Atomic layer deposition: an overview, *Chem. Rev.* 110 (2010) 111–131.
- [8] D.B. Chrisey, G.K. Hubler (Eds), *Pulsed Laser Deposition of Thin Films*, Wiley-Interscience, ISBN-13: 978–0471592181.
- [9] Special section on EUV sources for lithography, *SPIE J. Micro/Nanolithogr., MEMS, MOEMS* 11 (2) (April 2012).
- [10] Special section on photomasks for extreme ultraviolet lithography, *SPIE J. Micro/Nanolithogr., MEMS, MOEMS* 12 (2) (April 2013).

Further reading

- [1] C. Mack, *Fundamental Principles of Optical Lithography: The Science of Micro-fabrication*, John Wiley & Sons Ltd, NY, 2007, ISBN 978-0-470-01893-4.
- [2] C.-S. Kim, S.-H. Ahn, D.-Y. Jang, Review: developments in micro/nanoscale fabrication by focused ion beams, *Vacuum* 86 (2012) 1014–1035.

- [3] H. Schiff, Nanoimprint lithography: an old story in modern times? A review, *J. Vac. Sci. Technol. B* 26 (2) (March/April 2008).
- [4] K.R. Williams, K. Gupta, M. Wasilik, Etch rates for micromachining processing — Part II, *J. Microelectromech. Syst.* 12 (6) (December 2003).
- [5] H. Jansen, H. Gardeniers, M. de Boer, M. Elwenspoek, J. Fluitman, A survey on the reactive ion etching of silicon in microtechnology, *J. Micromech. Microeng.* 6 (1996) 14–28.