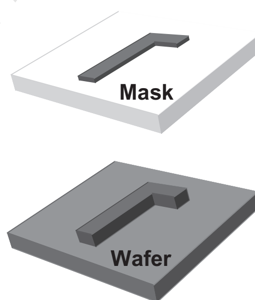
1

Introduction to Semiconductor Lithography

COPYRIGHTED MATERIAL



The fabrication of an integrated circuit (IC) involves a great variety of physical and chemical processes performed on a semiconductor (e.g. silicon) substrate. In general, the various processes used to make an IC fall into three categories: film deposition, patterning and semiconductor doping. Films of both conductors (such as polysilicon, aluminum, tungsten and copper) and insulators (various forms of silicon dioxide, silicon nitride and others) are used to connect and isolate transistors and their components. Selective doping of various regions of silicon allows the conductivity of the silicon to be changed with the application of voltage. By creating structures of these various components, millions (or even billions) of transistors can be built and wired together to form the complex circuitry of a modern microelectronic device. Fundamental to all of these processes is lithography, i.e. the formation of three-dimensional (3D) relief images on the substrate for subsequent transfer of the pattern into the substrate.

The word lithography comes from the Greek *lithos*, meaning stones, and *graphia*, meaning to write. It means quite literally writing on stones. In the case of semiconductor lithography, our stones are silicon wafers and our patterns are written with a light-sensitive polymer called a *photoresist*. To build the complex structures that make up a transistor and the many wires that connect the millions of transistors of a circuit, lithography and etch pattern transfer steps are repeated at least 10 times, but more typically 25 to 40 times to make one circuit. Each pattern being printed on the wafer is aligned to the previously formed patterns as slowly the conductors, insulators and selectively doped regions are built up to form the final device.

The importance of lithography can be appreciated in two ways. First, due to the large number of lithography steps needed in IC manufacturing, lithography typically accounts

Fundamental Principles of Optical Lithography: The Science of Microfabrication, Chris Mack.
© 2007 John Wiley & Sons, Ltd.

for about 30 % of the cost of manufacturing a chip. As a result, IC fabrication factories ('fabs') are designed to keep lithography as the throughput bottleneck. Any drop in output of the lithography process is a drop in output for the entire factory. Second, lithography tends to be the technical limiter for further advances in transistor size reduction and thus chip performance and area. Obviously, one must carefully understand the trade-offs between cost and capability when developing a lithography process for manufacturing. Although lithography is certainly not the only technically important and challenging process in the IC manufacturing flow, historically, advances in lithography have gated advances in IC cost and performance.

1.1 Basics of IC Fabrication

A *semiconductor* is not, as its name might imply, a material with properties between an electrical conductor and an insulator. Instead, it is a material whose conductivity can be readily changed by several orders of magnitude. Heat, light, impurity doping and the application of an electric field can all cause fairly dramatic changes in the electrical conductivity of a semiconductor. The last two can be applied locally and form the basis of a *transistor*: by applying an electric field to a doped region of a semiconductor material, that region can be changed from a good to a poor conductor of electricity, or vice versa. In effect, the transistor works as an electrically controlled *switch*, and these switches can be connected together to form digital logic circuits. In addition, semiconductors can be made to *amplify* an electrical signal, thus forming the basis of analog solid-state circuits.

By far the most common semiconductor in use is silicon, due to a number of factors such as cost, formation of a stable native oxide and vast experience (the first silicon IC was built in about 1960). A wafer of single-crystal silicon anywhere from 75–300 mm in diameter and about 0.6–0.8 mm thick serves as the substrate for the fabrication and interconnection of planar transistors into an IC. The most advanced circuits are built on 200- and 300-mm-diameter wafers. The wafers are far larger than the ICs being made so that each wafer holds a few hundred (and up to a few thousand) IC devices. Wafers are processed in lots of about 25 wafers at a time, and large fabs can have throughputs of greater than 10 000 wafers per week. The cycle time for making a chip, from starting bare silicon wafers to a finished wafer ready for dicing and packaging, is typically 30–60 days. Semiconductor processing (or IC fabrication) involves two major tasks:

- Creating small, interconnected 3D structures of insulators and conductors in order to manipulate local electric fields and currents
- Selectively doping regions of the semiconductor (to create p–n junctions and other electrical components) in order to manipulate the local concentration of charge carriers

1.1.1 Patterning

The 3D microstructures are created with a process called *patterning*. The common *subtractive* patterning process (Figure 1.1) involves three steps: (1) deposition of a uniform film of material on the wafer; (2) lithography to create a positive image of the pattern

Figure 1.1 *A simple subtractive patterning process*

that is desired in the film; and (3) etch to transfer that pattern into the wafer. An *additive* process (such as electroplating) changes the order of these steps: (1) lithography to create negative image of the pattern that is desired; and (2) selective deposition of material into the areas not protected by the lithographically produced pattern. Copper is often patterned additively using the damascene process (named for a unique decorative metal fill process applied to swords and developed in Damascus about 1000 years ago).

Deposition can use many different technologies dependent on the material and the desired properties of the film: oxide growth (direct oxidation of the silicon), chemical vapor deposition (CVD), physical vapor deposition (PVD), evaporation and sputtering. Common films include insulators (silicon dioxide, silicon nitride, phosphorous-doped glass, etc.) and conductors (aluminum, copper, tungsten, titanium, polycrystalline silicon, etc.). Lithography, of course, is the subject of this book and will be discussed at great length in the pages that follow. Photoresists are classed as positive, where exposure to light causes the resist to be removed, and negative, where exposed patterns remain after development. The goal of the photoresist is to resist etching after it has been patterned so that the pattern can be transferred into the film.

1.1.2 Etching

Etch involves both chemical and mechanical mechanisms for removal of the material not protected by the photoresist. *Wet etch*, perhaps the simplest form of etch, uses an etchant solution such as an acid that chemically attacks the underlying film while leaving the photoresist intact. This form of etching is *isotropic* and thus can lead to undercutting as the film is etched from underneath the photoresist. If *anisotropic* etching is desired (and most always it is), directionality must be induced into the etch process. *Plasma etching* replaces the liquid etchant with a plasma – an ionized gas. Applying an electric field causes the ions to be accelerated downward toward the wafer. The resulting etch is a mix of chemical etching due to reaction of the film with the plasma and physical sputtering due to the directional bombardment of the ions hitting the wafer. The chemical nature of the etch can lead to good *etch selectivity* of the film with respect to the resist (selectivity being defined as the ratio of the film etch rate to the resist etch rate) and with respect to the substrate below the film, but is essentially isotropic. Physical sputtering is very

directional (etching is essentially vertical only), but not very selective (the resist etches at about the same rate as the film to be etched). *Reactive ion etching* combines both effects to give good enough selectivity and directionality – the accelerated ions provide energy to drive a chemical etching reaction.

The photoresist property of greatest interest for etching is the etch selectivity, which is dependent both on the photoresist material properties and the nature of the etch process for the specific film. Good etch processes often have etch selectivities in excess of 4 (for example, polysilicon with a novolac resist), whereas poor selectivities can be as low as 1 (for example, when etching an organic bottom antireflection coating). Etch selectivity and the thickness of the film to be etched determine the minimum required resist thickness. Mechanical properties of the resist, such as adhesion to the substrate and resistance to mechanical deformation such as bending of the pattern, also play a role during etching.

For an etch process without perfect selectivity, the shape and size of the final etched pattern will depend on not only the size of the resist pattern but its shape as well. Consider a resist feature whose straight sidewalls make an angle θ with respect to the substrate (Figure 1.2). Given vertical and horizontal etch rates of the resist R_V and R_H, respectively, the rate at which the critical dimension (CD) shrinks will be

$$\frac{dCD}{dt} = 2(R_H + R_V \cot\theta) \tag{1.1}$$

Thus, the rate at which the resist CD changes during the etch is a function of the resist sidewall angle. As the angle approaches 90°, the vertical component ceases to contribute and the rate of CD change is at its minimum. In fact, Equation (1.1) shows three ways to minimize the change in resist CD during the etch: improved etch selectivity (making both R_H and R_V smaller), improved anisotropy (making R_H/R_V smaller) and a sidewall close to vertical (making $\cot\theta$ smaller).

The example above, while simple, shows clearly how resist profile shape can affect pattern transfer. In general, the ideal photoresist shape has perfectly vertical sidewalls. Other nonideal profile shapes, such as rounding of the top of the resist and resist footing, will also affect pattern transfer.

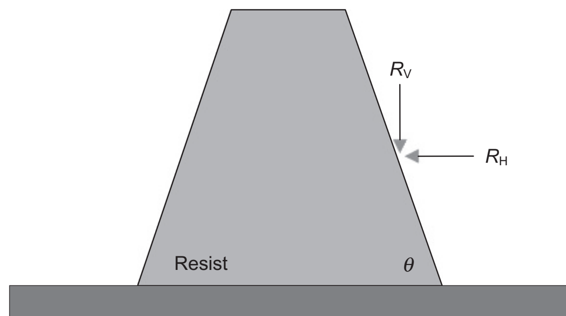


Figure 1.2 *Erosion of a photoresist line during etching, showing the vertical and horizontal etch rate components*

1.1.3 Ion Implantation

Selective doping of certain regions of the semiconductor begins with a patterning step (Figure 1.3). Regions of the semiconductor that are not covered by photoresist are exposed to a dopant impurity. *p-type* dopants like boron have three outer shell electrons and when inserted into the crystal lattice in place of silicon (which has four outer electrons) create mobile *holes* (empty spots in the lattice where an electron could go). *n-type* dopants such as phosphorous, arsenic and antimony have five outer electrons, which create excess mobile electrons when used to dope silicon. The interface between p-type regions and n-type regions of silicon is called a *p–n junction* and is one of the foundational structures in the building of semiconductor devices.

The most common way of doping silicon is with *ion implantation*. The dopant is ionized in a high-vacuum environment and accelerated into the wafer by an electric field (voltages of hundreds of kilovolts are common). The depth of penetration of the ions into the wafer is a function of the ion energy, which is controlled by the electric field. The force of the impact of these ions will destroy the crystal structure of the silicon, which then must be restored by a high-temperature annealing step, which allows the crystal to reform (but also causes diffusion of the dopant). Since the resist must block the ions in the regions where dopants are not desired (that is, in the regions covered by the resist), the resist thickness must exceed the penetration depth of the ions.

Ion implantation penetration depth is often modeled as a Gaussian distribution of depths, i.e. the resulting concentration profile of implanted dopants follows a Gaussian shape. The mean of the distribution (the peak of the concentration profile) occurs at a depth called the *projected range*, R_p. The standard deviation of the depth profile is called the *straggle*, ΔR_p. For photoresists, the projected range varies approximately linearly with implant energy, and inversely with the atomic number of the dopant (Figure 1.4a). A more accurate power-law model (as shown in Figure 1.4a) is described in Table 1.1. The straggle varies approximately as the square root of implant energy, and is about independent of the dopant (Figure 1.4b). For higher energies (1 MeV and above), higher atomic number dopants produce more straggle. When more detailed predictions are needed, Monte Carlo implantation simulators are frequently used.

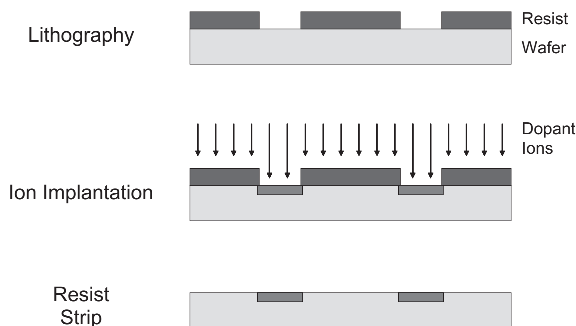


Figure 1.3 *Patterning as a means of selective doping using ion implantation*

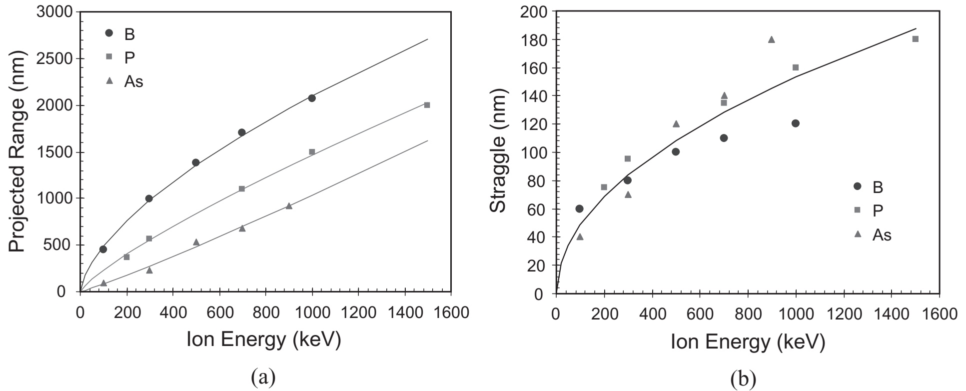


Figure 1.4 *Measured and fitted ion implantation penetration depths for boron, phosphorous and arsenic in AZ 7500 resist: (a) projected range and (b) straggle. Symbols are data*[1] *and curves are power-law fits to the data as described in Table 1.1. For the straggle data, the empirical model fit is* $\Delta R_p = 4.8E^{0.5}$ *where* E *is the ion energy in keV and* ΔR_p *is the straggle in nm*

Table 1.1 *Empirical model of ion implanted projected range* (R_p, *in nm) into photoresist versus ion energy* (E, *in keV) as* $R_p = aE^b$

Dopant	Coefficient *a*	Power *b*
Boron	26.9	0.63
Phosphorous	5.8	0.80
Arsenic	0.49	1.11

In order to mask the underlying layers from implant, the resist thickness must be set to at least

$$\textit{resist thickness} \geq R_{\mathrm{p}} + m\Delta R_{\mathrm{p}} \tag{1.2}$$

where m is set to achieve a certain level of dopant penetration through the resist. For example, if the dopant concentration at the bottom of the resist cannot be more than 10^{-4} times the peak concentration (a typical requirement), then m should be set to 4.3. While the resolution requirements for the implant layers tend not to be challenging, often the thickness required for adequate stopping power does pose real challenges to the lithographer. Carbonization of the resist during high energy and high dose implantation (as well as during plasma etching) can also result in a film that is very difficult to strip away at the end.

1.1.4 Process Integration

The combination of patterning and selective doping allows the buildup of the structures required to make transistors. Figure 1.5 shows a diagrammatical example of a pair of CMOS (complementary metal oxide semiconductor) transistors. Subsequent metal layers (up to 10 metal levels are not uncommon) can connect the many transistors into a full

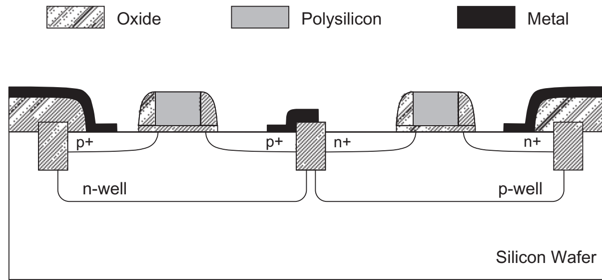


Figure 1.5 *Cross section of a pair of CMOS transistors showing most of the layers through metal 1*

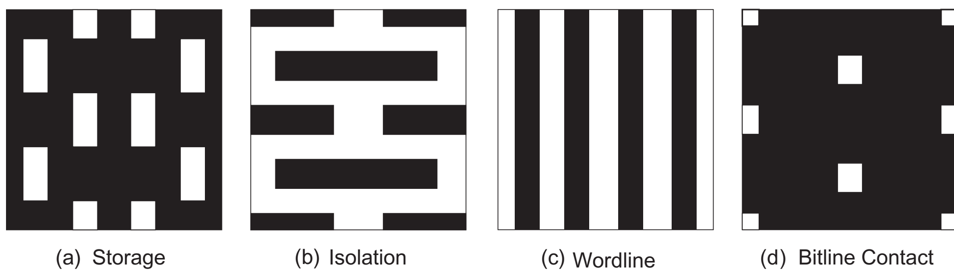


Figure 1.6 *Critical mask level patterns for a 1-Gb DRAM chip[2]. Each pattern repeats in both* x *and* y *many times to create the DRAM array*

circuit and the final metal layer will provide connections to the external pins of the device package.

Many lithographic levels are required to fabricate an IC, but about 1/3 of these levels are considered ‘critical’, meaning that those levels have challenging lithographic requirements. Which levels are critical depends on the process technology (CMOS logic, DRAM, BiCMOS, etc.). The most common critical levels of a CMOS process are active area, shallow trench isolation (STI), polysilicon gate, contact (between metal 1 and poly) and via (between metal layers), and metal 1 (the first or bottom most metal layer). For a large logic chip with 10 layers of metal, the first three will be ‘1×’, meaning the dimensions are at or nearly at the minimum metal 1 dimensions. The next three metal layers will be ‘2×’, with dimensions about twice as big as the 1× metal levels. The next few metal levels will be 4×, with the last few levels as large as 10×. For a DRAM device, some of the critical levels are known as storage, isolation, wordline and bitline contact (see Figure 1.6 for example design patterns for these four levels).

1.2 Moore’s Law and the Semiconductor Industry

The impact of semiconductor ICs on modern life is hard to overstate. From computers to communication, entertainment to education, the growth of electronics technology, fueled

by advances in semiconductor chips, has been phenomenal. The impact has been so profound that it is now often taken for granted: consumers have come to expect increasingly sophisticated electronics products at ever lower prices, and semiconductor companies expect growth and profits to improve continually. The role of optical lithography in these trends has been, and will continue to be, vital.

The remarkable evolution of semiconductor technology from crude single transistors to billion-transistor microprocessors and memory chips is a fascinating story. One of the first 'reviews' of progress in the semiconductor industry was written by Gordon Moore, a founder of Fairchild Semiconductor and later Intel, for the 35th anniversary issue of *Electronics* magazine in 1965.[3] After only 6 years since the introduction of the first commercial planar transistor in 1959, Moore observed an astounding trend – the number of electrical components per IC chip was doubling every year, reaching about 60 transistors in 1965. Extrapolating this trend for a decade, Moore predicted that chips with 64 000 components would be available by 1975! Although extrapolating any trend by three orders of magnitude can be quite risky, what is now known as Moore's Law proved amazingly accurate.

Some important details of Moore's remarkable 1965 paper have become lost in the lore of Moore's Law. First, Moore described the number of components per IC, which included resistors and capacitors, not just transistors. Later, as the digital age reduced the predominance of analog circuitry, transistor count became a more useful measure of IC complexity. Further, Moore clearly defined the meaning of the 'number of components per chip' as the number which minimized the cost per component. For any given level of manufacturing technology, one can always add more components – the problem being a reduction in yield and thus an increase in the cost per component. As any modern IC manufacturer knows, cramming more components onto ICs only makes sense if the resulting manufacturing yield allows costs that result in more commercially desirable chips. This 'minimum cost per component' concept is in fact the ultimate driving force behind the economics of Moore's Law.

Consider a very simple cost model for chip manufacturing as a function of lithographic feature size. For a given process, the cost of making a chip is proportional to the area of silicon consumed divided by the final yield of the chips. Will shrinking the feature sizes on the chip result in an increase or decrease in cost? The area of silicon consumed will be roughly proportional to the feature size squared. But yield will also be a function of feature size. Assuming that the only yield limiter will be the parametric effects of reduced feature size, a simple yield model might look something like

$$Yield = 1 - e^{-(w-w_0)^2/2\sigma^2}, \quad Cost \propto \frac{w^2}{Yield} \tag{1.3}$$

where w is the feature size (which must be greater than w_0 for this model), w_0 is the ultimate resolution (feature size at which the yield goes to zero), and σ is the sensitivity of yield to feature size. Figure 1.7 shows this yield model and the resulting cost function for arbitrary but reasonable parameters.

But minimizing cost is not really the goal of a semiconductor fab – it is maximizing profit. Feature size affects total profit in two ways other than chip cost. The number of chips per wafer is inversely proportional to the area of each chip, thus increasing the

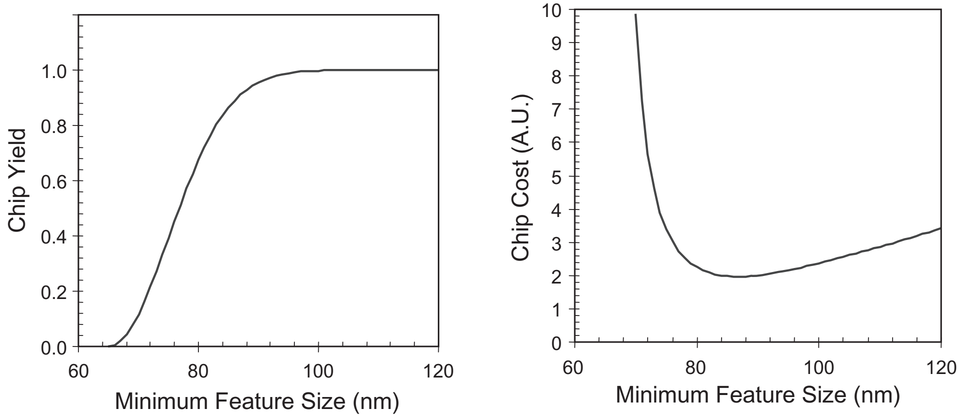


Figure 1.7 *A very simple yield and cost model shows the feature size that minimizes chip cost (w_0 = 65 nm, σ = 10 nm). Lowest chip cost occurs, in this case, when w = 87 nm, corresponding to a chip yield of about 90 %*

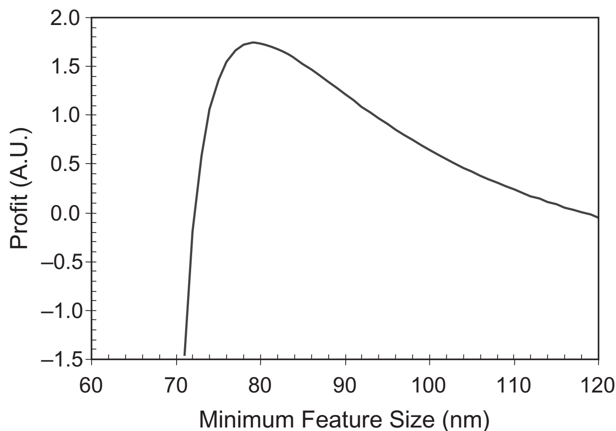


Figure 1.8 *Example fab profit curve using the yield and cost models of Figure 1.7 and assuming the value of the chip is inversely proportional to the minimum feature size. For this example, maximum profit occurs when w = 80 nm, even though the yield is only 65 %*

number of chips that can be sold (that is, the total possible throughput of chips for the fab). Also, the value of each chip is often a function of the feature size. Smaller transistors generally run faster and fit in smaller packages – both desirable features for many applications. Assuming the price that a chip can be sold for is inversely proportional to the minimum feature size on the chip, an example profit model for a fab is shown in Figure 1.8. The most important characteristic is the steep falloff in profit that occurs when trying to use a feature size below the optimum. For the example here, the profit goes to zero if one tries to shrink the feature size by 10 % below its optimum value (unless, of course,

the yield curve can be improved). It is a difficult balancing act for a fab to try to maximize its profit by shrinking feature size without going too far and suffering from excessive yield loss.

In 1975, Moore revisited his 1965 prediction and provided some critical insights into the technological drivers of the observed trends.[4] Checking the progress of component growth, the most advanced memory chip at Intel in 1975 had 32 000 components (but only 16 000 transistors). Thus, Moore's original extrapolation by three orders of magnitude was off by only a factor of 2. Even more importantly, Moore divided the advances in circuit complexity among its three principle components: increasing chip area, decreasing feature size, and improved device and circuit designs. Minimum feature sizes were decreasing by about 10 % per year (resulting in transistors that were about 21 % smaller in area, and an increase in transistors per area of 25 % each year). Chip area was increasing by about 20 % each year (Figure 1.9). These two factors alone resulted in a 50 % increase in the number of transistors per chip each year. Design cleverness made up the rest of the improvement (33 %). In other words, the 2× improvement = (1.25)(1.20)(1.33).

Again, there are important details in Moore's second observation that are often lost in the retelling of Moore's Law. How is 'minimum feature size' defined? Moore explained that both the linewidths and the spacewidths used to make the circuits are critical to density. Thus, his density-representing feature size was an average of the minimum linewidth and the minimum spacewidth used in making the circuit. Today, we use the equivalent metric, the minimum pitch divided by 2 (called the minimum half-pitch). Unfortunately, many modern forecasters express the feature size trend using features that do not well represent the density of the circuit. Usually, minimum half-pitch serves this purpose best.

By breaking the density improvement into its three technology drivers, Moore was able to extrapolate each trend into the future and predict a change in the slope of his

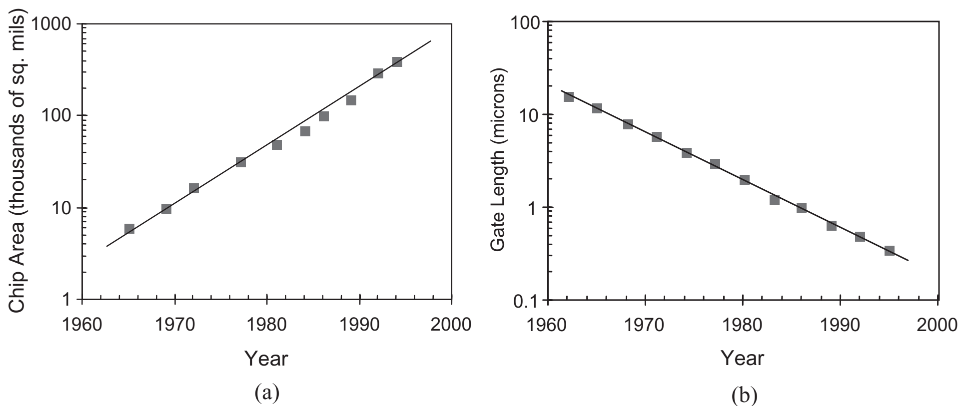


Figure 1.9 *Moore's Law showing (a) an exponential increase (about 15 % per year) in the area of a chip, and (b) an exponential decrease (about 11 % per year) in the minimum feature size on a chip (shown here for DRAM initial introduction)*

observation. Moore saw the progress in lithography allowing continued feature size shrinks to 'one micron or less'. Continued reductions in defect density and increases in wafer size would allow the die area trend to continue. But in looking at the 'device and circuit cleverness' component of density improvement, Moore saw a limit. Although improvements in device isolation and the development of the MOS transistor had contributed to greater packing density, Moore saw the latest circuits as near their design limits. Predicting an end to the design cleverness trend in 4 or 5 years, Moore predicted a change in the slope of his trend from doubling every year, to doubling every 2 years.

Moore's prediction of a slowdown was both too pessimistic and too generous. The slowdown from doubling each year had already begun by 1975 with Intel's 16-Kb memory chip. The 64-Kb DRAM chip, which should have been introduced in 1976 according to the original trend, was not available commercially until 1979. However, Moore's prediction of a slowdown to doubling components every 2 years instead of every year was too pessimistic. The 50 % improvement in circuit density each year due to feature size and die size was really closer to 60 % (according to Moore's retelling of the story[5]), resulting in a doubling of transistor counts per chip every 18 months or so (Figure 1.10). Offsetting the curve to switch from component counts to transistor counts and beginning with the 64-Kb DRAM in 1979, the industry followed the 'new' Moore's Law trend throughout the 1980s and early 1990s.

After 40 years, extrapolation of Moore's Law now seems less risky. In fact, predictions of future industry performance have reached such a level of acceptance that they have been codified in an industry-sanctioned 'roadmap' of the future. The *National Technology Roadmap for Semiconductors* (NTRS)[6] was first developed by the Semiconductor Industry Association in 1994 to serve as an industry-standard Moore's Law. It extrapolated then current trends to the year 2010, where 70-nm minimum feature sizes were predicted to enable 64-Gb DRAM chip production. This official industry roadmap has been updated many times, going international in 1999 to become the ITRS, the *International Technology Roadmap for Semiconductors*.

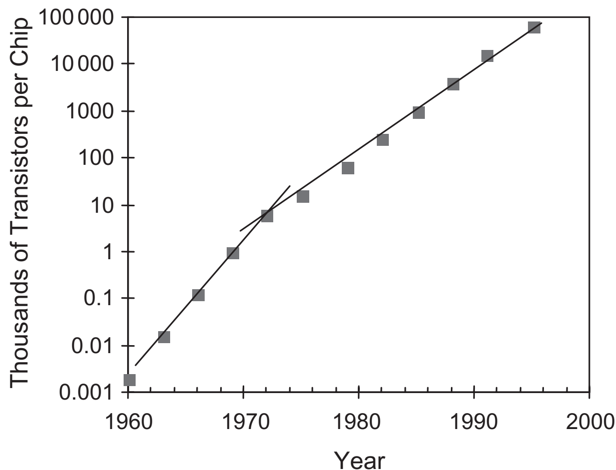


Figure 1.10 *Moore's Law showing an exponential increase in the number of transistors on a semiconductor chip over time (shown here for DRAM initial introduction)*

Ultimately, the drivers for technology development fall into two categories: push and pull. *Push drivers* are technology enablers, those things that make it possible to achieve the technical improvements. Moore described the three push drivers as increasing chip area, decreasing feature size and design cleverness. *Pull drivers* are the economic drivers, those things that make it worthwhile to pursue the technical innovations. Although the two drivers are not independent, it is the economic drivers that always dominate. As Bob Noyce, cofounder of Intel, wrote in 1977 '. . . further miniaturization is less likely to be limited by the laws of physics than by the laws of economics.'[7]

The economic drivers for Moore's Law are extraordinarily compelling. As the dimensions of a transistor shrink, the transistor becomes smaller, faster, consumes less power and in many cases is more reliable. All of these factors make the transistor more desirable for virtually every possible application. But there is more. Historically, the semiconductor industry has been able to manufacture silicon devices at an essentially constant cost per area of processed silicon. Thus, as the devices shrink, they enjoy a shrinking cost per transistor. As many have observed, it is a life without tradeoffs (unless, of course, you consider the stress on the poor engineers trying to make all of this happen year after year). Each step along the roadmap of Moore's Law virtually guarantees economic success. Advances in lithography, and in particular optical lithography, have been critical enablers to the continued rule of Moore's Law.

The death of optical lithography has been predicted so often by industry pundits, incorrectly so far, that it has become a running joke among lithographers. In 1979, conventional wisdom limited optical lithography to 1-μm resolution and a 1983 demise (to be supplanted by electron-beam imaging systems).[8] By 1985, the estimate was revised to 0.5-μm minimum resolution and a 1993 replacement by x-ray lithography.[9] The reality was quite a bit different. In 2006, optical lithography was used for 65-nm production (about 90-nm half-pitch). It seems likely that optical lithography will be able to manufacture devices with 45-nm half-pitch, and experts hedge their bets on future generations. Interestingly, the resolution requirements of current and future lithography processes are not so aggressive that they cannot be met with today's technology – electron beam and x-ray lithography have both demonstrated resolution to spare. The problem is one of cost. Optical lithography is unsurpassed in the cost per pixel (one square unit of minimum resolution) when printing micron-sized and submicron features on semiconductor wafers. To keep the industry on Moore's Law well into the 21st century, advances in optical lithography must continue.

1.3 Lithography Processing

Optical lithography is basically a photographic process by which a light-sensitive polymer, called a photoresist, is exposed and developed to form 3D relief images on the substrate. In general, the ideal photoresist image has the exact shape of the designed or intended pattern in the plane of the substrate, with vertical walls through the thickness of the resist. Thus, the final resist pattern should be binary: parts of the substrate are covered with resist while other parts are completely uncovered. This binary pattern is needed for pattern transfer since the parts of the substrate covered with resist will be protected from etching, ion implantation, or other pattern transfer mechanism.

The general sequence of processing steps for a typical optical lithography process is: substrate preparation, photoresist spin coat, post-apply bake, exposure, post-exposure bake, development and postbake. Metrology and inspection followed by resist strip are the final operations in the lithographic process, after the resist pattern has been transferred into the underlying layer. This sequence is shown diagrammatically in Figure 1.11, and most of these steps are generally performed on several tools linked together into a contiguous unit called a *lithographic cluster* or *cell* (Figure 1.12). A brief discussion of each

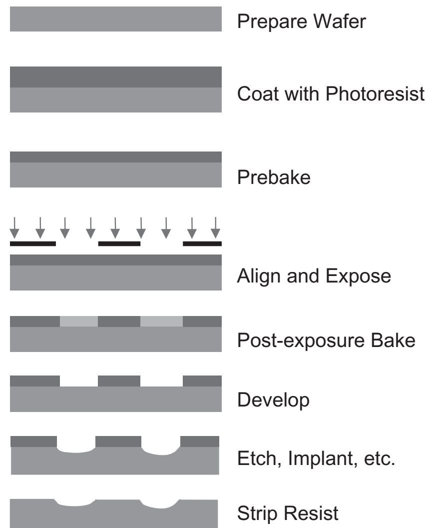


Figure 1.11 *Example of a typical sequence of lithographic processing steps, illustrated for a positive resist*

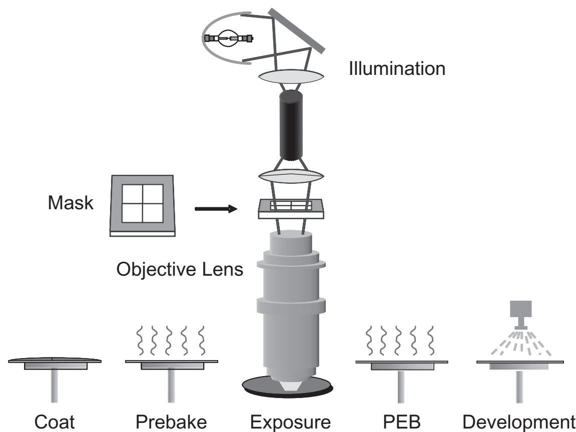


Figure 1.12 *Iconic representation of the integration of the various lithographic process steps into a photolithography cell. Many steps, such as chill plates after the bake steps, have been omitted*

step is given below, pointing out some of the practical issues involved in photoresist processing. More fundamental and theoretical discussions on these topics will be provided in subsequent chapters.

1.3.1 Substrate Preparation

Substrate preparation is intended to improve the adhesion of the photoresist material to the substrate and provide for a contaminant-free resist film. This is accomplished by one or more of the following processes: substrate cleaning to remove contamination, dehydration bake to remove water and addition of an adhesion promoter. Substrate contamination can take the form of particulates or a film and can be either organic or inorganic. Particulates result in defects in the final resist pattern, whereas film contamination can cause poor adhesion and subsequent loss of linewidth control. Particulates generally come from airborne particles or contaminated liquids (e.g. dirty adhesion promoter). The most effective way of controlling particulate contamination is to eliminate their source. Since this is not always practical, chemical/mechanical cleaning is used to remove particles. Organic films, such as oils or polymers, can come from vacuum pumps and other machinery, body oils and sweat, and various polymer deposits leftover from previous processing steps. These films can generally be removed by chemical, ozone, or plasma stripping. Similarly, inorganic films, such as native oxides and salts, can be removed by chemical or plasma stripping. One type of contaminant – adsorbed water – is removed most readily by a high-temperature process called a *dehydration bake*.

A dehydration bake, as the name implies, removes water from the substrate surface by baking at temperatures of 200 to 400 °C for up to 60 minutes. The substrate is then allowed to cool (preferably in a dry environment) and coated as soon as possible. It is important to note that water will re-adsorb on the substrate surface if left in a humid (nondry) environment. A dehydration bake is also effective in volatilizing organic contaminants, further cleaning the substrate. Often, the normal sequence of processing steps involves some type of high-temperature process immediately before coating with photoresist, for example, thermal oxidation. If the substrate is coated immediately after the high-temperature step, the dehydration bake can be eliminated. A typical dehydration bake, however, does not completely remove water from the surface of silica substrates (including silicon, polysilicon, silicon dioxide and silicon nitride). Surface silicon atoms bond strongly with a monolayer of water forming silanol groups (SiOH) and bake temperatures in excess of 600 °C are required to remove this final layer of water. Further, the silanol quickly reforms when the substrate is cooled in a nondry environment. As a result, the preferred method of removing this silanol is by chemical means.

Adhesion promoters are used to react chemically with surface silanol and replace the —OH group with an organic functional group that, unlike the hydroxyl group, offers good adhesion to photoresist. Silanes are often used for this purpose, the most common being hexamethyl disilizane (HMDS).[10] (As a note, HMDS adhesion promotion was first developed for fiberglass applications, where adhesion of the resin matrix to the glass fibers is important.) The HMDS can be applied by spinning a diluted solution (10–20 % HMDS in cellosolve acetate, xylene, or a fluorocarbon) directly on to the wafer and allowing the HMDS to spin dry (HMDS is quite volatile at room temperature). If the HMDS is not allowed to dry properly, dramatic loss of adhesion will result. Although direct spinning

is easy, it is only effective at displacing a small percentage of the silonal groups. By far the preferred method of applying the adhesion promoter is by subjecting the substrate to HMDS vapor at elevated temperatures and reduced pressure. This allows good coating of the substrate without excess HMDS deposition, and the higher temperatures cause more complete reaction with the silanol groups. Once properly treated with HMDS, the substrate can be left for up to several days without significant re-adsorption of water. Performing the dehydration bake and vapor prime in the same oven gives optimum performance. Such vapor prime systems are often integrated into the wafer processing tracks used for the subsequent steps of resist coating and baking.

A simple method for testing for adsorbed water on the wafer surface, and thus the likelihood of resist adhesion failure, is to measure the contact angle of a drop of water. If a drop of water wets the surface (has a low contact angle), the surface is hydrophilic and the resist will be prone to adhesion failure during development. For a very hydrophobic surface, water will have a large contact angle (picture water beading up on a waxed automobile). Contact angles can be easily measured on a primed wafer using a goniometer, and should be in the 50–70° range for good resist adhesion[11] (see Figure 1.13).

1.3.2 Photoresist Coating

A thin, uniform coating of photoresist at a specific, well-controlled thickness is accomplished by the seemingly simple process of *spin coating*. The photoresist, rendered into a liquid form by dissolving the solid components in a solvent, is poured onto the wafer, which is then spun on a turntable at a high speed producing the desired film. (For the case of DNQ/novolac resists, the resist solutions are often supersaturated, making them prone to precipitation.) Stringent requirements for thickness control and uniformity and low defect density call for particular attention to be paid to this process, where a large number of parameters can have significant impact on photoresist thickness uniformity and control. There is the choice between static dispense (wafer stationary while resist is dispensed) or dynamic dispense (wafer spinning while resist is dispensed), spin speeds and times, and accelerations to each of the spin speeds. Also, the volume of the resist dispensed and properties of the resist (such as viscosity, percent solids and solvent composition) and the substrate (substrate material and topography) play an important role in the resist thickness uniformity. Further, practical aspects of the spin operation, such as exhaust, ambient temperature and humidity control, resist temperature, spin cup geometry, point-of-use filtration and spinner cleanliness often have significant effects on the resist film. Figure 1.14a shows a generic photoresist spin coat cycle. At the end of this cycle, a thick, solvent-rich film of photoresist covers the wafer, ready for post-apply bake. By the end

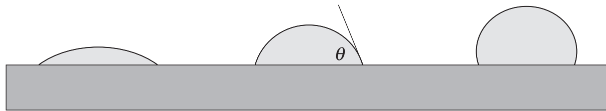


Figure 1.13 A water droplet on the surface of the wafer indicates the hydrophobicity of the wafer: the left-most drop indicates a hydrophilic surface, the right-most drop shows an extremely hydrophobic surface. The middle case, with a contact angle of 70°, is typically about optimum for resist adhesion

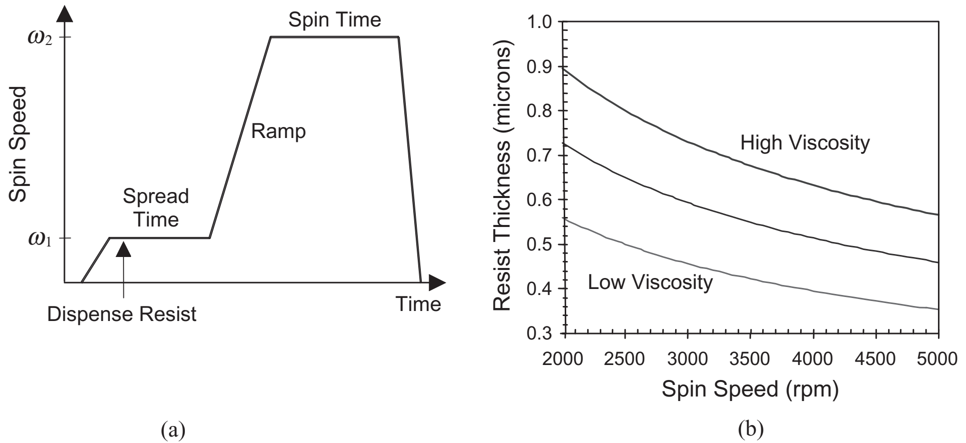


Figure 1.14 *Photoresist spin coat cycle: (a) pictorial representation (if $\omega_1 > 0$, the dispense is said to be dynamic), and (b) photoresist spins speed curves for different resist viscosities showing how resist thickness after post-apply bake varies as (spin speed)$^{-1/2}$*

of the post-apply bake, the film can have a thickness controlled to within 1–2 nm across the wafer and wafer-to-wafer.

The rheology of resist spin coating is complex and yet results in some simple, and seemingly unexpected, properties of the final resist film. Spinning results in centrifugal forces pushing the liquid photoresist toward the edge of the wafer where excess resist is flung off. The frictional force of viscosity opposes this centrifugal force. As the film thins, the centrifugal force (which is proportional to the mass of the resist on the wafer) decreases. Also, evaporation of solvent leads to dramatic increases in the viscosity of the resist as the film dries (as the resist transitions from a liquid to a solid, the viscosity can increase by 7–10 orders of magnitude). Eventually the increasing viscous force exceeds the decreasing centrifugal force and the resist stops flowing. This generally occurs within the first second of the spin cycle (often before the wafer has fully ramped to its final spin speed). The remaining portion of the spin cycle causes solvent evaporation without mass flow of the resist solids. The separation of the spin cycle into a very quick radial mass flow (coating stage) followed by a long evaporation of solvent (drying stage) provides for some of the basic and important properties of spin coating. Since the overall spin time is much longer than the coating stage time, the final thickness of resist is virtually independent of the initial volume of resist dispensed onto the wafer above a certain threshold. For laminar flow of air above the spinning wafer, the amount of drying (mass transfer of solvent) will be proportional to the square root of the spin speed. And since most of the thinning of the resist comes from the drying stage, the final thickness of the resist will vary inversely with the square root of the spin speed. Finally and most importantly, both the coating stage and drying stage produce a film whose thickness is not dependent on the radial position on the wafer.

Although theory exists to describe the spin coat process rheologically,[12,13] in practical terms the variation of photoresist thickness and uniformity with the process parameters are determined experimentally. The photoresist spin speed curve (Figure 1.14b) is an

essential tool for setting the spin speed to obtain the desired resist thickness. As mentioned above, the final resist thickness varies as one over the square root of the spin speed (ω) and is roughly proportional to the liquid photoresist viscosity (ν) to the 0.4–0.6 power:

$$thickness \propto \frac{\nu^{0.4}}{\omega^{0.5}} \tag{1.4}$$

For a given desired resist thickness, the appropriate spin speed is chosen according to the spin curve. However, there is a limited range of acceptable spin speeds. Speeds less than 1000 rpm are harder to control and do not produce uniform films. If the spin speed is too high, turbulent airflow at the edge of the wafer will limit uniformity. The onset of turbulence depends on the Reynolds number *Re*, which for a rotating disk is

$$Re = \frac{\omega r^2}{\nu_{air}} \tag{1.5}$$

where r is the wafer radius, ν_{air} is the kinematic viscosity of air (about $1.56 \times 10^{-5}\ \mathrm{m^2/s}$ at standard conditions). The onset of turbulence begins for Reynolds numbers of about 300 000.[14,15] Instabilities in the flow, in the form of spiral vortices, can occur at Reynolds numbers as low as 100 000 without careful design of the spin coat chamber. [Note that sometimes the square root of the expression (1.5) is used as the Reynolds number, so that the threshold for turbulence is 550.] For a 300-mm wafer, this means the maximum spin speed is on the order of 2000 rpm. If the desired resist thickness cannot be obtained over the acceptable range of spin speeds, a different viscosity resist formulation can be chosen. Typical resist viscosities range from 5 to 35 cSt (1 Stoke = 1 $\mathrm{cm^2/s}$). As a point of reference, water has a viscosity of about 1 cSt at room temperature.

Unfortunately, the forces that give rise to uniform resist coatings also cause an unwanted side effect: edge beads. The fluid flow discussion above described a balance of the centrifugal and viscous forces acting on the resist over the full surface of the wafer. However, at the edge of the wafer, a third force becomes significant. Surface tension at the resist–air interface results in a force pointing inward perpendicular to the resist surface. Over most of the wafer, this force is pointing downward and thus does not impact the force balance of spinning plus friction. However, at the edge of the wafer, this force must point inward toward the center of the wafer (Figure 1.15). The extra force adding to the viscous force will stop the flow of resist sooner at the edge than over the central portion of the wafer,

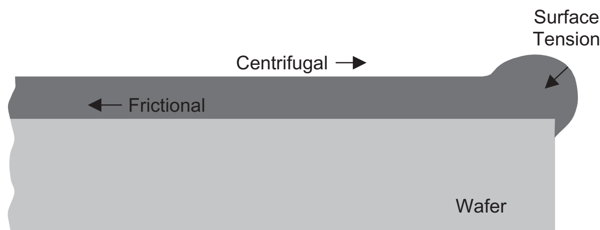


Figure 1.15 *A balance of spin-coat forces at the wafer edge leads to the formation of a resist edge bead*

resulting in an accumulation of resist at the edge. This accumulation is called an *edge bead*, which usually exists within the outer 1–2 mm of the wafer and can be 10–30 times thicker than the rest of the resist film.

The existence of an edge bead is detrimental to the cleanliness of subsequent wafer processing. Tools which grab the wafer by the edge will flake off the dried edge bead, resulting in very significant particulate contamination. Consequently, removal of the edge bead is required. Within the spin-coat chamber and immediately after the resist spin coating is complete, a stream of solvent (called an EBR, edge bead remover) is directed at the edge of the wafer while it slowly spins. Resist is dissolved off the edge and over the outer 1.5–2 mm of the wafer surface.

1.3.3 Post-Apply Bake

After coating, the resulting resist film will contain between 20 and 40 % solvent by weight. The post-apply bake (PAB) process, also called a softbake or a prebake, involves drying the photoresist after spin coat by removing most of this excess solvent. The main reason for reducing the solvent content is to stabilize the resist film. At room temperature, an unbaked photoresist film will lose solvent by evaporation, thus changing the properties of the film with time. By baking the resist, the majority of the solvent is removed and the film becomes stable at room temperature. There are four major effects of removing solvent from a photoresist film: (1) film thickness is reduced; (2) post-exposure bake and development properties are changed; (3) adhesion is improved; and (4) the film becomes less tacky and thus less susceptible to particulate contamination. Typical post-apply bake processes leave between 3 and 10 % residual solvent in the resist film (depending on resist and solvent type, as well as bake conditions), sufficiently small to keep the film stable during subsequent lithographic processing.

Unfortunately, there can be other consequences of baking photoresists. At temperatures greater than about 70 °C, the photosensitive component of a typical resist mixture, called the photoactive compound (PAC), may begin to decompose. Also, the resin, another component of the resist, can cross-link and/or oxidize at elevated temperatures. Both of these effects are undesirable. Thus, one must search for the optimum post-apply bake conditions that will maximize the benefits of solvent evaporation and minimize the detriments of resist decomposition. For chemically amplified resists, residual solvent can significantly influence diffusion and reaction properties during the post-exposure bake, necessitating careful control over the post-apply bake process. Fortunately, these modern resists do not suffer from significant decomposition of the photosensitive components during post-apply bake.

There are several methods that can be used to bake photoresists. The most obvious method is an oven bake. Convection oven baking of conventional photoresists at 90 °C for 30 minutes was typical during the 1970s and early 1980s, but currently the most popular bake method is the hot plate. The wafer is brought either into intimate vacuum contact with or close proximity to a hot, high-mass metal plate. Due to the high thermal conductivity of silicon, the photoresist is heated to near the hot plate temperature quickly (in about 5 seconds for hard contact, or about 20 seconds for proximity baking). The greatest advantage of this method is an order of magnitude decrease in the required bake time over convection ovens, to about 1 minute, and the improved uniformity of the bake.

In general, proximity baking is preferred to reduce the possibility of particle generation caused by contact with the backside of the wafer.

When the wafer is removed from the hot plate, baking continues as long as the wafer is hot. The total bake process cannot be well controlled unless the cooling of the wafer is also well controlled. In other words, the bake process should be thought of in terms of an integrated thermal history, from the start of the bake till the wafer has sufficiently cooled. As a result, hot plate baking is always followed immediately by a chill plate operation, where the wafer is brought in contact or close proximity to a cool plate (kept at a temperature slightly below room temperature). After cooling, the wafer is ready for its lithographic exposure.

1.3.4 Alignment and Exposure

The basic principle behind the operation of a photoresist is the change in solubility of the resist in a developer upon exposure to light. In the case of the standard diazonaphthoquinone positive photoresist, the PAC, which is not soluble in the aqueous base developer, is converted to a carboxylic acid on exposure to UV light in the range of 350–450 nm. The carboxylic acid product is very soluble in the basic developer. Thus, a spatial variation in light energy incident on the photoresist will cause a spatial variation in solubility of the resist in developer.

Contact and proximity lithography are the simplest methods of exposing a photoresist through a master pattern called a photomask (Figure 1.16). Contact lithography offers reasonably high resolution (down to about the wavelength of the radiation), but practical problems such as mask damage (or equivalently, the formation of mask defects) and resulting low yield make this process unusable in most production environments. Proximity printing reduces mask damage by keeping the mask a set distance above the wafer (e.g. 20 μm). Unfortunately, the resolution limit is increased significantly. For a mask–wafer gap of g and an exposure wavelength of λ,

$$Resolution \sim \sqrt{g\lambda} \tag{1.6}$$

Because of the high defect densities of contact printing and the poor resolution of proximity printing, by far the most common method of exposure is *projection printing*.

Projection lithography derives its name from the fact that an image of the mask is projected onto the wafer. Projection lithography became a viable alternative to contact/proximity printing in the mid-1970s when the advent of computer-aided lens design and

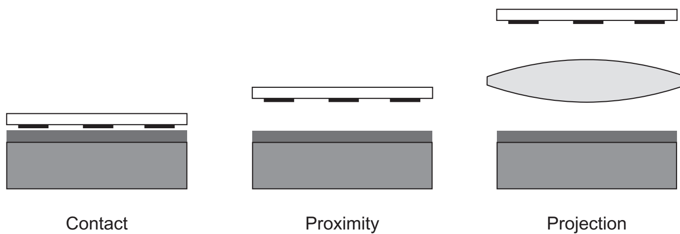


Figure 1.16 *Lithographic printing in semiconductor manufacturing has evolved from contact printing (in the early 1960s) to projection printing (from the mid-1970s to today)*

improved optical materials and manufacturing methods allowed the production of lens elements of sufficient quality to meet the requirements of the semiconductor industry. In fact, these lenses have become so perfect that lens defects, called aberrations, play only a small role in determining the quality of the image. Such an optical system is said to be *diffraction-limited*, since it is diffraction effects and not lens aberrations which, for the most part, determine the shape of the image.

There are two major classes of projection lithography tools – *scanning* and *step-and-repeat* systems. Scanning projection printing, as pioneered by the Perkin-Elmer company,[16] employs reflective optics (i.e. mirrors rather than lenses) to project a slit of light from the mask onto the wafer as the mask and wafer are moved simultaneously past the slit. Exposure dose is determined by the intensity of the light, the slit width and the speed at which the wafer is scanned. These early scanning systems, which use polychromatic light from a mercury arc lamp, are 1 : 1, i.e. the mask and image sizes are equal. Step-and-repeat cameras (called steppers for short), first developed by GCA Corp., expose the wafer one rectangular section (called the image field) at a time and can be 1 : 1 or reduction. These systems employ refractive optics (i.e. lenses) and are usually quasi-monochromatic. Both types of systems (Figure 1.17) are capable of high-resolution imaging, although reduction imaging is best for the highest resolutions in order to simplify the manufacture of the photomasks.

Scanners replaced proximity printing by the mid-70s for device geometries below 4 to 5 μm. By the early 1980s, steppers began to dominate as device designs pushed to 2 μm and below. Steppers have continued to dominate lithographic patterning throughout the 1990s as minimum feature sizes reached the 250-nm levels. However, by the early 1990s a hybrid *step-and-scan* approach was introduced by SVG Lithography, the successor to Perkin-Elmer. The step-and-scan approach uses a fraction of a normal stepper field (for example, 26×8 mm), then scans this field in one direction to expose the entire 4× reduction mask (Figure 1.18). The wafer is then stepped to a new location and the scan is repeated. The smaller imaging field simplifies the design and manufacture of the lens, but at the expense of a more complicated reticle and wafer stage. Step-and-scan technology is the technology of choice today for below 250-nm manufacturing.

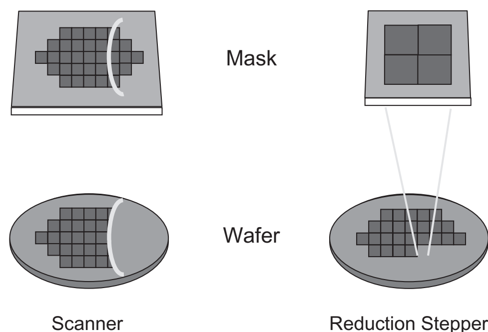


Figure 1.17 *Scanners and steppers use different techniques for exposing a large wafer with a small image field*

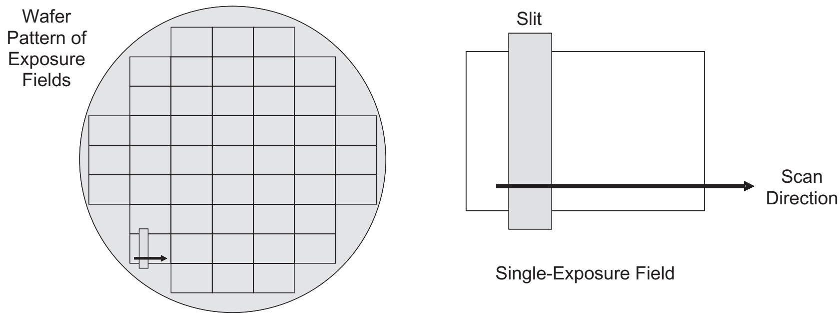


Figure 1.18 *In step-and-scan imaging, the field is exposed by scanning a slit that is about 25 × 8 mm across the exposure field*

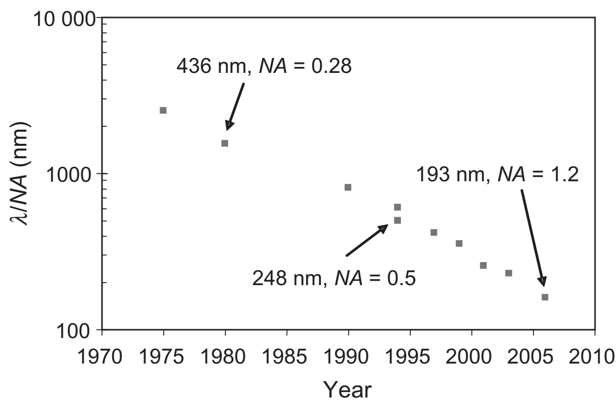


Figure 1.19 *The progression of* λ/NA *of lithographic tools over time (year of first commercial tool shipment)*

Resolution, the smallest feature that can be printed with adequate control, has two basic limits: the smallest image that can be projected onto the wafer, and the resolving capability of the photoresist to make use of that image. From the projection imaging side, resolution is determined by the wavelength of the imaging light (λ) and the numerical aperture (NA) of the projection lens according to the Rayleigh resolution criterion:

$$Resolution \propto \frac{\lambda}{NA} \quad (1.7)$$

Lithography systems have progressed from blue wavelengths (436 nm) to UV (365 nm) to deep-UV (248 nm) to today's mainstream high-resolution wavelength of 193 nm (see Figure 1.19 and Table 1.2). In the meantime, projection tool numerical apertures have risen from 0.16 for the first scanners to amazingly high 0.93 NA systems producing features well under 100 nm in size. In addition, immersion lithography, where the bottom of

Table 1.2 *The change in projection tool specifications over time*

	First Stepper (1978)	Immersion Scanner (2006)
Wavelength	436 nm	193 nm
Numerical Aperture	0.28	1.2
Field Size	10 × 10 mm	26 × 33 mm
Reduction Ratio	10	4
Wafer Size	4″ (100 mm)	300 mm
Throughput	20 wafers per hour (0.44 cm^2/s)	120 wafers per hour (24 cm^2/s)

the lens is immersed in a high refractive index fluid such as water, enables numerical apertures greater than one, with the first such 'hyper NA' tools available in 2006.

The main imaging lens of a stepper or scanner is the most demanding application of commercial lens design and fabrication today. The needs of microlithographic lenses are driving advances in lens design software, spherical and aspherical lens manufacturing, glass production and lens metrology. There are three competing requirements of lithographic lens performance – higher resolution, large field size and improved image quality (lower aberrations). Providing for any two of these requirements is rather straightforward (for example, a microscope objective has high resolution and good image quality but over a very small field). Accomplishing all three means advancing the state-of-the-art in optics. The first stepper in 1978 employed an imaging wavelength of 436 nm (the g-line of the mercury spectrum), a lens numerical aperture of 0.28 and a field size of 14 mm in diameter. Today's tools use an ArF excimer laser at 193 nm, a lens with a numerical aperture of 0.93 dry, and up to 1.35 with water immersion, and a field size of 26 × 33 mm. The 'hyper-NA' lens systems (NA > 1) are catadioptric, employing both mirrors and refractive lenses in the optical system. As might be expected, these modern high-performance imaging systems are incredibly complex and costly.

Before the exposure of the photoresist with an image of the mask can begin, this image must be aligned with the previously defined patterns on the wafer. This alignment process, and the resulting overlay of the two or more lithographic patterns, is critical since tighter overlay control means circuit features can be packed closer together. Closer packing of devices through better overlay is nearly as critical as smaller devices through higher resolution in the drive toward more functionality per chip. Along with alignment, wafer focus is measured at several points so that each exposure field is leveled and brought into proper focus.

Another important aspect of photoresist exposure is the *standing wave* effect. Monochromatic light, when projected onto a wafer, strikes the photoresist surface over a range of angles, approximating plane waves. This light travels down through the photoresist and, if the substrate is reflective, is reflected back up through the resist. The incoming and reflected light waves interfere to form a standing wave pattern of high and low light intensity at different depths in the photoresist. This pattern is replicated in the photoresist, causing ridges in the sidewalls of the resist feature as seen in Figure 1.20. As pattern dimensions become smaller, these ridges can significantly affect the quality of the feature. The interference that causes standing waves also results in a phenomenon called *swing curves*, the sinusoidal variation in linewidth with changing resist thickness. These detri-

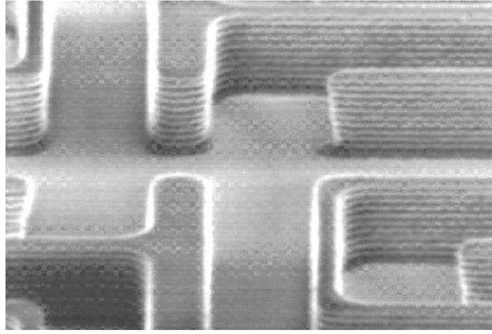


Figure 1.20 *Photoresist pattern on a silicon substrate (i-line exposure pictured here) showing prominent standing waves*

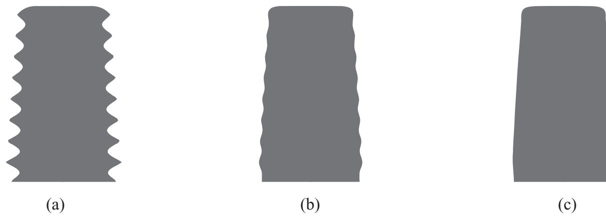


Figure 1.21 *Diffusion during a post-exposure bake is often used to reduce standing waves. Photoresist profile simulations as a function of the PEB diffusion length: (a) 20 nm, (b) 40 nm and (c) 60 nm*

mental effects are best cured by coating the substrate with a thin absorbing layer called a *bottom antireflective coating* (BARC) that can reduce the reflectivity of the substrate as seen by the photoresist to much less than 1 %.

1.3.5 Post-exposure bake

One method of reducing the standing wave effect is called the post-exposure bake (PEB).[17] The high temperatures used (100–130 °C) cause diffusion of the photoactive compound, thus smoothing out the standing wave ridges (Figure 1.21). It is important to note that the detrimental effects of high temperatures on photoresist, as discussed above concerning PAB, also apply to the PEB. Thus, it becomes very important to optimize the bake conditions. Also, the rate of diffusion of the exposure products is dependent on the PAB conditions – the presence of solvent enhances diffusion during a PEB. Thus, a low-temperature post-apply bake results in greater diffusion for a given PEB temperature.

For a conventional resist, the main importance of the PEB is diffusion to remove standing waves. For another class of photoresists, called chemically amplified resists, the PEB is an essential part of the chemical reactions that create a solubility differential between exposed and unexposed parts of the resist. For these resists, exposure generates a small amount of a strong acid that does not itself change the solubility of the resist. During the

post-exposure bake, this photogenerated acid catalyzes a reaction that changes the solubility of the polymer resin in the resist. Control of the PEB is extremely critical for chemically amplified resists.

1.3.6 Development

Once exposed, the photoresist must be developed. Most commonly used photoresists employ aqueous bases as developers. In particular, tetramethyl ammonium hydroxide (TMAH) is used in concentrations of 0.2–0.26 N. Development is undoubtedly one of the most critical steps in the photoresist process. The characteristics of the resist–developer interactions determine to a large extent the shape of the photoresist profile and, more importantly, the linewidth control.

The method of applying developer to the photoresist is important in controlling the development uniformity and process latitude. In the past, batch development was the predominant development technique. A boat of some 10–20 wafers or more is developed simultaneously in a large beaker, usually with some form of agitation. With the push toward in-line processing in the late 1970s, however, other methods have become prevalent. During *spin development* wafers are spun, using equipment similar to that used for spin coating, and developer is poured onto the rotating wafer. The wafer is also rinsed and dried while still spinning. *Spray development* uses a process identical to spin development except the developer is sprayed, rather than poured, on the wafer by a nozzle that produces a fine mist of developer over the wafer (Figure 1.22). This technique reduces developer usage significantly and gives more uniform developer coverage.

Another in-line development strategy is called *puddle development*. Again using developers specifically formulated for this process, the developer is poured onto a slowly spinning wafer that is then stopped and allowed to sit motionless for the duration of the development time. The wafer is then spin rinsed and dried. Note that all three in-line processes can be performed in the same piece of equipment with only minor modifications, and combinations of spray and puddle techniques are frequently used. Puddle development has the advantage of minimizing developer usage but can suffer from developer depletion – clear regions (where most of the resist is being dissolved) result in excessive dissolved resist in the developer, which depletes the developer and slows down development in these clear regions relative to dark regions (where most of the resist

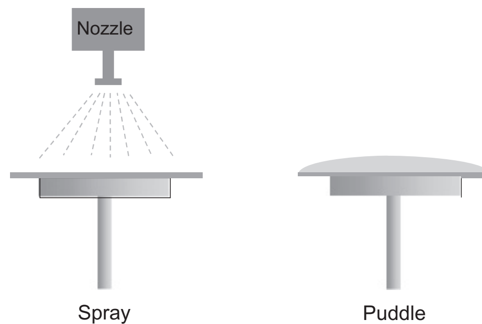


Figure 1.22 *Different developer application techniques are commonly used*

remains on the wafer). When this happens, the development cycle is often broken up into two shorter applications of the puddle in what is called a double-puddle process.

1.3.7 Postbake

The postbake (not to be confused with the post-exposure bake that comes before development) is used to harden the final resist image so that it will withstand the harsh environments of implantation or etching. The high temperatures used (120–150 °C) will cross-link the resin polymer in the photoresist, thus making the image more thermally stable. If the temperature used is too high, the resist will flow causing degradation of the image. The temperature at which flow begins is essentially equal to the glass transition temperature of the resist and is a measure of its thermal stability. In addition to cross-linking, the postbake can remove residual solvent, water, and gasses, and will usually improve adhesion of the resist to the substrate. Removal of these volatile components makes the resist more vacuum compatible, an important consideration for ion implantation.

Other methods are also used to harden a photoresist image. Exposure to high intensity deep-UV light cross-links the resin at the surface of the resist forming a tough skin around the pattern. Deep-UV hardened photoresist can withstand temperatures in excess of 200 °C without dimensional deformation. Plasma treatments and electron beam bombardment have also been shown to effectively harden photoresist. Commercial deep-UV hardening systems are available and widely used. Most of these hardening techniques are used simultaneously with high-temperature baking.

1.3.8 Measure and Inspect

Either before or after the postbake step, some sample of the resist patterns are inspected and measured for quality control purposes. Critical features and test patterns are measured to determine their dimensions (called a *critical dimension*, CD) and the overlay of the patterns with respect to previous lithographically defined layers. Wafers can also be inspected for the presence of random defects (such as particles) that may interfere with the subsequent pattern transfer step. This step of measurement and inspection is called ADI, *after develop inspect*, as opposed to measurements taken after pattern transfer, which are called *final inspect* (FI).

Inspection and measurement of the wafers before pattern transfer offer a unique opportunity: wafers (or entire lots) that do not meet CD or overlay specifications can be *reworked*. When a wafer is reworked, the patterned resist is stripped off and the wafers are sent back to the beginning of the lithography process. Wafers that fail to meet specifications at FI (after pattern transfer is complete) cannot be reworked and must be scrapped instead. Since reworking a wafer is far more cost-beneficial than scrapping a wafer, significant effort is put into verifying the quality of wafers at ADI and reworking any wafers that have potential lithography-limited yield problems.

1.3.9 Pattern Transfer

After the small patterns have been lithographically printed in photoresist, these patterns must be transferred into the substrate. As discussed in section 1.1.1, there are three basic pattern transfer approaches: subtractive transfer (etching), additive transfer (selective

deposition) and impurity doping (ion implantation). Etching is the most common pattern transfer approach. A uniform layer of the material to be patterned is deposited on the substrate. Lithography is then performed such that the areas to be etched are left unprotected (uncovered) by the photoresist. Etching is performed either using wet chemicals such as acids, or more commonly in a dry plasma environment. The photoresist 'resists' the etchant and protects the material covered by the resist. When the etching is complete, the resist is stripped leaving the desired pattern etched into the deposited layer. Additive processes are used whenever workable etching processes are not available, for example, for copper interconnects (copper does not form volatile etching by-products, and so is very difficult to etch in a plasma). Here, the lithographic pattern is used to open areas where the new layer is to be grown (by electroplating, in the case of copper). Stripping of the resist then leaves the new material in a negative version of the patterned photoresist. Finally, doping involves the addition of controlled amounts of contaminants that change the conductive properties of a semiconductor. Ion implantation uses a beam of dopant ions accelerated at the photoresist-patterned substrate. The resist blocks the ions, but the areas uncovered by resists are embedded with ions, creating the selectively doped regions that make up the electrical heart of the transistors. For this application, the 'stopping power' of the resist (the minimum thickness of resist required to prevent ions from passing through) is the parameter of interest.

1.3.10 Strip

After the imaged wafer has been pattern transferred (e.g. etched, ion implanted, etc.), the remaining photoresist must be removed. There are two classes of resist stripping techniques: wet stripping using organic or inorganic solutions, and dry (plasma) stripping. A simple example of an organic stripper is acetone. Although commonly used in laboratory environments, acetone tends to leave residues on the wafer (scumming) and is thus unacceptable for semiconductor processing. Most commercial organic strippers are phenol-based and are somewhat better at avoiding scum formation. However, the most common wet strippers for positive photoresists are inorganic acid-based systems used at elevated temperatures.

Wet stripping has several inherent problems. Although the proper choice of strippers for various applications can usually eliminate gross scumming, it is almost impossible to remove the final monolayer of photoresist from the wafer by wet chemical means. It is often necessary to follow a wet strip by a plasma 'descum' step to completely clean the wafer of resist residues.[18] Also, photoresist which has undergone extensive hardening (e.g. deep-UV hardening) and been subjected to harsh processing conditions (e.g. high-energy ion implantation) can be almost impossible to strip chemically. For these reasons, plasma stripping has become the standard in semiconductor processing. An oxygen plasma is highly reactive toward organic polymers but leaves most inorganic materials (such as are mostly found under the photoresist) untouched.

Problems

1.1. When etching an oxide contact hole with a given process, the etch selectivity compared to photoresist is found to be 2.5. If the oxide thickness to be etched is 140 nm

and a 50 % overetch is used (that is, the etch time is set to be 50 % longer than that required to just etch through the nominal oxide thickness), what is the minimum possible photoresist thickness (that is, how much resist will be etched away)? For what reasons would you want the resist to be thicker than this minimum?

1.2. For a certain process, a 300-keV phosphorous implant is masked well by a 1.0-μm-thick photoresist film.
   (a) For this implant process, how many multiples of the straggle does this resist thickness represent?
   (b) If the implant energy is increased to 450 keV, how much should the photoresist thickness be increased?
   (c) If the dopant is also changed to arsenic (with the energy at 450 keV), what resist thickness will be needed?

1.3. A photoresist gives a final resist thickness of 320 nm when spun at 2800 rpm.
   (a) What spin speed should be used if a 290-nm-thick coating of this same resist is desired?
   (b) If the maximum practical spin speed for 200-mm wafers is 4000 rpm, at what thickness would a lower viscosity formulation of the resist be required?

1.4. Resolution in optical lithography scales with wavelength and numerical aperture according to a modified Rayleigh criterion:

$$R = k_1 \frac{\lambda}{NA}$$

where k_1 can be thought of as a constant for a given lithographic approach and process. Assuming $k_1 = 0.35$, plot resolution versus numerical aperture over a range of NAs from 0.5 to 1.0 for the common lithographic wavelengths of 436, 365, 248, 193 and 157 nm. From this list, what options (NA and wavelength) are available for printing 90-nm features?

References

1 Glawischnig, H. and Parks, C.C., 1996, SIMS and modeling of ion implants into photoresist, *Proceedings of the 11th International Conference on Ion Implantation Technology*, 579–582.
2 Wong, A.K., Ferguson, R., Mansfield, S., Molless, A., Samuels, D., Schuster, R. and Thomas, A., 2000, Level-specific lithography optimization for 1-Gb DRAM, *IEEE Transactions on Semiconductor Manufacturing*, **13**, 76–87.
3 Moore, G.E., 1965, Cramming more components onto integrated circuits, *Electronics*, **38**, 114–117.
4 Moore, G.E., 1975, Progress in digital integrated electronics, *IEDM Technical Digest*, **21**, 11–13.
5 Moore, G.E., 1995, Lithography and the future of Moore's law, *Proceedings of SPIE: Optical/Laser Microlithography VIII*, **2440**, 2–17.
6 *The National Technology Roadmap for Semiconductors*, 1994, Semiconductor Industry Association (San Jose, CA).
7 Noyce, R., 1977, Microelectronics, *Scientific American*, **237**, 63–69.
8 Tobey, A.C., 1979, Wafer stepper steps up yield and resolution in IC lithography, *Electronics*, **52**, 109–112.
9 Lyman, J., 1985, Optical lithography refuses to die, *Electronics*, **58**, 36–38.

10 Collins, R.H. and Deverse, F.T., 1970, U.S. Patent No. 3,549,368.
11 Levinson, H.J., 2005, *Principles of Lithography*, second edition, SPIE Press (Bellingham, WA), p. 59.
12 Meyerhofer, D., 1978, Characteristics of resist films produced by spinning, *Journal of Applied Physics*, **49**, 3993–3997.
13 Bornside, D.E., Macosko, C.W. and Scriven, L.E., 1989, Spin coating: one-dimensional model, *Journal of Applied Physics*, **66**, 5185–5193.
14 Kobayashi, R., 1994, 1994 review: laminar-to-turbulent transition of three-dimensional boundary layers on rotating bodies, *Journal of Fluids Engineering*, **116**, 200–211.
15 Gregory, N., Stuart, J.T. and Walker, W.S., 1955, On the stability of three-dimensional boundary layers with application to the flow due to a rotating disk, *Philosophical Transactions of the Royal Society of London Series A*, **248**, 155–199.
16 Markle, D.A., 1974, A new projection printer, *Solid State Technology*, **17**, 50–53.
17 Walker, E.J., 1975, Reduction of photoresist standing-wave effects by post-exposure bake, *IEEE Transactions on Electron Devices*, **ED-22**, 464–466.
18 Kaplan, L.H. and Bergin, B.K., 1980, Residues from wet processing of positive resists, *Journal of The Electrochemical Society*, **127**, 386–395.