1

Lithography

Miniaturization science is the science of making very small things. In top-down micro- and nanomachining, this is done by building down from bigger chunks of material; in bottom-up nanochemistry, it is by building up from smaller building blocks. Both require an excellent understanding of the intended application, different manufacturing options, materials properties, and scaling laws. The resulting three-dimensional structures, ranging in size from subcentimeters to subnanometers, include sensors, actuators, microcomponents, and microsystems.

Why must our bodies be so large compared with the atom?

Erwin Schrödinger, *What Is Life?*

Discovery consists of looking at the same thing as everyone else and thinking something different.

Albert-Szent Györgi

Introduction

Many top-down miniaturization methods start with lithography, the technique used to transfer copies of a master pattern onto the surface of a solid material such as a silicon wafer. In this chapter, we review different forms of lithography, detailing those that differ most from the miniaturization processes used to fashion integrated circuits (ICs).

A short historical note about the origins of lithography is followed by a description of photolithography, including developments that have allowed the printing of the ever-shrinking features of modern ICs. After reviewing the limits of photolithography, we detail alternatives, including x-ray and charged-particle (electron and ion) lithographies, followed by promising techniques in the early research and development (R&D) stage.

Historical Note: Lithography's Origins

After experimenting with various resins in sunlight, Nicéphore Niépce managed to copy an etched print on oiled paper by placing it over a glass plate coated with bitumen (asphalt) dissolved in lavender oil (France, 1822). After two or three hours of sunlight, the unshaded areas in the bitumen became hard as compared to the shaded areas, which remained more soluble and could be washed away with a mixture of turpentine and lavender oil. Niépce's concoction, we will learn below, corresponds to a negative-type photoresist. Five years later, in 1827 (talk about fast turnaround time!), using strong acid, the Parisian engraver Lemaître made an etched copy of an engraving of Cardinal d'Amboise (Inset 1.1) from a plate developed by Niépce. The latter copy represents the earliest example of pattern transfer by photolithography and chemical milling. The accuracy of the technique was 0.5 to 1 mm.[1]

The word *lithography* (Greek for the words *stone* [*lithos*] and *to write* [*gráphein*]) refers to the process invented in 1796 by Aloys Senefelder. Senefelder found that stone (he used Bavarian limestone), when properly inked and treated with chemicals, could transfer a carved image onto paper. Due to the chemical treatment of the stone, image and nonimage areas became oil receptive (water repellent) and oil repellent (water receptive), respectively, attracting ink onto the image area and attracting water on nonimage areas.[2]

The Niépce process heralded the advent of photography. Much later, photomasking, followed by chemical processing, led to the photolithography now used in fabricating ICs and in miniaturization science. Not until World War II, more than 100 years after Niépce and Lemaître, did the first applications of the printed circuit board come about. Interconnections were made by soldering separate electronic components to a pattern of "wires" produced by photoetching a layer of copper foil that
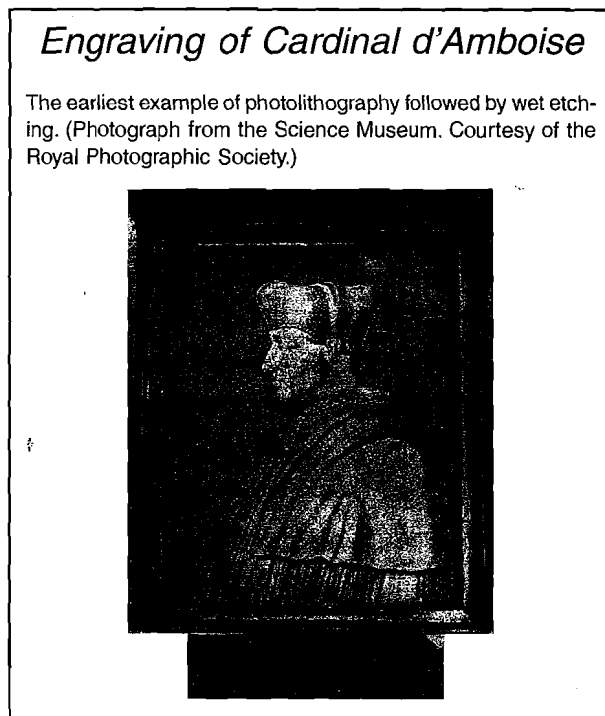
was laminated to a plastic board. By 1961, methods were devised whereby a photoetching process produced large numbers of transistors on a thin slice of silicon (Si). At that time, pattern resolution was no better than 5 μm.[1] Today, photolithography, x-ray lithography, and charged-particle lithography all achieve submicron printing accuracy. The Coppermine PIIIs chip from Intel (introduced in late 1999), for example, incorporates a 0.18 μm minimum feature size and is fabricated using UV photolithography.[3]

Photolithography Overview

Introduction

The most widely used form of lithography is photolithography. In the IC industry, pattern transfer from masks onto thin films is accomplished almost exclusively via photolithography. The combination of accurate registration and exposing a series of successive patterns leads to complex multilayered ICs. This essentially two-dimensional process has a limited tolerance for nonplanar topography, creating a major constraint for building non-IC miniaturized systems, which often exhibit extreme topographies. Photolithography has matured rapidly and constantly improved in its ability to resolve ever smaller features. For the IC industry, this continued improvement in resolution has impeded the adaptation of alternative, higher-resolution lithography techniques such as x-ray lithography. Research in high-aspect-ratio resist features, to satisfy the needs of both IC and non-IC miniaturization, is also finally improving photolithography's capacity to cover wide ranges of topography.

Photolithography and pattern transfer involve a set of process steps as summarized in Figure 1.1. As an example, we use an
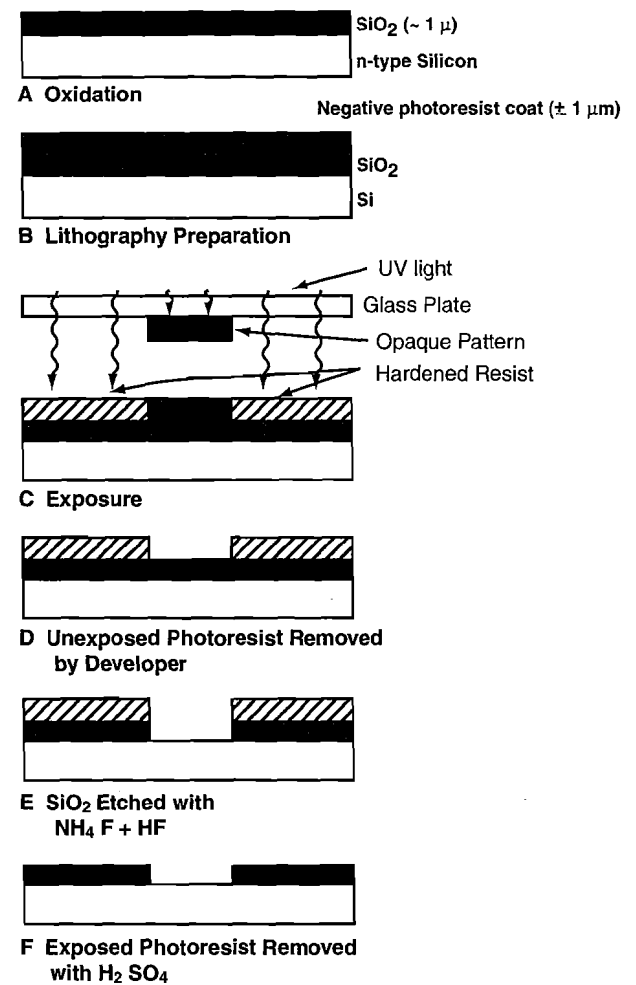

Figure 1.1   Basic photolithography and pattern transfer. Example uses an oxidized Si wafer and a negative photoresist system. Process steps include exposure, development, oxide etching, and resist stripping. Steps A through F are explained in the text.

oxidized Si wafer and a negative photoresist system. For simplicity, not all of the steps are detailed in this figure, as they will be covered in subsequent text. A short preview follows.

An oxidized wafer (A) is coated with a 1 μm thick negative photoresist layer (B). After exposure (C), the wafer is rinsed in a developing solution or sprayed with a spray developer, which removes the unexposed areas of photoresist and leaves a pattern of bare and photoresist-coated oxide on the wafer surface (D). The photoresist pattern is the negative image of the pattern on the photomask. In a typical next step after development, the wafer is placed in a solution of HF or HF + NH_4F, meant to attack the oxide but not the photoresist or the underlying silicon (E). The photoresist protects the oxide areas it covers. Once the exposed oxide has been etched away, 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 include organic solvent strippers and alkaline strippers (with or without oxidants). 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.

Masks

The stencil used to repeatedly generate a desired pattern on resist-coated wafers is called a *mask*. In use, a photomask—a nearly optically flat glass (transparent to near ultraviolet [UV]) or quartz plate (transparent to deep UV) with an absorber pattern metal (e.g., an 800-Å-thick chromium layer)—is placed in direct contact with the photoresist-coated surface, and the wafer is exposed to UV radiation. The absorber pattern on the photomask is opaque to UV light, whereas glass and quartz are transparent. The absorber pattern on the mask is generated by e-beam lithography, a technique that produces higher resolution than photolithography. A light field or dark field image, known as *mask polarity* (Inset 1.2), is then transferred to the semiconductor surface. This procedure results in a 1:1 image of the entire mask onto the silicon wafer.

The described masks, making direct physical contact (also referred to as *hard contact*) with the substrate, are called *contact masks*. Unfortunately, these masks degrade faster through wear than do noncontact, proximity masks (also referred to as *soft contact masks*), which are slightly raised, say 10 to 20 µm, above the wafer. The defects resulting from hard contact masks on both the wafer and the mask make this method of optical pattern transfer unsuitable for very large scale integration (VLSI) manufacturing. In VLSI, integrated circuits have between 100,000 and 1 million components. We review hard contact masks because they are still in use in R&D, in mask making itself, and for prototyping. Contact mask and proximity mask printing are collectively known as *shadow printing*. A more reliable method of masking is *projection printing* where, rather than placing a mask in direct contact with (or in proximity of) a wafer, the photomask is imaged by a high-resolution lens system onto the resist-coated wafer. In the latter case, the only limit to

the mask lifetime results from operator handling. The imaging lens can reduce the mask pattern by 1:5 or 1:10, making mask fabrication less challenging.
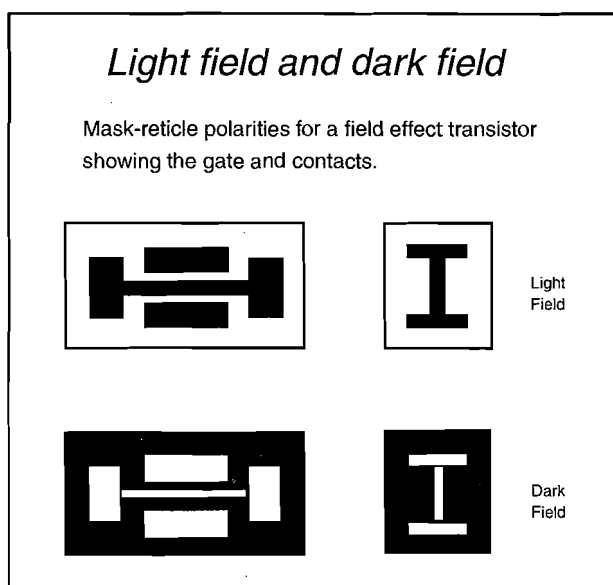
The design of electron beam generated masks for ICs and miniaturized machines is generally fairly straightforward and requires some suitable computer aided design (CAD) software and a platform on which to run it. Electron-beam lithography is discussed later in this chapter, and mask design and suitable CAD software will be addressed in Chapter 8. In miniaturization science, one often is looking for low-cost and fast-turnaround methods to fabricate masks. This may involve manually drawn patterns on cut-and-peel masking films and photo reduction, affording fast turnaround without relying on outside photomask services. Alternatively, it may involve direct writing on a photoresist-coated plate with a laser plotter (~2 µm resolution).[4] Simpler yet, using a drawing program such as Canvas (Deneba Systems, Inc.), Freehand (Macromedia, Inc.), Illustrator® (Adobe Systems Inc.), or L-Edit (Tanner Research, Inc.), a mask design can be created on a computer and saved as a PostScript® file to be printed with a high-resolution printer (say 4000 dpi) on a transparency.[5] The transparency may then be clamped between a presensitized chrome-covered mask plate (i.e., a vendor such as Nanofilm has preapplied the resist) and a blank plate. After exposure and development, the exposed plate is put in a chrome etch for a few minutes to generate the desired metal pattern, and the remaining resist is stripped off.
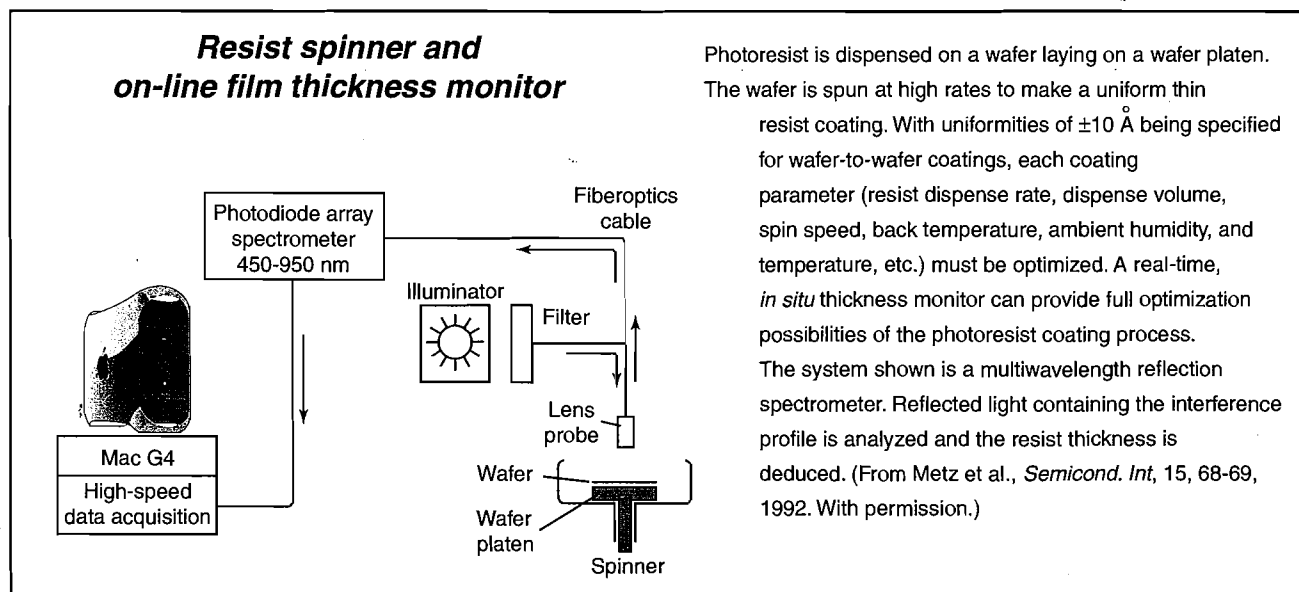
Spinning Resist and Soft Baking

A common step before spinning on a resist with Si as a substrate is to grow a thin layer of oxide on the wafer surface by heating it to between 900 and 1150°C in steam or in a humidified oxygen stream (see Figure 1.1A). Dry oxygen also works, but wet oxygen and steam produce faster results. The oxide can serve as a mask for a subsequent wet etch or boron implant. As the first step in the lithography process itself, a thin layer of an organic polymer, a photoresist sensitive to ultraviolet radiation, is deposited on the oxide surface (see Figure 1.1B). The photoresist is dispensed from a viscous solution of the polymer onto the wafer lying on a wafer platen in a resist spinner (Inset 1.3).[6] A vacuum chuck holds the wafer in place. The wafer is then spun at high speed, between 1500 and 8000 rotations per minute (rpm), depending on the viscosity and required film thickness, to make a uniform film. At these speeds, centrifugal force causes the solution to flow to the edges, where it builds up until expelled when surface tension is exceeded. The resulting polymer thickness, T, is a function of spin speed, solution concentration, and molecular weight (measured by intrinsic viscosity). The empirical expression for T is given by:

$$T = \frac{KC^\beta \eta^\gamma}{\omega^\alpha} \tag{1.1}$$

where K = overall calibration constant
       C = polymer concentration in g/100 mL solution
       η = intrinsic viscosity
       ω = rotations per minute (rpm)

Light field and dark field

Mask-reticle polarities for a field effect transistor showing the gate and contacts.

Light Field

Dark Field

Inset 1.2

**Resist spinner and
on-line film thickness monitor**

Photoresist is dispensed on a wafer laying on a wafer platen. The wafer is spun at high rates to make a uniform thin resist coating. With uniformities of ± 10 Å being specified for wafer-to-wafer coatings, each coating parameter (resist dispense rate, dispense volume, spin speed, back temperature, ambient humidity, and temperature, etc.) must be optimized. A real-time, *in situ* thickness monitor can provide full optimization possibilities of the photoresist coating process. The system shown is a multiwavelength reflection spectrometer. Reflected light containing the interference profile is analyzed and the resist thickness is deduced. (From Metz et al., *Semicond. Int*, 15, 68-69, 1992. With permission.)

Inset 1.3

Once the various exponential factors (α, β, and γ) have been determined, Equation 1.1 can be used to predict the thickness of the film that can be spun for various molecular weights and solution concentrations of a given polymer and solvent system.[7]

The spinning process is of primary importance to the effectiveness of pattern transfer. The quality of the resist coating determines the density of defects transferred to the device under construction. The resist film uniformity across a single substrate and from substrate to substrate must be ± 5 nm (for a 1.5 μm film, this is $\pm 0.3\%$) to ensure reproducible line widths and development times in subsequent steps. The coating thickness of the thin, glassy resist film depends on the chemical resistance required for image transfer and the fineness of the lines and spaces to be resolved. The application of too much resist results in edge covering or run-out, hillocks, and ridges, reducing manufacturing yield. Too little resist may leave uncovered areas. For silicon integrated circuits, the resist thickness after prebaking (see below) typically ranges between 0.5 and 2 μm. For miniaturized 3D structures, much greater resist thicknesses are often required. In the latter case, techniques such as casting, plasma polymerizing of the resist, and the use of thick sheets of dry photoresist replace the ineffective resist spinners. Layers 1 cm thick and above have been produced experimentally.[8] The challenges involved in making thicker resist coats for high-aspect-ratio miniaturized machinery will be discussed further in Chapter 6. Optimization of the "regular" photoresist coating process in terms of resist dispense rate, dispense volume, spin speed, ambient temperature, and humidity presents a growing challenge as a 1 μm resist thickness with a repeatability of ± 10 Å ($\pm 0.1\%$) is becoming the norm. The microelectronic industry in 2001 was dealing with feature sizes of 0.18 μm and below, and for these smaller feature sizes, control over thinner and thinner resist layers is needed. An on-line film thickness monitor, possibly a technique based on reflection spectroscopy (Inset 1.3), will become essential for statistical process control of such demanding photoresist coatings.[6]

After spin coating, the resist still contains up to 15% solvent and may contain built-in stresses. The wafers are therefore soft baked (also prebaked) at 75 to 100°C for, say, 10 min to remove solvents and stress and to promote adhesion of the resist layer to the wafer. The optimization of the prebaking step may substantially increase device yield.

Exposure and Postexposure Treatment

After soft baking, the resist-coated wafers are transferred to some type of illumination or exposure system where they are aligned with the features on the mask (see Figure 1.1C). For any lithographic technique to be of value, it must provide an alignment technique capable of a superposition precision of mask and wafer that is a small fraction of the minimum feature size. In the simplest case, an exposure system consists of a UV lamp illuminating the resist-coated wafer through a mask without any lenses between the two. The purpose of the illumination is to deliver light with the proper intensity, directionality, spectral characteristics, and uniformity across the wafer, allowing a nearly perfect transfer or printing of the mask image onto the resist in the form of a latent image.

In photolithography, wavelengths of the light source used for exposure of the resist-coated wafer range from the very short wavelengths of extreme ultraviolet (EUV) (10 to 14 nm) to deep ultraviolet (DUV) (150 to 300 nm) to near ultraviolet (UV) (350 to 500 nm). In near UV, one typically uses the g-line (435 nm) or i-line (365 nm) of a mercury lamp. The brightness of most shorter-wavelength sources is severely reduced in comparison to that of longer-wavelength sources, and the addition of lenses further reduces the efficiency of the exposure system. For example, the total collected DUV power for a 1 kW mercury-xenon lamp in the 200 to 250 nm range is only 30 to 40 mW; the additional optics absorb more energy of the short wavelengths passing through them. As a conse-

quence, with shorter wavelengths, higher resist sensitivity is required, and newer DUV sources that produce a higher flux of DUV radiation must be used. For example, a KrF excimer laser with a short wavelength of 248 nm and a power of 10 to 20 W at that wavelength is an option. In general, the smallest feature that can be printed using projection lithography is roughly equal to the wavelength of the exposure source, in this example, 248 nm would be expected. The same laser, in combination with some sophisticated resolution-enhancing techniques (RETs), is used to produce the most advanced circuits with transistor gate features as small as 160 nm. RET methods (see below) enable one to go quite a bit beyond the conventional Rayleigh diffraction limit. Other candidate exposure systems of the future, but not yet out of the laboratory, include two DUV excimer lasers: the ArF at 193 nm and the F_2 at 157 nm. The consensus candidate for the next generation of lithography is photolithography using 193 nm light[9] from ArF lasers. In the case of EUV, a plasma or synchrotron source and reflective optics are used. Refractive optical elements are too absorbing at those wavelengths.[10] The prospect of using EUV as a commercial tool is still in the early research phase and is discussed below under *Next-Generation Lithographies* (page 48).

The incident light intensity (in W/cm^2) multiplied by the exposure time (in seconds) gives the incident energy (J/cm^2) or dose, D, across the surface of the resist film. Radiation induces a chemical reaction in the exposed areas of the photoresist, altering the solubility of the resist in a solvent either directly or indirectly via a sensitizer.

During the latent-image-forming reaction, the sensitizer in the resist usually *bleaches;* in other words, exposed resist is rendered transparent to the incoming wavelength. This bleaching allows the use of thick films with high absorbency, since light will reach the substrate through the bleached resist. The absorbency of the unexposed resist should not reach 40% so as to avoid degradation of the image profile through the resist depth, as too large a percentage of the light is absorbed in the top layer. On the other hand, with the absorbency far below 40%, exposure times required to form the image become too long. The smaller the dose needed to "write" or "print" the mask features onto the resist layer with good resolution, the better the lithographic sensitivity of the resist.

Postexposure treatment is often desired, because the reactions initiated during exposure might not have run to completion. To halt the reactions or to induce new ones, several postexposure treatments are in use: postexposure baking, flood exposure with other types of radiation, treatment with reactive gas, and vacuum treatment. Postexposure baking (sometimes in a vacuum) and treatment with reactive gas are used in image reversal and dry resist development. In the case of a chemically amplified resist, the postexposure bake is most critical. Although reactions induced by the catalyst that forms during exposure take place at room temperature, their rate is highly increased by baking at 100°C. The precise control of this type of postexposure bake critically determines the subsequent development itself. Image reversal, dry resist development, and chemically amplified resists will be treated below.

Development

Development transforms the latent resist image formed during exposure into a relief image that will serve as a mask for further subtractive and additive steps (see Chapters 2 and 3, respectively). During the development process, selective dissolving of resist takes place (see Figure 1.1D). Two main technologies are available for development: wet development, widely used in circuit and miniaturization manufacture in general, and dry development, which is starting to replace wet development for some of the ultimate line width resolution applications.

Wet development by solvents may be based on at least three different types of exposure-induced changes: variation in molecular weight of the polymers (by cross-linking or by chain scission), reactivity change, and polarity change.[11] Two main types of wet development setups are used: immersion and spray (Inset 1.4). During immersion developing, cassette-loaded wafers are batch-immersed for a timed period in a developer bath and agitated at a specific temperature.

During spray development, fan-type sprayers direct fresh developing solution across wafer surfaces. Positive resists are typically developed in aqueous alkaline solutions, and negative resists in organic ones. Aqueous development is highly favored for health reasons. The aqueous development rate depends on the pH of the developer and the temperature, which needs to be controlled to within ±0.5°C.[7] Surfactants and other wetting agents added to the developer ensure uniform wetting, and buffers provide a more stable operating window and a longer lifetime. The newer negative resist formulations also may be developed in aqueous solvents.

The use of organic solvents leads to some swelling of the resist (especially for negative resists; see below) and loss of adhesion

Spray developer

Fresh developing solution is directed across wafer surfaces by a fan-type spraying nozzle. The renewal of developer allows a uniform bath strength to be maintained.
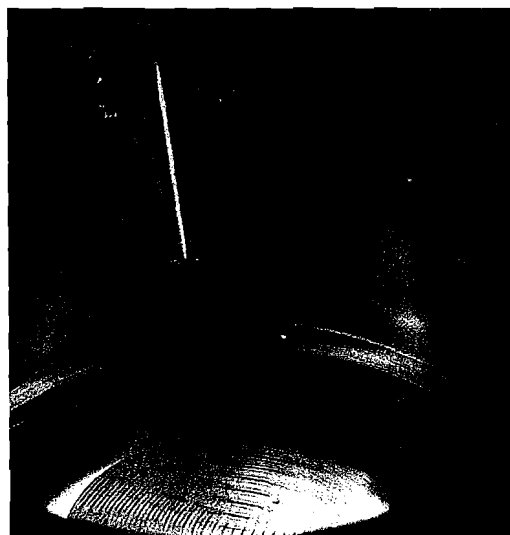
Inset 1.4

of the resist to the substrate. Dry development overcomes these problems, as it is based either on a vapor phase process or a plasma.[12] In the latter, oxygen-reactive ion etching (O_2-RIE) is used to develop the latent image. The image formed during exposure exhibits a differential etch rate to O_2-RIE rather than differential solubility to a solvent.[7] Dry developed resists should not be confused with dry film resists, which are resists that come in film form and are laminated onto a substrate rather than spin coated. Dry-developed resists, such as the DESIRE process, where the surface of the exposed resist is treated with a silicon-containing reagent, will be discussed below.

With continued pressure by the U.S. Environmental Protection Agency (EPA) for a cleaner environment, dry development as well as dry etching (see Chapter 2) are bound to get increasing attention.

De-scumming and Postbaking

A mild oxygen plasma treatment, so-called *de-scumming*, removes unwanted resist left behind after development. Negative, and to a lesser degree positive, resists leave a thin polymer film at the resist/substrate interface. The problem is most severe in small (<1 μm) high-aspect-ratio structures where the mass transfer of a wet developer is poor. Patterned resist areas are also thinned in the de-scumming process, but this is usually of little consequence.

Before etching the substrate or adding a material, the wafer must be postbaked. Postbaking or hard baking removes residual solvents and anneals the film to promote interfacial adhesion of the resist that has been weakened either by developer penetration along the resist/substrate interface or by swelling of the resist (mainly for negative resists). Hard baking also improves the hardness of the film. Improved hardness increases the resistance of the resist to subsequent etching steps. Postbaking frequently occurs at higher temperatures (120°C) and for longer times (say 20 min) than soft or prebaking. The major limitation for heat application is excessive flow or melt, which degrades wall profile angles and makes it more difficult to remove the resist. Special care needs to be taken with the baking temperature above the glass transition temperature, T_g, at which impurities are easily incorporated into the resist. Positive resists withstand higher heating temperatures than negative resists, but their stripping proves more difficult. De-scumming and postbaking both follow step (D) in Figure 1.1.

Resist does not withstand long exposure to etchants well. As a consequence, with 1:7 buffered HF (BHF) (a mixture of one part 49% aqueous HF-solution and seven parts NH_4F that is used to strip SiO_2), the postbake sometimes is repeated after 5 min of etching to prolong the lifetime of the resist layer (see Figure 1.1E). Also, postbaking should be prolonged before electroplating.

Resists

The principal components of photoresists are a polymer (base resin), a sensitizer, and a casting solvent. The polymer changes structure when exposed to radiation; the solvent allows spin application and formation of thin layers on the wafer surface; sensitizers control the photochemical reactions in the polymeric phase. Resists without sensitizers are single-component or one-component systems, whereas sensitizer-based resists are two-component systems. Solvent and other potential additives do not directly relate to the photoactivity of the resist.

Resist Tone

If the photoresist is of the type called *positive* (also *positive tone*), the photochemical reaction during exposure of a resist typically weakens the polymer by rupture or scission of the main and side polymer chains, and the exposed resist becomes more soluble in developing solutions (say ten times more soluble). In other words, the development rate, R, for the exposed resist is about ten times faster than the development rate, R_0, for the unexposed resist. If the photoresist is of the type called *negative* (also *negative tone*), the reaction strengthens the polymer by random cross-linkage of main chains or pendant side chains, becoming less soluble (slower dissolving). Exposure, development, and pattern-transfer sequences for negative and positive resists are shown in Figure 1.2.

Positive Resists

Two well-known families of positive photoresists are the single-component poly(methylmethacrylate) (PMMA) (Inset 1.5) resists and the two-component DQN (Inset 1.6) resists composed of a photoactive diazoquinone ester (DQ) (20 to 50 wt%) and a phenolic novolak resin (N).
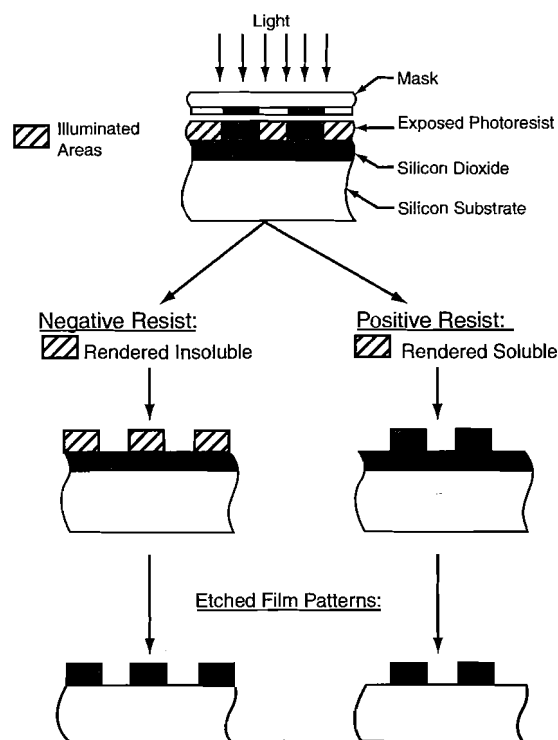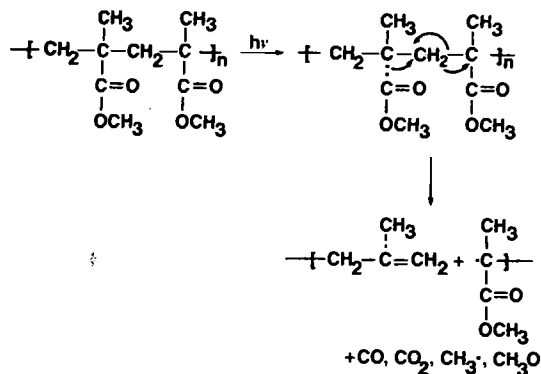

Figure 1.2   Positive and negative resist: exposure, development, and pattern transfer. Positive resists develop in the exposed region. Negative resists remain in the exposed region.

Poly(methylmethacrylate) or PMMA

Photo-induced chain scission of PMMA resist.

Inset 1.5

Diazoquinone ester (DQ) and phenolic novolak resin (N), i.e., DQN

The novolak matrix resin (N) is prepared by acid copolymerization of cresol and formaldehyde. The base insoluble sensitizer, a diazoquinone (DQ), undergoes photolysis to produce a carbene which then undergoes a rearrangement to form a ketene. The ketene reacting with water present in the film forms a base-soluble, indenecarboxylic acid photoproduct.
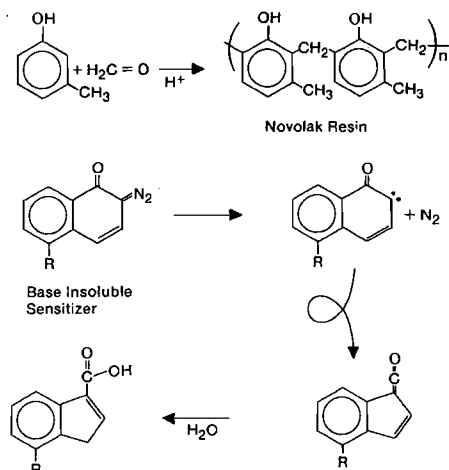
Inset 1.6

PMMA becomes soluble through chain scission under DUV illumination. The maximum sensitivity of PMMA lies at 220 nm; above 240 nm, the resist becomes insensitive. PMMA resin by itself constitutes a rather insensitive or slow DUV photoresist, requiring doses >250 mJ/cm^2. Exposure times of tens of minutes were required with the earliest DUV PMMAs available.[12] By adding a photosensitizer such as t-butyl benzoic acid, the UV spectral absorbency of PMMA is increased, and a 150 mJ/cm^2

lithographic sensitivity can be obtained. PMMA is also used in electron beam, ion beam, and x-ray lithography.

The DQN system is a "workhorse," near-UV, two-component positive resist, which photochemically transforms into a polar, base-soluble product.[9] The hydrophilic novolak resin (N) is in itself alkali soluble but is rendered insoluble by the addition of 20 to 50 wt% DQ, which forms a complex with the phenol groups of the novolak resin. The resist is rendered soluble again through the photochemical reaction of DQ. The matrix novolak resin is a copolymer of a phenol and formaldehyde. A novolak resin absorbs light below 300 nm, and the DQ addition adds an absorption region around 400 nm. The 365, 405, and 435 nm mercury lines can all be used for exposure of DNQ. The intense absorption of aromatic molecules prevents the use of this resist at exposing wavelengths less than about 300 nm; at those shorter wavelengths, linear acrylate and methacrylate copolymers have the advantage.

Most positive resists are soluble in strongly alkaline solutions (a fact that is taken advantage of for stripping of the resist—see Figure 1.1F and text below) and develop in mildly alkaline ones (as shown in Figure 1.1D). Some typical industrial developers for positive resists are KOH (e.g., a 0.05 to 0.5 N aqueous solution and a surfactant), tetramethylammonium hydroxide (TMAH), ketones, and acetates. As mentioned above, besides changing the molecular weight of the resist, radiation-induced reactions may also change the resist's solubility by altering its hydrophilicity (polarity) or its reactivity. Typical casting solvents for positive resists are Cellosolve* acetate, methyl Cellosolve, and aromatic hydrocarbons.

Negative Resists

The first negative photoresists were based on free-radical-initiated photo-cross-linking processes of main or pendant polymer side chains rendering the exposed parts insoluble. They were the very first types of resists used to pattern semiconductor devices and still constitute the largest segment of the photoresist industry, being widely used to define circuitry in printed wiring boards (PWBs).[13] As illustrated in Figure 1.3, a negative photoresist becomes insoluble in organic (more traditional negative resists) or water-based developers (newer negative resist systems) upon exposure to UV radiation. The insoluble layer forms a "negative" pattern that is used as a stencil (usually temporarily) to delineate many levels of circuitry in semiconductors, microelectromechanical systems (MEMS), and printed wiring boards (PWBs). The insolubilization of radiated negative resists can be achieved in one of two ways: the negative resist material increases in molecular weight through UV exposure (traditional negative resists), or it is photochemically transformed to form new insoluble products (newer negative resist products). The increase in molecular weight is generally accomplished through photoinitiators that generate free radicals or strong acids facilitating polymeric cross-linking or the photopolymerization of monomeric or oligomeric species. Photochemical transformation of negative photoresists may also lead to hydrophobic or

---

* Cellosolve is a trade name for solvents based on esters of ethylene glycol; these solvents have been identified as possible carcinogens.
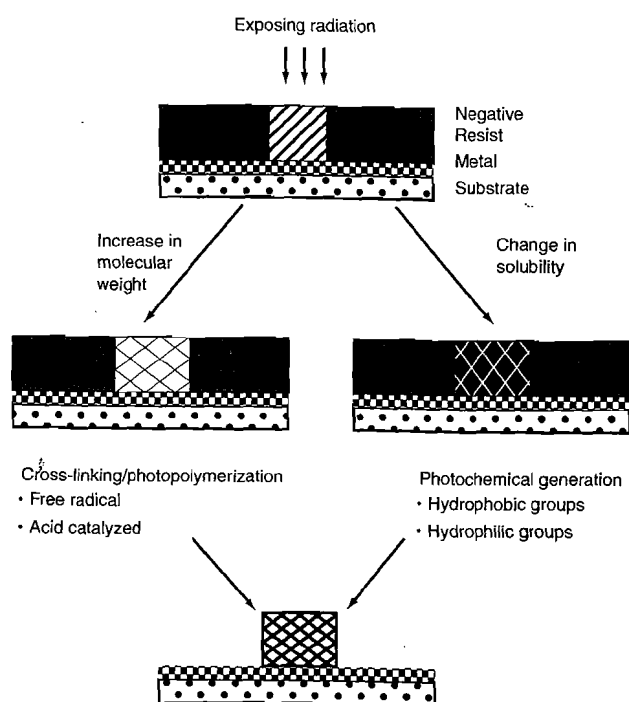
Figure 1.3 Negative resist system. Exposure induces different types of changes in the photopolymer.

Inset 1.7

hydrophilic groups, which provide another means of inducing preferential solubility between the exposed and unexposed resist film.

Commonly used negative-acting, two-component resists are bis(aryl)azide rubber resists (Inset 1.7), whose matrix resin is cyclized poly(cis-isoprene), a synthetic rubber. Bis(aryl)azide sensitizers lose nitrogen and generate a highly reactive nitrene upon photolysis. The nitrene intermediate undergoes a series of reactions that results in the cross-linking of the resin. Oxidation, with oxygen from the ambient or dissolved in the polymer, often is a competing reaction for polymerization. In other words, polymerization can be inhibited by the quenching of the cross-linking reactions through scavenging of the nitrene photoproduct by oxygen. This competing reaction represents a disadvantage, as exposure has to be carried out under a nitrogen blanket or in a vacuum. Another disadvantage of this type of negative resist is that the resolution is limited by film thickness. The cross-linking process starts topside, where the light hits the resist first. Consequently, overexposure is needed to render the resist insoluble at the substrate interface. The thicker the resist wanted, the greater the overdose needed for complete polymerization and the larger the scattered radiation. Scattered radiation at the resist/substrate interface in turn reduces the obtainable resolution. Moreover, the organic solvent developer swells the cross-linked negative image, which further degrades the resolution. In a practical situation, this leads to a 2 to 3 μm maximum resolution in a 1 μm thick resist layer. To improve the resolution of a negative resist, thinner resist layers can be used; however, when using thin layers of negative resist, pinholes become problematic. Xylene is the most commonly used aromatic solvent for negative resists, although almost any
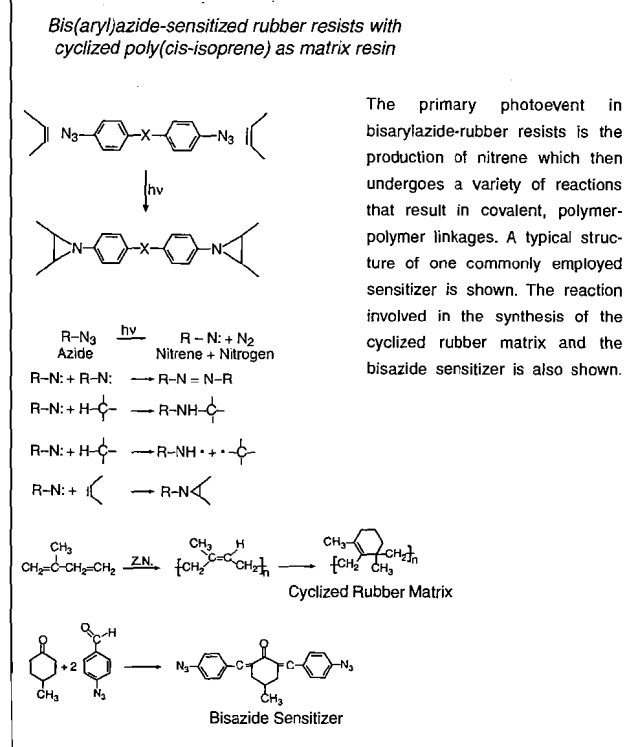
organic solvent will do. Aromatic solvent developers may pose environmental, health, and safety concerns. Newer negative resists are water developable.

An example of a commercial, two-component negative photoresist is the Kodak KTFR [an azide-sensitized poly(isoprene) rubber] with a lithographic sensitivity (also photospeed) of 75 to 125 mJ/cm^2. Negative photoresists, in general, adhere very well to the substrate, and a vast amount of compositions are available (stemming from R&D work in paints, UV curing inks, and adhesives, all based on polymerization hardening). Negative resists are highly resistant to acid and alkaline aqueous solutions as well as to oxidizing agents. As a consequence, a given thickness of negative resist is more resistant than a corresponding thickness of positive resist. This chemical resistance ensures better retention of resist features even during a long, aggressive wet or dry etch. Negative resists also are more sensitive than positive resists but exhibit a lower contrast (γ smaller; see *Contrast and Experimental Determination of Lithographic Sensitivity*, page 19).

A comparison of negative and positive photoresist features is presented in Table 1.1. This table is not exhaustive and is meant only as a practical guide for selection of a resist tone. The choice of whether to use a negative or a positive resist system depends on the needs of the specific application, such as resolution, ease of processing, speed, and cost. The choice of resist tone will even depend on the specific intended pattern geometry, which is known as the *optical proximity effect*. For example, an isolated single line most easily resolves in a negative resist (higher-resolution line), whereas an isolated hole or trench is most easily defined in a positive resist. Because traditional negative resists used to have a line width limit of only

TABLE 1.1 Comparison of Traditional Negative and Positive Photoresists[*]

Characteristic	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	± 2 μm
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 μm 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 Metal steps	Acid Simple solvents	Acid Chlorinated solvent compounds
Swelling in developer	No	Yes
Thermal stability	Good	Fair
Wet chemical resistance	Fair	Excellent

The "Resist type" header spans both the Positive and Negative columns.

[*]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} C/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 layers[15] 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/cm²
	APEX-E (IBM and Shipley)	Positive	75 mJ/cm²
	XP-2198 (Shipley)	Positive	30 mJ/cm²
	KRF (from UCB-JSR)	Negative	20–30 mJ/cm²
E-beam	COP [copolymer-(α-cyano ethyl acrylate-α-amido ethyl acrylate)]	Negative	0.5 μC/cm²
	GeSe [germanium selenide]	Negative	80 μC/cm²
	PBS [poly-(butene-1-sulfone)]	Positive	1 μC/cm²
	PMMA	Positive	100 μC/cm²
X-ray	COP	Negative	100 mJ/cm²
	DCOPA	Negative	14 mJ/cm²
	PBS	Positive	170 mJ/cm²
	PMMA	Positive	6500 mJ/cm²

of biosensors and microfluidics were recognized as well.[17] For many applications, the poorer resolution of dry resists (25 to 75 μm) 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 Å 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

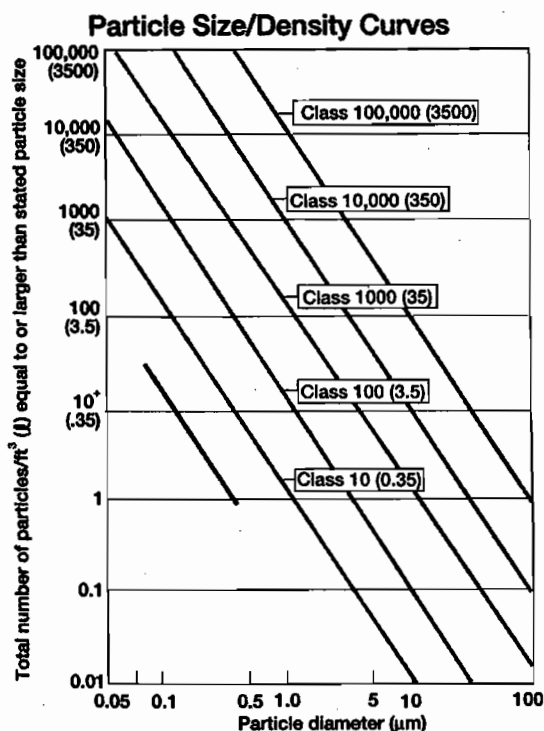**Particle Size/Density Curves**

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 NH_3 (25% aqueous solution) to five parts of DI water; heat to boiling and add one part of H_2O_2. 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 H_2O_2. Immerse the wafer for ten minutes. This procedure removes metal ions.
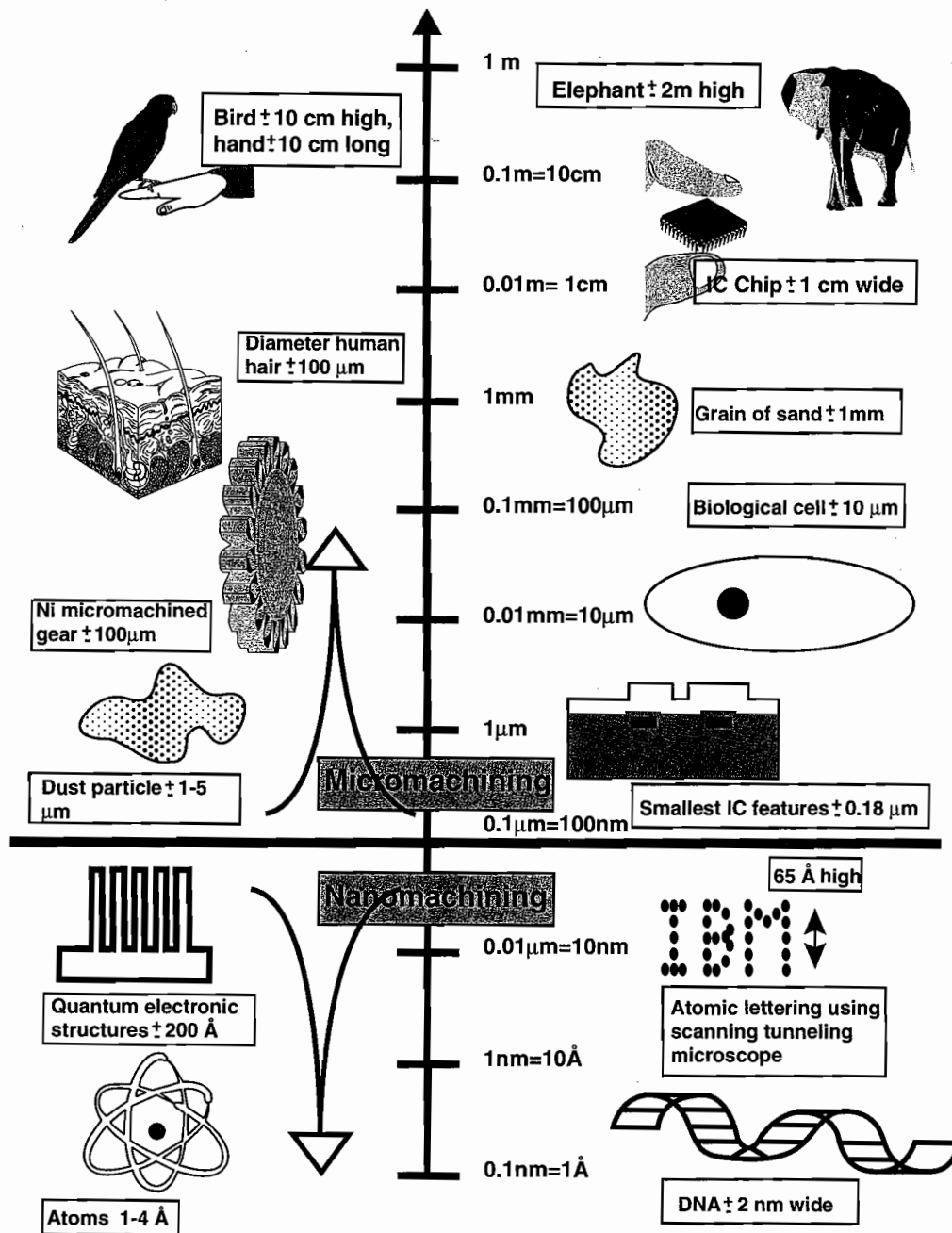
Figure 1.5   Various objects and their linear size. (This figure also appears in the color plate section following page 394.)

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 CO_2 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 pressure pulsation. (A) 1200 psi and (B) 800 psi. (From Bok et al., *Solid State Technol.*, 35, 117–119, 1992. With permission.)
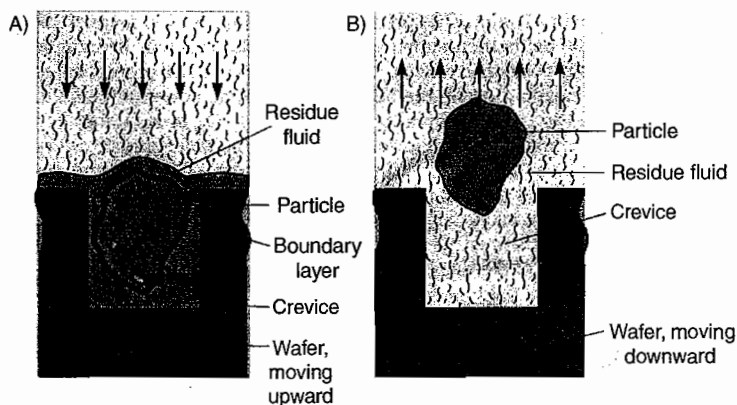
Inset 1.8

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 ($\varepsilon_r < 3$) 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 submicron 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-

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.)

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 ±0.1 μm (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 ±.02 μm (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 ±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
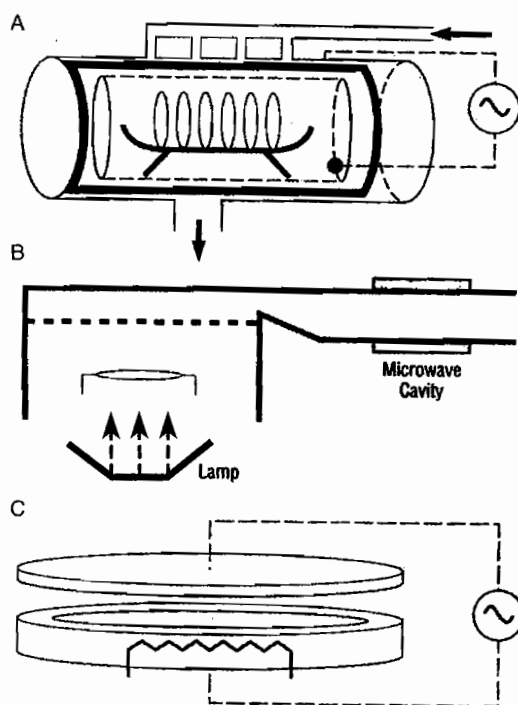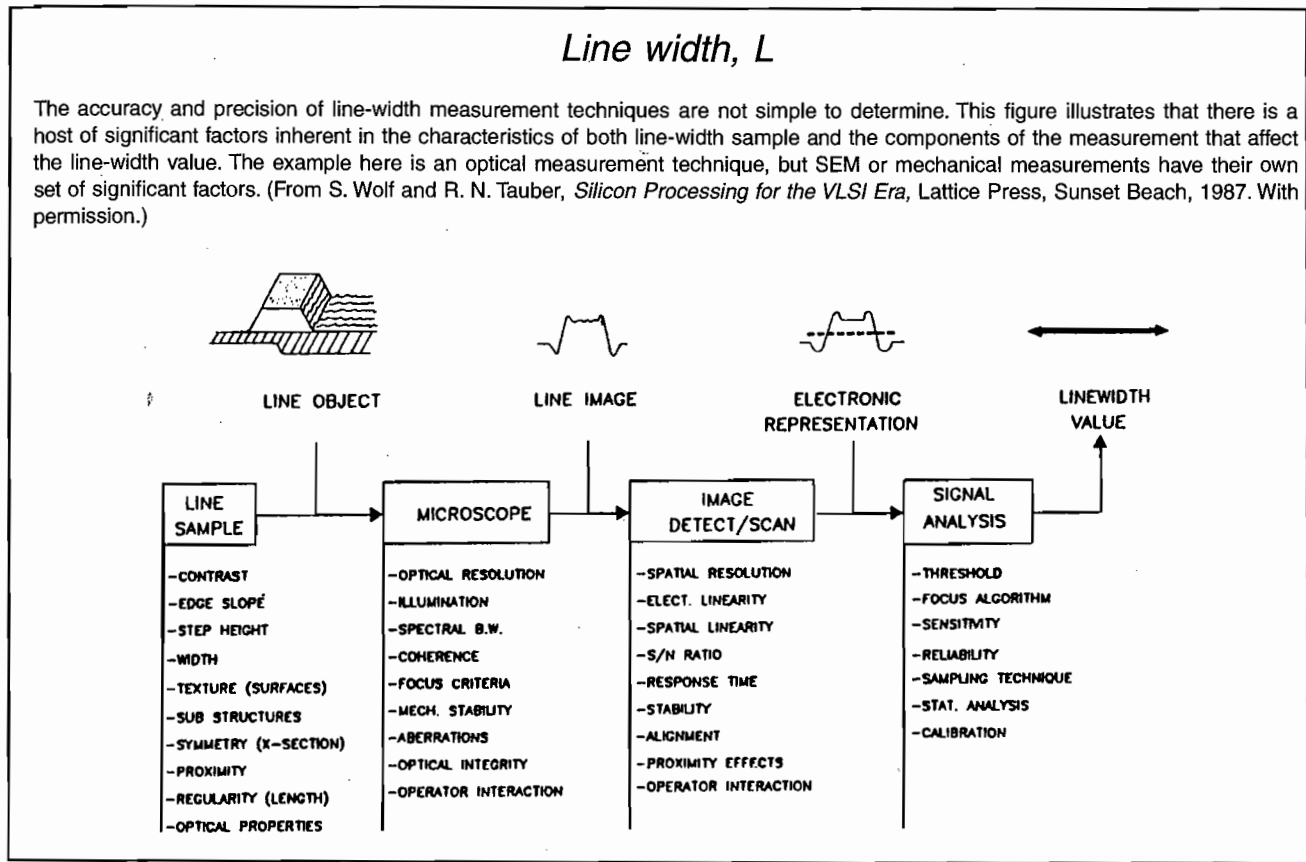

Line width, L

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.)

LINE OBJECT   LINE IMAGE   ELECTRONIC REPRESENTATION   LINEWIDTH VALUE

LINE SAMPLE	MICROSCOPE	IMAGE DETECT/SCAN	SIGNAL ANALYSIS
−CONTRAST	−OPTICAL RESOLUTION	−SPATIAL RESOLUTION	−THRESHOLD
−EDGE SLOPE	−ILLUMINATION	−ELECT. LINEARITY	−FOCUS ALGORITHM
−STEP HEIGHT	−SPECTRAL B.W.	−SPATIAL LINEARITY	−SENSITIVITY
−WIDTH	−COHERENCE	−S/N RATIO	−RELIABILITY
−TEXTURE (SURFACES)	−FOCUS CRITERIA	−RESPONSE TIME	−SAMPLING TECHNIQUE
−SUB STRUCTURES	−MECH. STABILITY	−STABILITY	−STAT. ANALYSIS
−SYMMETRY (X−SECTION)	−ABERRATIONS	−ALIGNMENT	−CALIBRATION
−PROXIMITY	−OPTICAL INTEGRITY	−PROXIMITY EFFECTS	
−REGULARITY (LENGTH)	−OPERATOR INTERACTION	−OPERATOR INTERACTION	
−OPTICAL PROPERTIES			

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:

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

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 (»10 μm 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]

$$M_n^0 = \frac{wN_A}{N_o} \qquad (1.3)$$

where N_A = Avogadro's number.

Rearranging Equation 1.3 yields:

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

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:

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

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:

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

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:

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

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:

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

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

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

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.
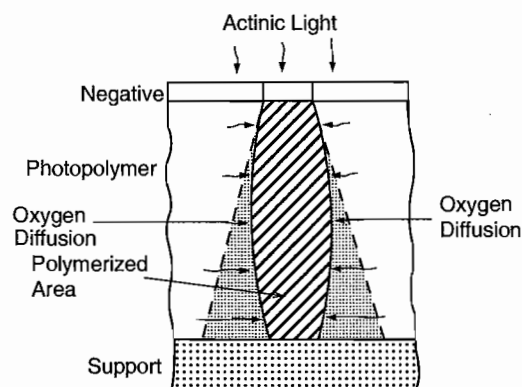
Time-independent organic solvent development of overexposed negative resists from outside the scattered region reveals the scattered radiation zone, because cross-linking and further swelling inhibit its removal. On the other hand, development of overexposed positive resists using aqueous alkaline is 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 μm. 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, e.g., 95–110°). 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

DEVELOPMENT STEP

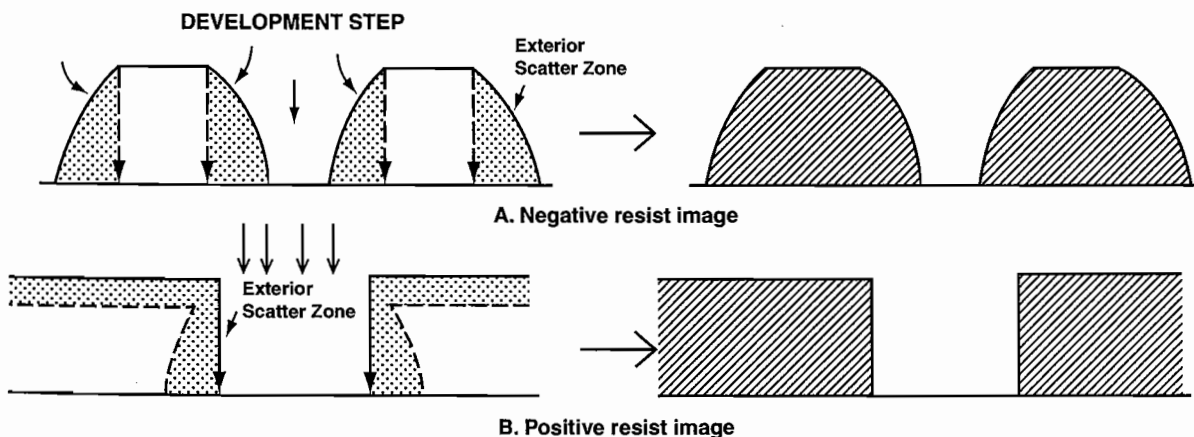**A. Negative resist image**

B. Positive resist image

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.)
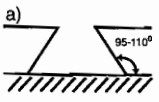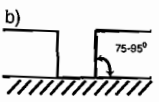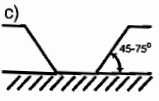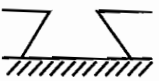
Profile	Dose	Developer Influence	R/R_0	γ	Uses
A. Positive resists Undercut a) 95-110°	High (often with back scatter radiation)	Low	> 10	> 6	Ion implant, lift-off Not good for plasma etching. Often only obtained through image reversal
Vertical b) 75-95°	Normal dose	Moderate	5 - 10	< 4	Lift-off, Reactive ion etch wet etch Ion beam etch Perfect fidelity
Normal or overcut c) 45-75°	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° *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 (~90°, 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
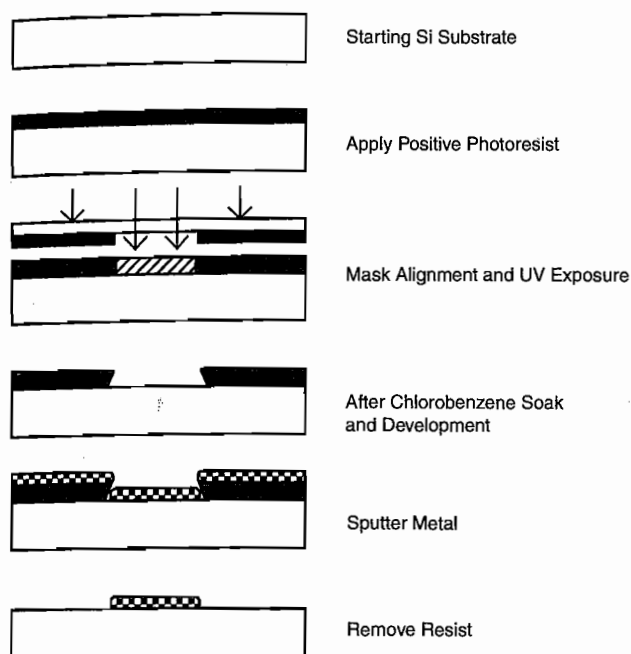
Figure 1.9 Example of lift-off sequence with positive resist for the construction of a Pt-based electrochemical sensor electrode. Rounding of deposited features through shadowing is observed (see text).

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 "presoak" 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 hexamethyldisizilane (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:

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

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,
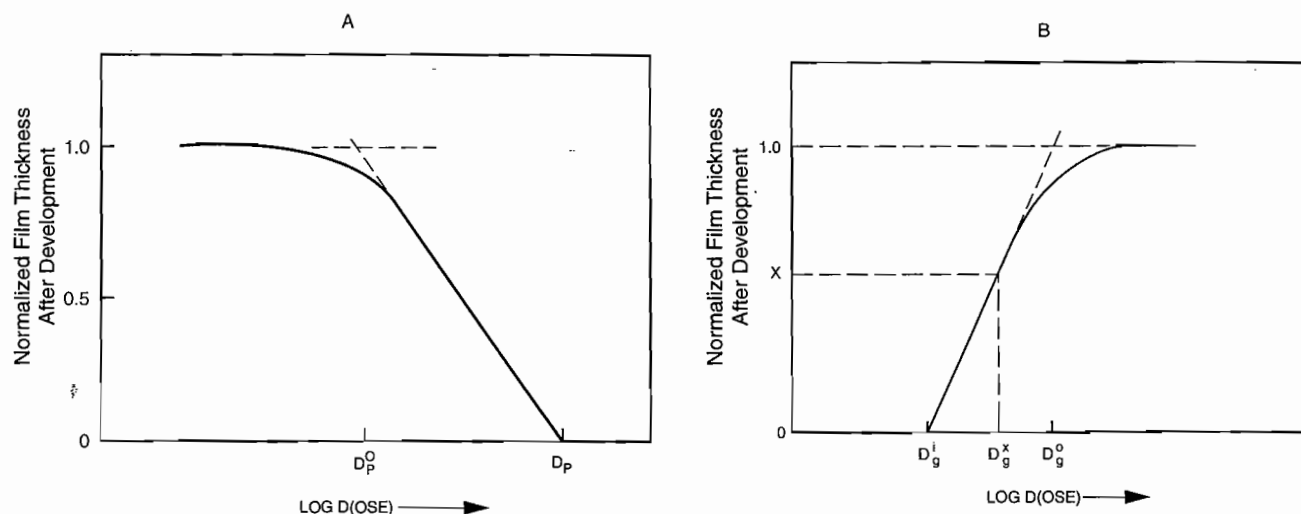
A

B

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:

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

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
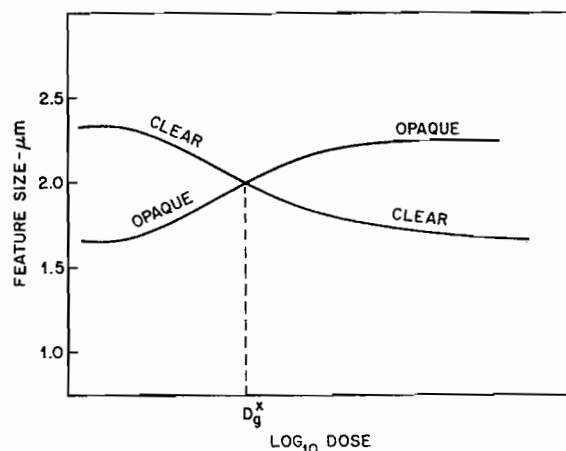

Figure 1.11 Size of a clear and opaque 2.0-μm feature (on the mask) as a function of the exposure dose for a negative resist. The dose (D_g^x) resulting in the correct feature size (same size as on the mask) is called the *lithographic sensitivity*. (From C.G. Willson in *Introduction to Microlithography*, 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:

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

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

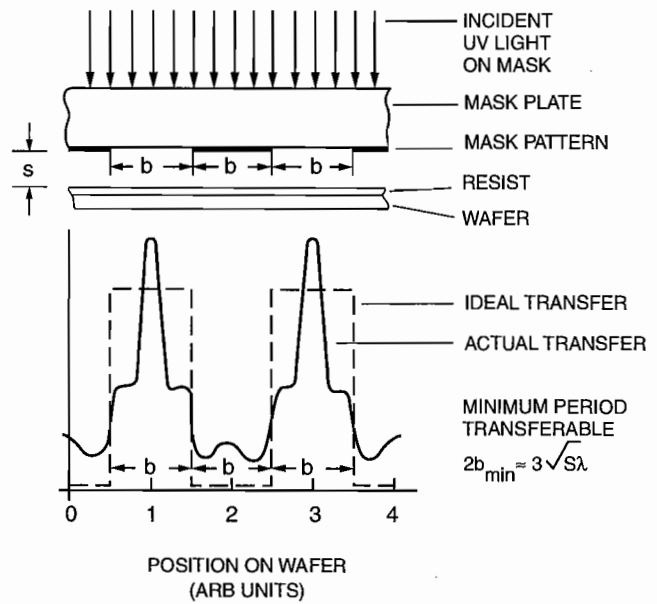s = gap between the mask and the photoresist surface

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

$$R = b_{min} = \frac{3}{2}\sqrt{\frac{\lambda z}{2}} \tag{1.13}$$

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:

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

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 °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:

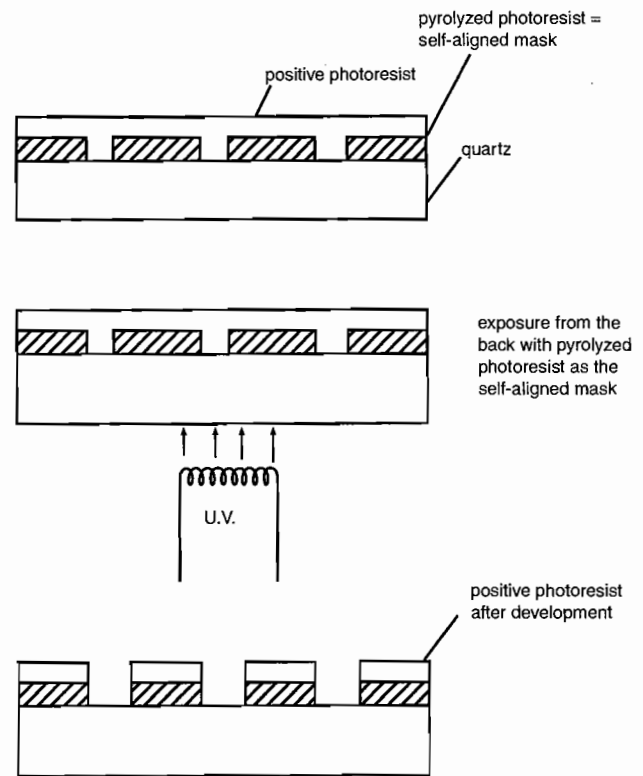$$R = \frac{k_1 \lambda}{NA} \qquad (1.15)$$

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:

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

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).
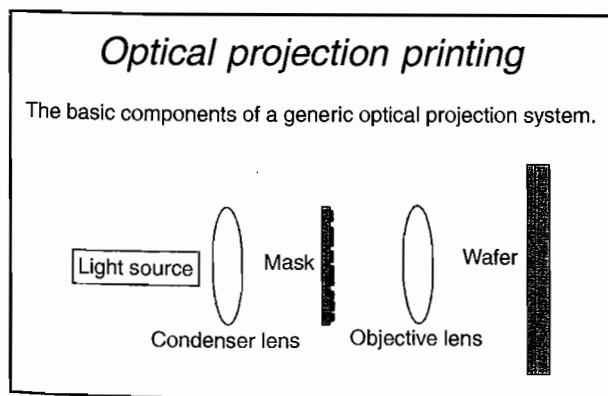
A stepper system, also step-and-repeat, exposes one small part of the wafer, followed by a new exposure at the next position. By stepping and repeating, the entire wafer is covered with the reticle pattern. Steppers have the ability to align to each field and make adjustments to x, y, rotation and focus, 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.

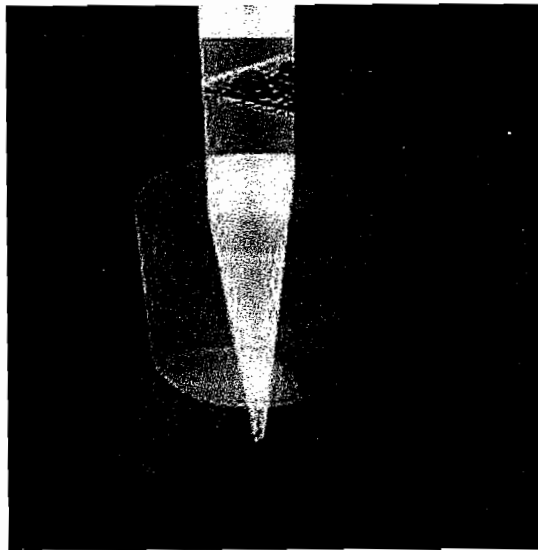Light source — Condenser lens — Mask — Objective lens — Wafer

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.
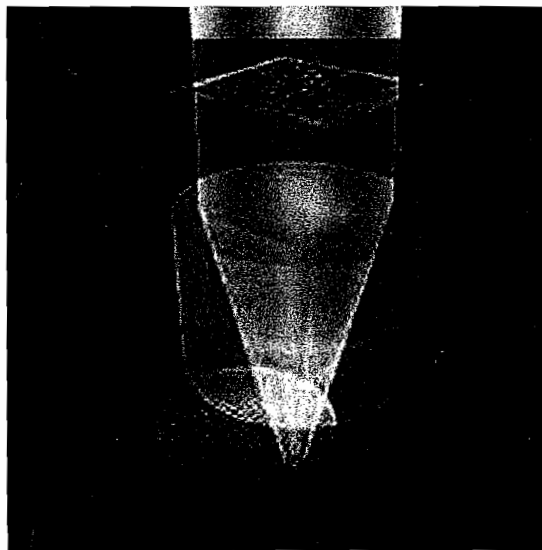
Stepper Principle   Step & Scan Principle

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:

$$M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{1.17}$$

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 (ν) 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 $M_{im}(ν)$ 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 $ν_{max} = NA/λ$, the *MTF* curve for the incoherent case decreases monotonically as the frequency increases up to a maximum value $ν_{max}$ of $2NA/λ$. Since $ν_{max} = 1/2b_{min}$ and $R = b_{min}$, the maximum resolution for coherent light is predicted to be $0.5λ/NA$ (see also the derivation of Equation 1.24, below); for incoherent light it is $0.25λ/NA$ (see also the derivation Equation 1.27, below). Earlier, we saw that the Rayleigh criterion for two incoherent light sources predicted $0.61λ/NA$ for a separation of $2b_{min}$ or $0.3λ/NA$ for resolution *R*. The latter is a close enough approximation of $0.25λ/NA$. 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]
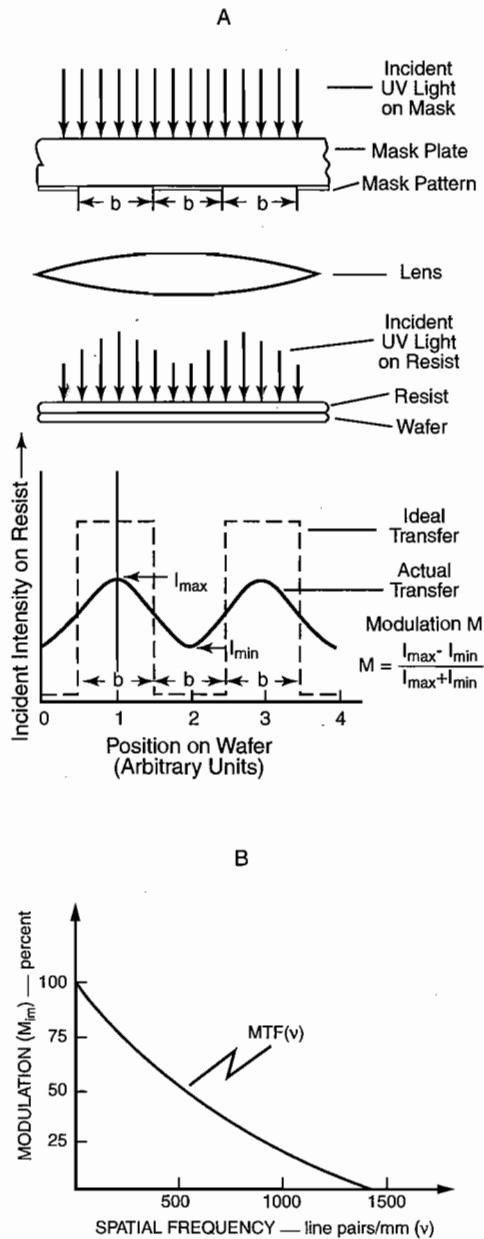
A

B

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:
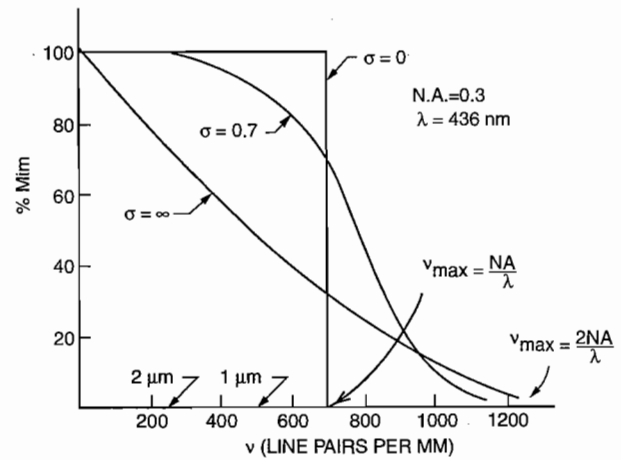
Figure 1.15 Percent modulation in the image plane, % M_{im}, modulation of an image as a function of spatial frequency ν for different coherency factors. (A) σ = 0 (coherent); (B) σ = 0.7 (partially coherent) and σ = infinite (incoherent). (From I. Brodie and J.J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, 1982.[28] Reprinted with permission.)

$$CMTF_{resist} = \frac{10^{\frac{1}{\gamma}} - 1}{10^{\frac{1}{\gamma}} + 1} \tag{1.18}$$

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:

$$NA = n\sin\theta_{max} \tag{1.19}$$

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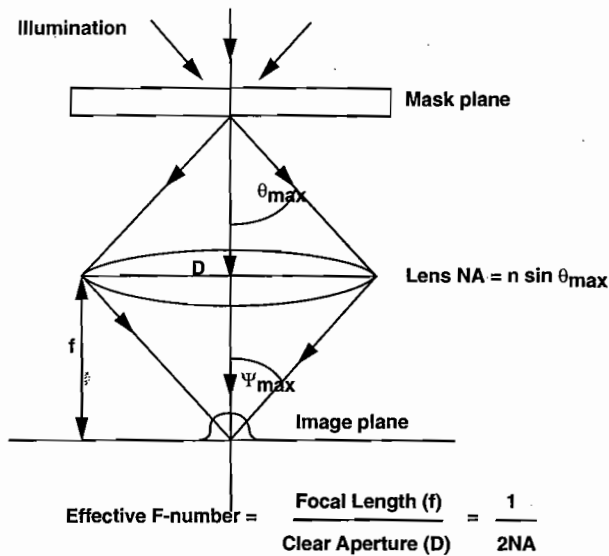
Figure 1.16 Relationship between the object, image, and focal length and diameter of a lens to define the numerical aperture.

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

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:

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

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:

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

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)*.
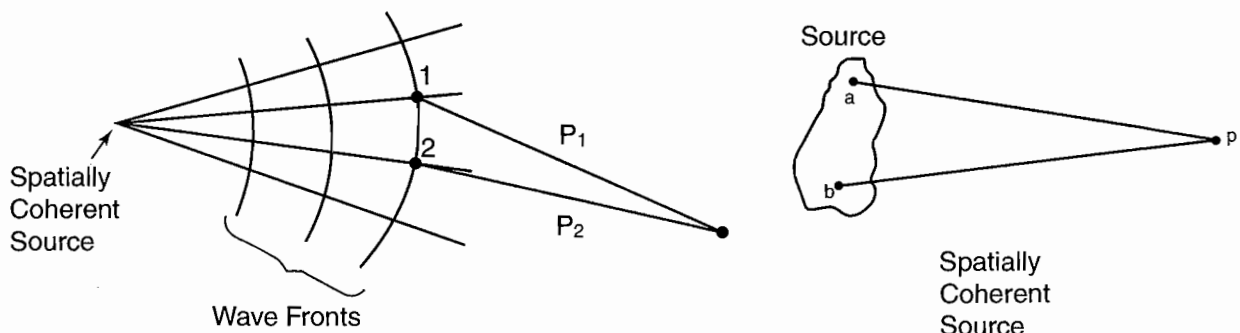
Inset 1.13

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:

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

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:

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

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:

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

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:

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

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× 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.)
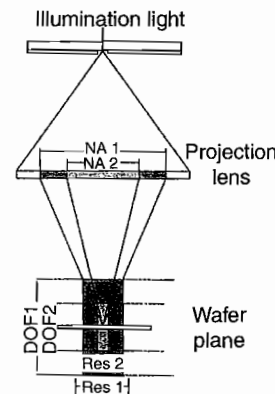
Inset 1.14

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 ±0.25 μm 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
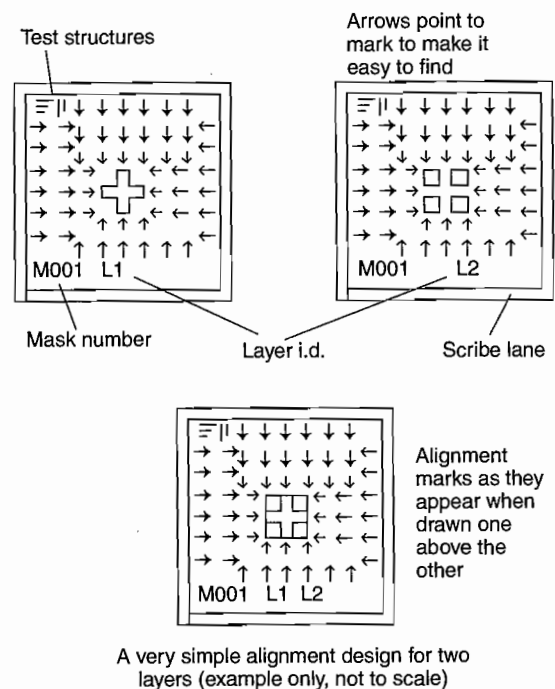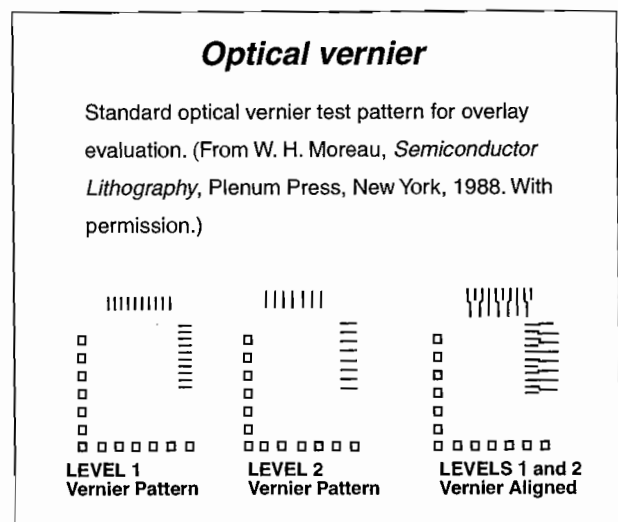

Figure 1.17 A simple set of fiducial marks and other important guiding features often found on masks. (Based on Banks at http://www.dbanks.demon.co.uk/ueng/.)

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

Optical vernier

Standard optical vernier test pattern for overlay evaluation. (From W. H. Moreau, *Semiconductor Lithography*, Plenum Press, New York, 1988. With permission.)
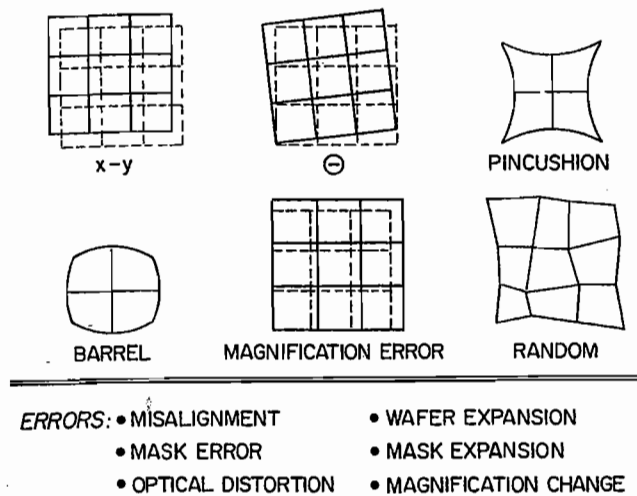
Inset 1.15

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 desirable 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
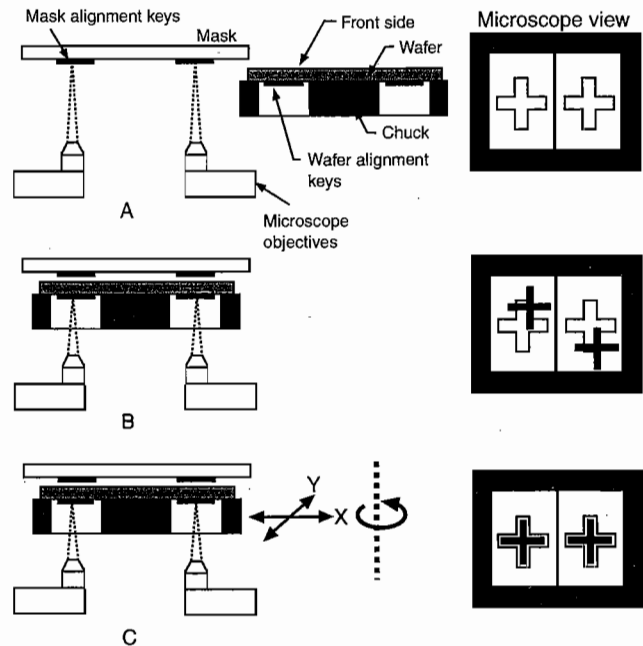

Figure 1.19 Double-sided alignment scheme for the Karl Suss MA-150 production mode system: (A) the image of mask alignment marks is electronically stored; (B) the alignment marks on the back side of the wafer are brought in focus; (C) the position of the wafer is adjusted by translation and rotation 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 Suss 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 Wenzel[37] 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

Next-Generation Lithographies

Introduction

Having set a rather optimistic forecast for the future of miniaturization, we now introduce alternatives for DUV photolithography, including extreme ultraviolet lithography (EUVL), x-ray lithography, and charged particle-beam lithography, including scattering with angular limitation projection electron-beam lithography (SCALPEL) and ion projection lithography (IPL). These methods are all options for next-generation lithography (NGL).

In the IC industry, continuous improvements in optical lithography have postponed the industrial adoption of alternative lithographies because of the huge financial investment required in new photolithography equipment. Some critics of alternative technologies have even said that photolithography will never be displaced by another method, since cost and technical obstacles will make it pointless to go smaller than 0.1 μm.[87] More likely, at a 0.1 μm feature size (somewhere after the year 2005) in which i-line and g-line light have long since ceased to be used, and neither 248 nm nor 193 nm excimer lasers are effective anymore, commercial incentives will beckon, and one of the alternatives discussed below or perhaps yet another method will be introduced.

IC and miniaturization science are taking increasingly separate paths in adopting preferred lithography strategies. With ICs, throughput and finer geometries are needed; batch processing is a prerequisite. In miniaturization science, modularity, good depth of focus, extending the z-direction [that is, the height of features (skyscraper-type structures)], incorporating nontraditional materials (e.g., gas-sensitive ceramic layers, polymers), and replication methods catch the spotlight; batch fabrication is not always a prerequisite. While reading the following sections, these important differences in characteristics of IC manufacture and miniaturization science should be kept in mind.

Extreme Ultraviolet Lithography

As mentioned above, somewhere around 2005, ICs will feature CDs below 100 nm. Photolithography using 193 nm light and exploiting resolution enhancing techniques (RETs) alone will not be able to write smaller patterns, and next-generation lithographies (NGLs) will have to come on-line. Extreme ultraviolet lithography is one of the most promising.

Extreme ultraviolet lithography (EUVL), using wavelengths in the range of 10 to 14 nm to carry out projection imaging, is perhaps the most natural extension of optical projection lithography as, in principle, it only differs in terms of the wavelength. This type of radiation is also referred to as *soft-x-ray radiation* and *vacuum UV*. Sources for this type of radiation are laser-produced plasmas and synchrotrons.

To appreciate the merits of EUVL, we take as a starting point the expressions for resolution (R) and depth of focus (DOF), that is Equations 1.15 and 1.28. We learned that, in practice, these parameters are experimentally determined and are influenced by both resist and exposure systems. Typically, in large-scale IC manufacture, k_1 and k_2 values in Equations 1.15 and 1.28 are greater than 0.6 and DOFs larger than 0.5 μm. These conditions lead to the desired CD control within a tolerable process window.[34] The push for higher resolution with 248 nm and 193 nm radiation has made it necessary to accept values for k_1 and k_2 that are smaller than 0.5. This led to processing problems, among others; it is more difficult to control the CD and creates an intolerably low DOF. The resolution enhancing technologies (RETs) discussed above enhance resolution and the effective DOF but, as we discussed, they are often not manufacturer friendly. EUVL, on the other hand, surmounts these problems because of its shorter wavelength. The technique is capable of printing sub-100-nm features while maintaining a DOF of 0.5 μm or larger, and it has k-values that make the process control less demanding. Unfortunately, things are not that simple—EUV is strongly absorbed in virtually all materials and, consequently, imaging must be carried out in vacuum; also, all camera optics as well as masks used must be reflective rather than refractive. New resists and processing techniques must be developed as well.

With respect to EUV reflectivity, it is well known that most materials have very low reflectivity for near-normal incidences at those short wavelengths, so multilayer Bragg reflectors with multiple reflecting coatings are under development. These coatings consist of a large number (e.g., 80) of alternating layers of materials (e.g., Mo and Si) having dissimilar EUV optical constants, and they provide a resonant reflectivity when the period of the layers is close to λ/2. With magnetron sputtered Mo/Si based reflectors, peak reflectivities of up to 70% at 13.4 nm have been achieved. The multilayer stacks in EUVL masks today have a demonstrated defect level of ~10^{14} defects/cm², still two orders of magnitude above target. Prototype EUVL cameras designed for 13.4 nm radiation have an NA = 0.1 and a four-mirror set with multilayer Mo/Si reflecting coatings.[34] These step-and-scan 4× cameras have a resolution better than 100 nm with a ring field of 26 × 1.5 mm. In the step-and scan system, the mask and wafer are simultaneously scanned in opposite directions, with the mask moving four times faster than the wafer. The camera wave-front error at 13.4 nm needs to be less than 1 nm so that the surface figure (that is, the basic shape of each of the mirrors) needs to be accurate to 0.25 nm rms or better. Achieving these accuracies will also require improved metrology techniques, as current ones cannot even confirm the accuracies required now. The fact that EUVL masks are reflective has advantages and disadvantages. Advantages are that no thin membranes are needed as in the case of x-ray lithography—just a solid substrate (e.g., a silicon wafer) coated with the required reflective coatings and a patterned EUV absorber; because of the 4:1 reduction, mask making is also easier than with x-rays. A major disadvantage is that the defect density in the reflective coating needs to be made so small that there are few techniques available yet to accomplish such defect-free films. Current resists absorb all of the EUV photons within less than 100 nm. Approaches for EUV resists include silylated single layer resist, refractory bilayer resists, and trilayer resists.
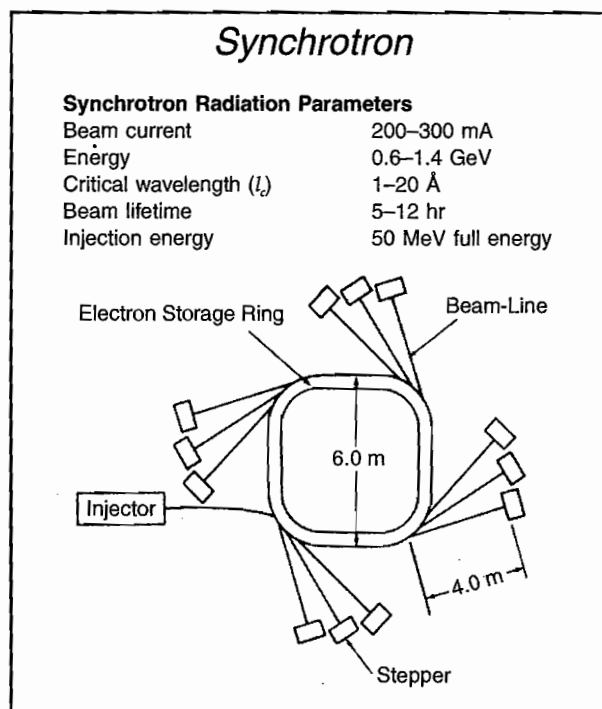
X-Ray Lithography

Introduction

In contrast with electron lithography and ion-beam lithography, no charged particles are directly involved in x-ray exposures, which eliminates the need for vacuum (see Figure 1.38).[28] Another advantage of x-rays is that one can use flood exposure of resist-coated wafers, ensuring higher throughput than when writing with a thin electron or ion beam. The method is also referred to as deep x-ray lithography (DXRL).

The three main classes of sources for x-ray lithography are electron impact tubes, laser-based plasmas, and synchrotrons. In miniaturization science, we mainly will be concerned with synchrotron radiation (Inset 1.21). For reasons that are explained below, in the case of ICs, less expensive alternative x-ray sources might suffice.

Major Features of X-Ray Lithography

X-ray lithography is superior to optical lithography because of the use of shorter wavelengths and a very large DOF, and because exposure time and development conditions are not as stringent. Reproducibility is high, as results are independent of substrate type, surface reflections, and wafer topography. Another important benefit is that x-ray lithography is immune to low-atomic-number (Z) particle contamination (dust). With an x-ray wavelength on the order of 10 Å or less, diffraction effects generally are negligible, and proximity masking can be used, increasing the lifetime of the mask. With a standard 50 μm proximity gap and using synchrotron x-rays, 0.25 μm patterns can be printed; by decreasing the proximity gap to 25 μm, patterns of 0.15 μm can be resolved.[88] The obtainable aspect

Figure 1.38 A comparison of photolithography, electron-beam, ion-beam, and x-ray lithography. (Based on I. Brodie and J.J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, 1982.[28])

Inset 1.21

ratio, defined as the structural height or depth to the minimum lateral dimension, reaches more than 100. (With UV photolithography, under special conditions, an aspect ratio of about 25 is possible at most; see SU-8.) An aspect ratio of 100 corresponds to the aspect ratio attainable by wet chemical anisotropic etching of monocrystalline Si (see Chapter 4).

In x-ray lithography, there are essentially no optics involved and, although this sounds like an advantage, it also presents one major disadvantage—one can only work with 1:1 shadow printing. No image reduction is possible, so the mask fabrication process is very complicated. In the United States, IBM remains the only major champion of x-ray lithography for next-generation lithography. At the end of 1999, IBM fabricated several PowerPC 604e microprocessor batches to demonstrate the viability of the method. In Japan, there are still many players involved such as NTT, Toshiba, Mitsubishi, and NEC.[89]

LIGA

The LIGA technique (a German acronym for Lithographie, Galvanoformung, Abformung) was invented about 20 years ago.[90] LIGA exploits all of the advantages of x-ray lithography listed above, and the process is schematically illustrated in Figure 6.1. It involves a thick layer of resist (from micrometers to centimeters), high-energy x-ray radiation, and development to make a resist mold. By applying galvanizing techniques, the mold is filled with a metal. The resist structure is removed, and metal products result. Alternatively, the metal part itself can serve as a mold for precision plastic injection molding. Several types of plastic molding processes have been tested, including reaction injection molding, thermoplastic injection molding, and hot embossing. The so-formed plastic part, just like the original resist structure, may again serve as a mold for fast and cheap mass production, since one does not rely on a new x-ray expo-
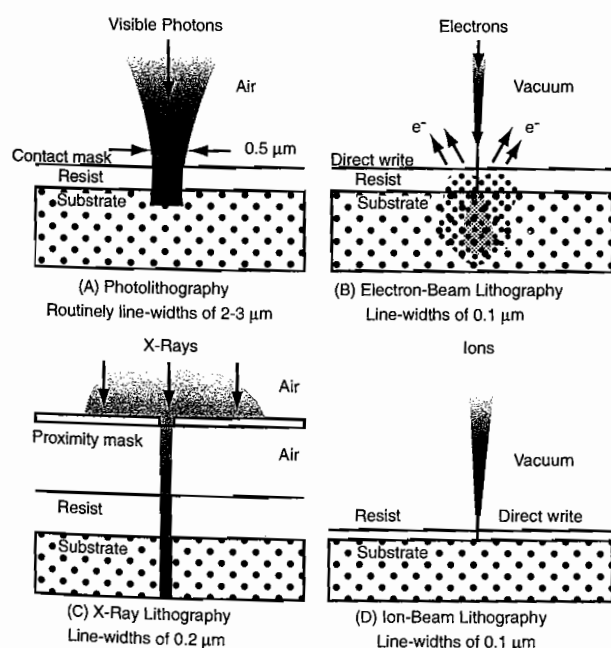
sure. LIGA enables new building materials and a wider dynamic range of dimensions and possible shapes. A showpiece structure for LIGA technology is pictured in Figure 1.39 in which an ant holds an Ni gear.

Of particular interest to miniaturization science is the possibility of creating three-dimensional shapes with slanted sidewalls and step-like structures as shown in Figure 1.40.[90] Two German companies, IMM (http://www.imm-dueck.de/) and Jenoptik Mikrotechnik GmbH (http://www.jo-mikrotechnik.com/), have developed an x-ray scanner that allows continuous tilt angles of the mask/substrate assembly with respect to the x-ray beam and rotation of the mask/substrate, and it contains an internal mask alignment system for different masks for multiple exposures. The unprecedented precision attainable with LIGA makes this technique stand out against other 3D lithography methods such as DUV with SU-8 (see above). Using the IMM/Jenoptik scanner, a vertical resist wall with a precision of less than 0.05 μm was demonstrated, and even resist walls inclined at a 45° angle were made with an accuracy of 1 μm over a height of up to 500 μm.[90] LIGA is further discussed in Chapter 6.

X-Ray Resists

An x-ray resist should have high sensitivity to x-rays; high resolution and resistance to chemical, ion, and/or plasma etching; thermal stability of >140°C; and a matrix or resin absorption of less than 0.35 μm^{-1} at the wavelength of interest. No present resist meets all those requirements. One material predominantly used in x-ray lithography is positive poly(methylmethacrylate) or PMMA, a material better known by its trade names Plexiglas™ and Lucite™. Another name commonly used for polymers based on PMMA is *acrylics*. Clear sheets of the material

Figure 1.39 Ant with gear. This figure also appears in the color plate section following page 394, and on the back cover of the book. (From Forschungszentrum Karlsruhe, Program Microsystem Technologies. Reprinted with permission.)
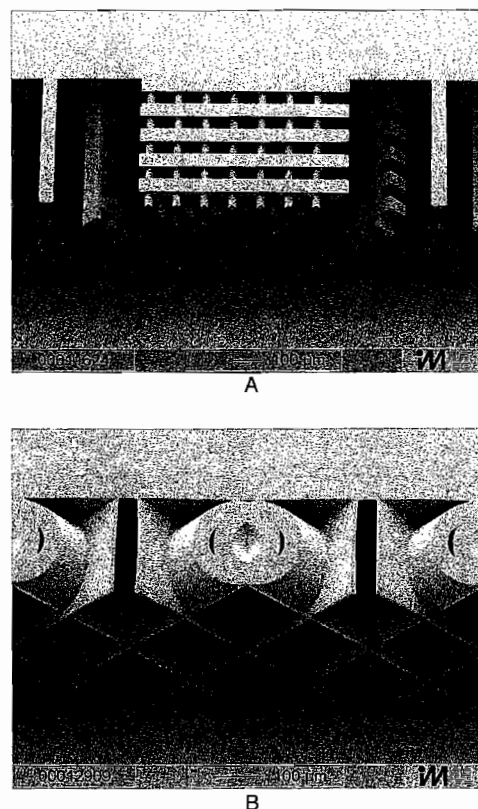

Figure 1.40 LIGA structures obtained at IMM using an x-ray scanner enabling continuous tilt angles of the mask/substrate assembly.[90] (Courtesy of IMM.)

are used, for example, to fabricate "unbreakable" windows, inexpensive lenses, machine guards, clear lacquers on decorative parts, etc. At a wavelength of 8.34 Å, the lithographic sensitivity of PMMA typically hovers about 2–6.5 J/cm^2, a rather low sensitivity implying a small throughput. A possible approach to make PMMA more x-ray sensitive is the incorporation of x-ray-absorbing high-atomic-number atoms. Another approach, discussed earlier, involves the use of chemically amplified photoresists. More recent x-ray resists explored for LIGA applications are poly(lactides). These resists show a considerably enhanced sensitivity and reduced stress corrosion compared to PMMA.[91]

Negative x-ray resists exhibit inherently higher sensitivities as compared to positive x-ray resists, although their resolution capability is limited by swelling. Poly(glycidyl methacrylate-co-ethyl acrylate) (PGMA), a negative e-beam resist, has been used in x-ray lithography. In general, resists materials sensitive to e-beam exposure are also sensitive to x-rays and function in the same way; that is, materials that are positive in tone for electron-beam radiation typically also are positive in tone for x-ray radiation. A strong correlation exists between the resist sensitivities observed with these two radiation sources, suggesting that the reaction mechanisms might be similar for both types of irradiation.[14]

The IC industry requires a typical resist layer not more than 1 to 2 μm thick. Thicker layers, say between 10 and 1000 μm, are dictated by the need for high-aspect-ratio micromachines. The technology of applying thicker layers of photoresist remains challenging. Spin coating of multiple resist layers (for relatively

thin coats), resist casting with *in situ* polymerization of mildly cross-linked PMMA (for layers above 500 μm), and plasma polymerization of PMMA with a possibility of subsequent diamond grinding of the resulting layers are some of the techniques currently in use.

In the case of exposures with very high energy x-rays (hard x-rays), the associated wavelength is measured in angstroms, not in nanometers. At those energies, almost every type of polymer becomes a "resist" and even "resistless" lithography becomes possible as thin films can be etched, vaporized, or ion-implanted directly.

X-Ray Masks

Another challenge in x-ray lithography, besides the low sensitivity of the resists and the high cost of sufficiently bright x-ray sources, is the mask making—already complex for producing DRAMs, but even more complex for 3D structures with high aspect ratios. In Table 1.6, the procedure for making an optical mask is compared to the procedure involved in making an x-ray mask. Technologies developed for manufacturing masks for submicron circuitry with x-ray lithography do not directly transfer to the fabrication of x-ray masks for building micromachines. An x-ray mask basically consists of a pattern of x-ray-absorbing material (a material with a high atomic number, Z, such as gold, tungsten, or tantalum silicide) on a membrane substrate that is transparent to x-rays (a low-Z material, e.g., Ti, Si, SiC, Si_3N_4, BN, Be). The membrane must have windows or other features in certain areas for alignment purposes. Yet the material also needs to be thermomechanically stable to a few parts in 10^6.[92] Mechanical stress in the absorber pattern can cause in-plane distortion of the supporting thin membrane, requiring a high Young's modulus material. Also, humidity or the x-ray exposure itself might distort the membrane. A fabrication sequence for an x-ray mask is shown in Figure 1.41. In this case, the x-ray absorber pattern is ion-beam etched in an Au film after the pattern has been first written in a 300 Å PMMA layer with an e-beam. The absorber film typically consists of two metal layers: a thin layer of chromium for adhesion to the substrate, topped by a thicker layer of gold. The higher the required aspect ratio of the exposed resist, the thicker the gold layer of the mask absorber pattern must be to maintain a good

contrast. The x-ray mask shown in Figure 1.41 only has a 400 Å thick gold layer, which adequately covers DRAM manufacture but not high-aspect-ratio 3D structures where 5 to 15 μm of Au absorber might be required for a 500 μm thick resist and up to 50 μm for 10 cm thick resists. A proximity x-ray mask as shown in Figure 1.41 is placed close to the substrate and its pattern is reproduced by exposure to x-rays.

Various other schemes to make LIGA x-ray masks are explored in Chapter 6.

Why Use a Synchrotron to Generate X-Rays?

The full power of x-ray lithography for miniaturization science materializes only when using hard, collimated synchrotron radi-
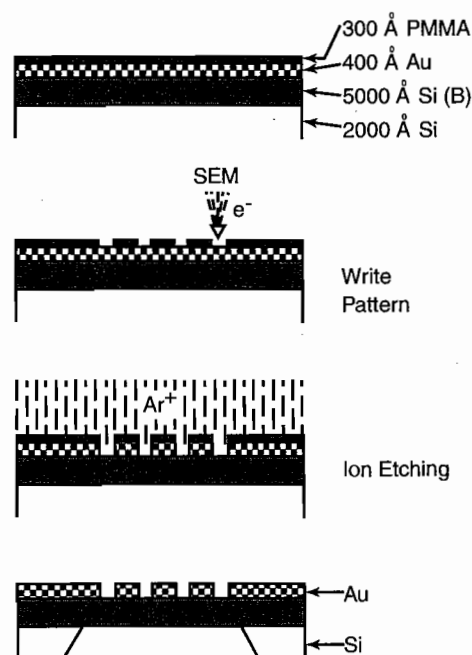

Figure 1.41 Fabrication of a silicon-membrane-based x-ray mask with a gold absorber pattern. For use in high-aspect-ratio micromachining, the gold absorber layer must be between 5 and 15 μm. (Based on I. Brodie and J.J. Muray, *The Physics of Microfabrication,* Plenum Press, New York, 1982.[28])

TABLE 1.6 Optical vs. X-Ray Mask

Optical mask	X-ray mask
Mask design: CAD	Mask design: CAD

Substrate preparation:
    Quartz
    Thin metal film deposition
	Substrate preparation:
	    Thin membrane substrate (Si, Be, Ti,)
	    Deposit plating base (50 Å Cr then 300 Å Au)

Pattern delineation:
    Coat substrate with resist
    Expose pattern (optical, e-beam)
    Develop pattern etch Cr layer
    Strip resist
	Pattern delineation:
	    Coat with resist
	    Expose pattern (optical, e-beam)
	    Develop pattern
	    Absorber definition:
	    Electroplate Au (~15 μm for hard x-rays)
	    Strip resist

Cost: $1k–3k	Cost: $4k–$12k
Duration: 3 days	Duration: 10 days

ation. To appreciate this, we will explore the procedure using a less intense, less collimated beam from an electron-beam bombardment source. A schematic diagram of an x-ray exposure system using an electron-beam-bombardment-based x-ray source is shown in Figure 1.42.

The mask typically is offset above the wafer by about 10 μm after alignment. A proximity scheme rather than a contact mask is a useful feature, given the fact that an x-ray mask can cost up to $13,000. Since the x-ray source is finite in size and separated by a distance, D, the edge of the mask does not cast a sharp shadow but rather has a region associated with it known as penumbral blur, d. Image blurring limits the ultimate resolution power of an x-ray exposure system, as shown in Figure 1.42. As diffraction effects can be ignored, simple geometric considerations can be used for relating the image to the pattern on the mask. From Figure 1.42, we estimate that the blurring, δ, at the resist plane is given by:

$$\delta = s\left(\frac{d}{D}\right) \qquad (1.38)$$

where s = mask-to-wafer gap
         d = source diameter
         D = source-to-substrate distance

In a high-resolution system, δ should be controlled to within 0.1 μm. Spacing, s, should allow the accommodation of large-diameter masks while avoiding the high risk of contacting the resist and greatly increasing the occurrence of defects. The x-ray source must be sufficiently collimated. In practice, this translates into a small source diameter d (e.g., a few millimeters) and a large source-to-mask distance (D). All these factors contribute to minimizing penumbral blurring. Conventional e-beam-generated x-ray sources have sizes of a few millimeters

and are about 40 cm away from the mask. Unfortunately, a large distance required for adequate collimation results in prohibitively long exposure times (e.g., hours) due to the weak intensity of these sources. With synchrotron radiation, on the other hand, penumbral blurring does not limit the spatial resolution. Because of the high collimation of synchrotron radiation, rather large distances between the mask and the wafer can be tolerated (about 1 mm for 1 μm line width patterns). In the electron storage ring or synchrotron, a magnetic field constrains electrons to follow a circular orbit and the radial acceleration of the electrons causes electromagnetic radiation to be emitted forward. The radiation is thus strongly collimated in the forward direction and can be assumed to be parallel for lithographic applications. Because of the much higher flux of usable collimated x-rays, shorter exposure times become possible. The Advanced Light Source (ALS) synchrotron in Berkeley, for example, can deliver a flux of 0.4 W/cm^2 at 30 m for 3 to 9 keV radiation. Especially if one wants to generate a highly collimated photon flux in the spectral range required for precise deep-etch x-ray lithography in thick resist layers, synchrotron radiation comes close to being the ideal source because of its intensity, tunability, small source size, and small divergence.[93]

As mentioned before, the IC industry favors improved photolithography over x-rays. The prohibitive cost of introducing a new type of industrial lithography remains a strong deterrent, pushing existing photolithography to its absolute resolution limit. For example, by using UV phase-shifting mask lithography, planar IC features approaching x-ray lithography resolution have been made. Even in the micromachining field, there are continued attempts to squeeze more out of classical photolithography. Using techniques such as deep UV lithography and deep dry etching with dense plasmas, LIGA-like, high-aspect-ratio features have been produced, hemming in the potential for x-ray lithography even for building 3D miniaturized machines.
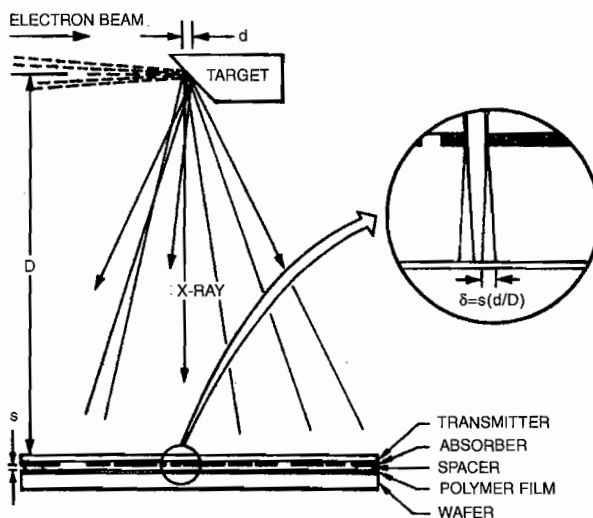
Charged-Particle-Beam Lithography

Introduction

The following is a brief introduction to lithographies based on charged-particle beams. Charged-particle vs. non-charged-particle approaches are compared in Figure 1.38. Both narrow-beam direct writing and flood exposure projection systems will be considered. The mask fabrication process is significantly simpler in the case of narrow-beam lithography as compared to flood-exposure-based lithography. In direct write systems, the computer-stored pattern is directly converted to address the writing particle beam, enabling the pattern to be exposed sequentially, point by point, over the whole wafer. In other words, the mask is a *software mask*. Electron-beam (e-beam) and ion-beam (i-beam) lithographies involve high current density in narrow electron or ion beams. The smaller the beam sizes, the better the resolution, but more time is spent writing the pattern. This sequential (scanning) type system exposes one pattern element or pixel at a time. Within that area, the charged-particle beam delivers maximum current (i), which is limited

Figure 1.42 Fabrication of a silicon-membrane-based x-ray mask with a gold absorber pattern. For use in high-aspect-ratio micromachining, the gold absorber layer must be between 5 and 15 μm. (Based on I. Brodie and J.J. Muray, *The Physics of Microfabrication,* Plenum Press, New York, 1982.[28])

primarily by the source brightness and column design. The experimental setup imposes a limit on the speed at which the writing beam can be moved and modulated, resulting in a "flash" time in seconds (t). The maximum dose (in coulombs per square centimeter) deliverable by a particular beam is given by:

$$D_{max} = \frac{it}{A} . \qquad (1.39)$$

with A the pixel area in cm^2. It will then be necessary to work with resists that react sufficiently fast at D_{max} to produce a lithographically useful, three-dimensional image (latent or direct image). The e-beam method displays a large depth of focus, as active focusing over various topographies is possible. The continued development of better charged-particle-beam sources keeps widening the possibilities for nanoscale engineering through lithography, etching, depositing, analyzing, and modifying a wide range of materials, well beyond the capability of classical photolithography. Table 1.7 lists some ion-beam and electron-beam applications.

TABLE 1.7 Electron- and Ion-Beam Applications

Electron-beam applications	Ion-beam applications
Nanoscale lithography	Micromachining and ion milling
Low-voltage scanning electron microscopy	Microdeposition of metals
Critical dimension measurements	Maskless ion implantation
Electron-beam-induced metal deposition	Microstructure failure analysis
Reflection high-energy electron diffraction (RHEED)	Secondary ion mass spectroscopy
Scanning auger microscopy	

Flood exposure of a mask in a projection system (that is, parallel exposure of all pattern elements at the same time, as we saw in DUV) is possible with ion and electron beams as well. In principle, these methods can take advantage of the excellent resolution of charged-particle beams while providing the throughput levels necessary for IC manufacture. Exposure masks are fabricated from heavy metals on semitransparent organic or inorganic membranes. The high cost of mask fabrication and the instability of the mask due to heating have postponed commercial acceptance of these high-energy exposure systems. Moreover, with ion and electron beams, flood exposure is limited to chip-size fields due to difficulties in obtaining broad, collimated, charged-particle beams. From the high-energy sources, only x-ray flood exposure and EUV, applicable over a large wafer area, approach the commercial application stage. The most prevalent use of charged-particle beams remains the scanning mode. We will learn how scattering with angular limitation projection electron-beam lithography (SCALPEL) and ion projection lithography (IPL) are starting to address these roadblocks.

Electron-Beam Lithography

Overview

Electron-beam lithography (EBL) is a high-resolution patterning technique in which high-energy electrons (10 to 100 keV) are focused into a narrow beam and are used to expose electron-sensitive resists. A typical setup involves 30 keV electrons from a Hitachi HL-700F EBL. The e-beam lithography method, like x-ray lithography, does not limit the obtainable feature resolution by diffraction, because the quantum mechanical wavelengths of high-energy electrons are exceedingly small. E-beam lithography exhibits some other attractive attributes compared to photolithography. These include

1. Precise control of the energy and dose delivered to a resist-coated wafer
2. Deflection and modulation of electron beams with speed and precision by electrostatic or magnetic fields
3. Imaging of electrons to form a small point of <100 Å as opposed to a spot of 5000 Å for light
4. No need for a physical mask; only a *software mask* is required
5. The ability to register accurately over small areas of a wafer
6. Lower defect densities
7. Large depth of focus because of continuous focusing over varying topography

At 30 keV, electrons will travel on average > 14 μm deep into a PMMA resist layer.

Some of the disadvantages of electron-beam lithography include

1. Electrons scatter quickly in solids, limiting practical resolution to dimensions greater than 10 nm.
2. Electrons, being charged particles, need to be held in a vacuum, making the apparatus more complex than for photolithography.
3. The slow exposure speed—an electron beam must be scanned across the entire wafer (for a 4-in wafer with a high feature density, this requires ~1h).
4. System cost is high.

The resolution of e-beam lithography tools is not simply the spot size of the focused beam; it also is affected by scattering of the e-beam inside the resist and substrate and by backscattering from the substrate exposing the resist over a greater area than the beam spot size. *Proximity effects* are created by scattered electrons, partially exposing the resist far beyond the point of impact. Line width variations due to local feature density are an immediate result. Proximity correction algorithms are used to achieve more uniform resist exposure with EBL. Such corrections are computer intensive and time consuming, however, and make a slow technique even slower.

As a result, the use of electron-beam lithography has been limited to mask making and direct writing on wafers for specialized applications, for example, small batches of custom ICs. As a research solution, several groups have been "Rube Gold-

berging" their standard scanning electron microscope to create customized electron-beam writing systems. Rosolen, for example, modified a Hitachi S2500 with a specifically built pattern generator and alignment system. The instrument does not require alignment marks on the sample and is able to compensate for positional errors caused by the sample stage and mask tolerances.[94]

Writing with an e-beam can be additive or subtractive. As an example of additive e-beam writing, e-beam-induced metal deposition from a metal organic gas [e.g., W deposition from $W(CO)_6$] has been used for the formation of microstructures of various geometries (see Figure 7.21). These devices are made one by one rather than in a large batch. Usually, this type of slow, expensive fabrication technique prohibits commercial acceptance. Some microstructures, especially intricate microsystems, might be worth the higher price tag. In that case, serial microfabrication techniques may not necessarily be as prohibitive as they would be in the case of ICs. In Chapter 7, direct write microfabrication methods are reviewed in more detail.

Electron-Beam Resists

Numerous commercial e-beam resists are produced for mask making and direct write applications. Bombardment of polymers by electrons causes bond breakage and, in principle, any material can function as a resist. However, the important considerations include sensitivity, tone, resolution, and etching resistance. PMMA exemplifies an inexpensive positive e-beam resist with a high resolution capability and a moderate glass transition temperature T_g (114°C), and Microposit SAL601 is an often used negative e-beam resist. The latter, being novolak based, has much better dry etch resistance than the former. The same materials act as an x-ray resist as well. This is not coincidental, as there is a strong relation between x-ray and e-beam sensitivity. A copolymer of glycidyl methacrylate and ethyl acrylate (COP) is another frequently used negative resist in mask manufacture. This material, although exhibiting good thermal stability, has (as is typical for acrylates) poor plasma-etching resistance. The measured $G(x)$ value of representative polymers in the COP family is about 10. Whelan et al. developed a low-energy electron beam top surface imaging chemically amplified (AXT) resist.[95] This AXT positive-mode resist incorporates a poly(hydroxystyrene) base resin and achieves sub-100 nm resolution with 2 keV electrons and features a sensitivity below 1 μC/cm².

Streblenchenko et al. patterned FeF_2 and CoF_2 with 100-keV electrons from a scanning transmission electron microscope (TEM) using a 0.5 nm diameter electron probe.[96] The fluoride resist films are prepared by thermal evaporation on thin carbon films. During electron bombardment, fluorine escapes, and the transition metal coalesces. The resolution in a 20-nm thick, very small-grained CoF_2 film is about 5 nm. An electron dose of 1000 C/cm² at 100 keV removes 90% of all the fluorine in the 20 nm thick CoF_2 film. Arbitrarily shaped nanometer scale magnetic structures have been written in such CoF_2 films.

Electron Emission Sources

Electron emission (field, thermionic, and photoemission) (Inset 1.22) underwrites the principle for electron emission source

---

Field and thermionic emission and photoemission

(From Lindquist et al., *Research and Development*, June, 91–98, 1990. With permission.)

Electron emission in a water bucket

THE THREE MECHANISMS used by field emission sources all basically involve emitting electrons and ions from a metal surface under the influence of a strong electric field.

Understanding these mechanisms is where the water bucket comes in.

In this analogy, the water level in a bucket represents the Fermi level—the highest occupied energy level in a cathode material. The work function is the energy required to get the water droplets (electrons) from the top of the liquid out of the bucket. This is the distance equivalent to the potential energy barrier.

In photoemission, photon energy excites electrons at the Fermi level of the cathode material and can impart enough kinetic energy to allow the electrons to escape from the bucket.

In thermionic emission, heat thermally excites the electrons, providing enough energy to boil the electrons off and out of the bucket.

In field emission a high electric field can thin the side of the bucket enough so that the electrons can tunnel through it.
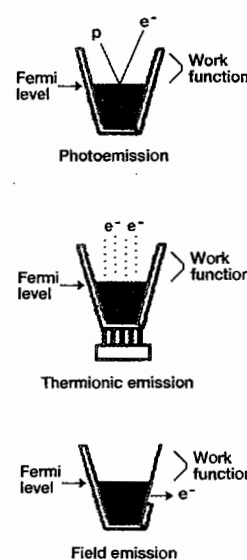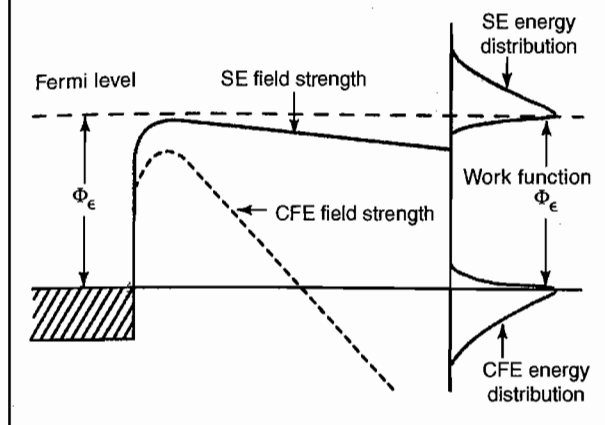
Inset 1.22

---

construction.[97] Schottky emission (SE) and cold field emission (CFE) have been in common use, especially for nanometer-sized beams for electron focusing systems. Emission of electrons from a metal under the influence of a field occurs in both SE and CFE. During SE emission, a blunt tungsten emitter tip coated with a low work function material (ZrO) is heated to 1800°K, and thermionic emission takes place; that is, heat thermally excites the electrons enough to bring them out of the material. In other words, the field emission is helped along by thermal excitation of the electrons. During CFE, a much smaller tungsten wire (radius <0.1 μm) is used, and a very high field causes electrons to tunnel out of the material. In CFE sources, electrons tunnel from various energies below the Fermi level. With SE cathodes, thermally excited electrons (nontunneling electrons) escape over a field-lowered potential energy barrier. Both SE and CFE sources display similar energy spreads, but their energy distributions are mirror images (Inset 1.23).[97]

Attaining high current levels in a submicron electron beam at low voltages (500 eV to 1 keV) is of interest for e-beam lithography as well as in scanning electron microscopy (SEM). When sensitive biological samples or electron beam sensitive resists are involved, SEM pictures must be made at voltages below 1 keV. At these low voltages, high currents are required to attain the needed detail and to minimize edge effects. Traditional SEM cathodes using tungsten hairpin filaments are very limited at low voltages; they cannot supply enough current in submicron beams at the low-voltage end. The high brightness SE and CFE sources, including LaB_6 and CeB_6, can do the job

SE and CFE energy distributions

Fermi level — SE field strength — SE energy distribution

Φ_ϵ

CFE field strength

Work function Φ_ϵ

CFE energy distribution

Inset 1.23

(Inset 1.24). SEM instruments with beam diameters of 1 nm have been made using such cathodes. For further reading about e-beam lithography, refer to Brewer.[98]

Micromachined Electron Emission Sources

There have been many attempts at decreasing the accelerating voltage in electron-beam lithography (EBL), since lower-energy electrons have a more confined lateral backscattering range. Moreover, it was found that low-energy EBL leads to higher sensitivity—that is, for a given resist, a lower exposure dose is required than with higher energy electrons.[99] Low-energy e-beams require thinner resist layers, as the penetration depth falls off very quickly (e.g., in a typical resist, 2 keV will only penetrate about 120 nm). Micromachining is ideally suited to deliver these low-energy electron-writing tools. We present only some examples here of micromachined e-beam tools. A first example involves the work by Zlatkin et al., who developed an elegant array of focused electron writing beams operating at 300 eV or less.[100] The emitters used are cold field-emission (CFE) sharpened tungsten tips, although thermionic or Schottky emitters would be feasible as well (for a description of different types of emission sources, see above). The emitters are positioned several millimeters above the micromachined extraction holes of a lens array fashioned in a single crystal substrate. Because of the relatively large distance between the emitter and the extraction anode (a 1 μm aperture), no precise alignment is necessary. The setup and a detail of the lens system are shown in Figure 1.43.

Each lens has a total area of 1 × 1 cm on a 500-μm thick Si wafer and is principally built up from two 0.5-μm thin Si layers. The two thin Si layers are separated from each other and the Si substrate by insulating silicon oxide. An anisotropic etch into the single crystal Si side opposite the thin layered sandwich structure exposes the oxide/Si/oxide/Si/oxide thin layered structure. A short isotropic etch of the exposed oxide is followed by a 1-μm dry-etch bore in the first thin Si layer (first aperture, 1 μm diameter). The small aperture is used to extract electrons from the emitter and to spatially confine the initial electron beam. Another isotropic etch through the second oxide is followed by the larger dry etched aperture (50 μm) in the second thin Si layer. This second aperture is used to focus the electron trajectories and is well aligned with the first aperture. By changing the voltage on the thin Si layers, incorporating extraction and focusing electrodes, the focal length can be changed. To deflect the electron beam for scanning, a set of four electrodes symmetrically placed around the extraction and focus apertures is energized. The four electrodes are placed on top of another insulator layer on the focusing membrane. Applying potentials to opposite pairs of deflection quadrants deflects the electron beam laterally, and the focused spot can be swept over an area either to create an SEM image or to pattern a design in EBL mode. The aim is to eventually be able to image and write in the sub-100 nm range. Imaging in the 30 nm range was demonstrated already, but writing is still quite inadequate (so far limited to >200 nm).

Comparison of electron sources

| Parameter	Field Emission (Shottky ZrO/W)	Field Emission (Cold)	LaB$_6$	CeB$_6$	Tungsten Filament
Brightness (A/cm^2 sr)	5×10^8	10^9	10^7	10^7	10^6
Emitting surface area (μm^2)	>0.3	0.03	>>1	>>1	>>1
Typical service life (hr)	5000	2000	1000	1500+	100
Operating vacuum	10^{-8}	10^{-10}	10^{-7}	10^{-7}	10^{-5}
Energy spread (eV)	0.3–1.0	0.2–0.3	1.0	1.0	1.0
Evaporation rate (g/cm^{2}s)			2.9×10^{-9}	2.1×10^{-9}	NA
Work function (eV)			2.6	2.4	4.5
Short-term beam stability (% RMS)	<1	4 to 6	<1	<1	<1
Standard cost			$775	$850	$20

Inset 1.24
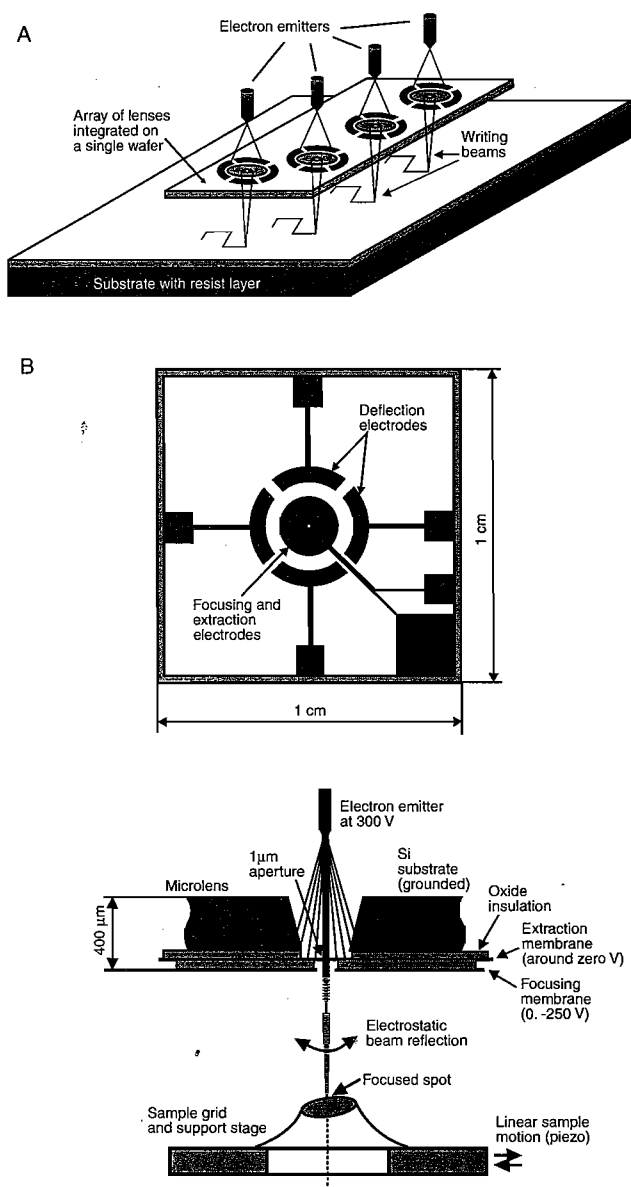
Figure 1.43 Array of focused electron writing beams operating at 300 eV or less (A). Setup and detail of the lens system (B). (Redrawn from Zlatkin and Garcia, *Microelectron. Eng.*, 46, 213–217, 1999.[100]).

Micromachining presents other potential solutions to make "nanolithography" a cost-effective and adequately fast proposition in the future. The next example addresses the issue of speed in writing with small e-beam systems by introducing massive parallelism. At Cornell's National Nanofabrication Facility (NNF), a research group has been working on arrays of microfabricated, miniaturized electron lithography systems based on scanning tunneling microscopes (STMs). In this STM aligned field emission (SAFE) system, the physical dimensions of the electron beam column (length and diameter) are in the range of millimeters. A field emission tip is mounted onto an STM; the STM feedback principle is used for precision x, y, and z piezoelectric alignment of the tip to a miniaturized electron lens to form a focused probe of electrons as shown in Figure 1.44.[101] Since many electron-optic aberrations scale with size, microfab-
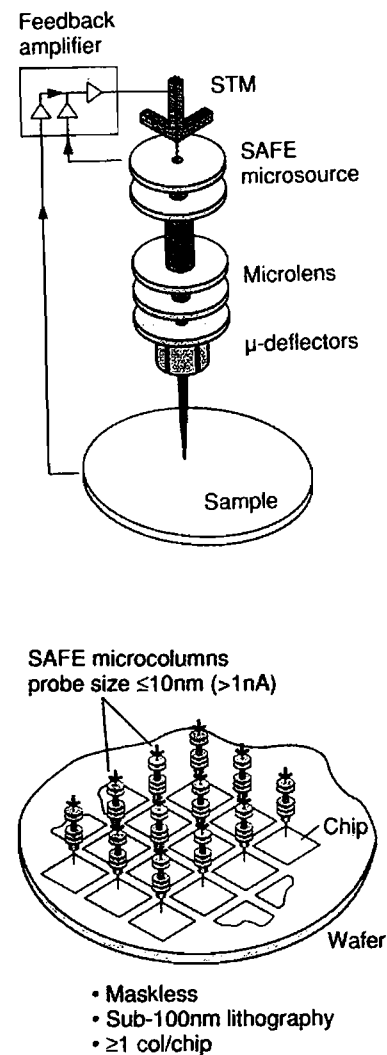

Figure 1.44 A micro column based on STM aligned field emission (SAFE) and arrayed micro-column lithography. (From an editorial in *Solid State Technol.*, 36, 25–26, 1993.[101] Reprinted with permission.)

rication techniques enable lenses with negligible aberrations, resulting in exceptionally high brightness and resolution. STM controls also allow for stability of the emission by automatically adjusting the z-position through a piezoelement.

An array of these micro columns, each with a field emission tip as the source and with individual STM sensors and controls, can generate patterns in parallel, one or more columns per chip (see Figure 1.44). The low voltage of operation of these tips obviates the need for proximity effect corrections, as low voltage operations have proven to eliminate proximity effects.[101,102] For example, low-energy electrons (15 to 50 eV) from an STM have been used to write patterns with 23 nm feature sizes, more than four times smaller than can be written on the same substrate and in the same resist with a tightly focused 50 kV e-beam.[102] Using a similar approach to the Cornell team, Wada et al. are well on their way to demonstrating arrays of ~10^4 to ~10^6 micromachined STMs on a nanolithographic subsystem.[103,104] About 25 of these subsystems will eventually form a 50-wafer-per-hour nanolithography system.

Nanolithography with ultrasharp field emitters is discussed further under scanning probe lithography (SPL), although it is an emission technology rather than a tunneling one. Tunneling is the mechanism in scanning probe lithography, the key difference being the distance between tip and substrate (1 nm with SPL vs. 100 nm or more with sharp field emitter tips) and the electron generation mechanism (tunneling vs. emission). The reason ultrasharp field emitters are discussed in that context is to compare their performance with that of SPL methods.

In addition to the application described above, micromachined electron emission sources are also used or are intended to be used for flat panel displays, high-temperature and high-radiation amplifiers, vacuum gauges and RF oscillators and amplifiers.[105]

Scattering with Angular Limitation Projection Electron-Beam Lithography (SCALPEL)

Scattering with angular limitation projection electron-beam lithography (SCALPEL) is a projection electron beam technique employing a 4× reduction and a step-and-scan writing strategy. The mask is a continuous thin membrane (~100 nm) made of a low atomic number material such as silicon nitride, patterned with a 25- to 50-nm high atomic number material such as W. The thin silicon nitride membrane is supported by thick silicon struts to minimize image distortion. High-energy electrons (~100 keV) pas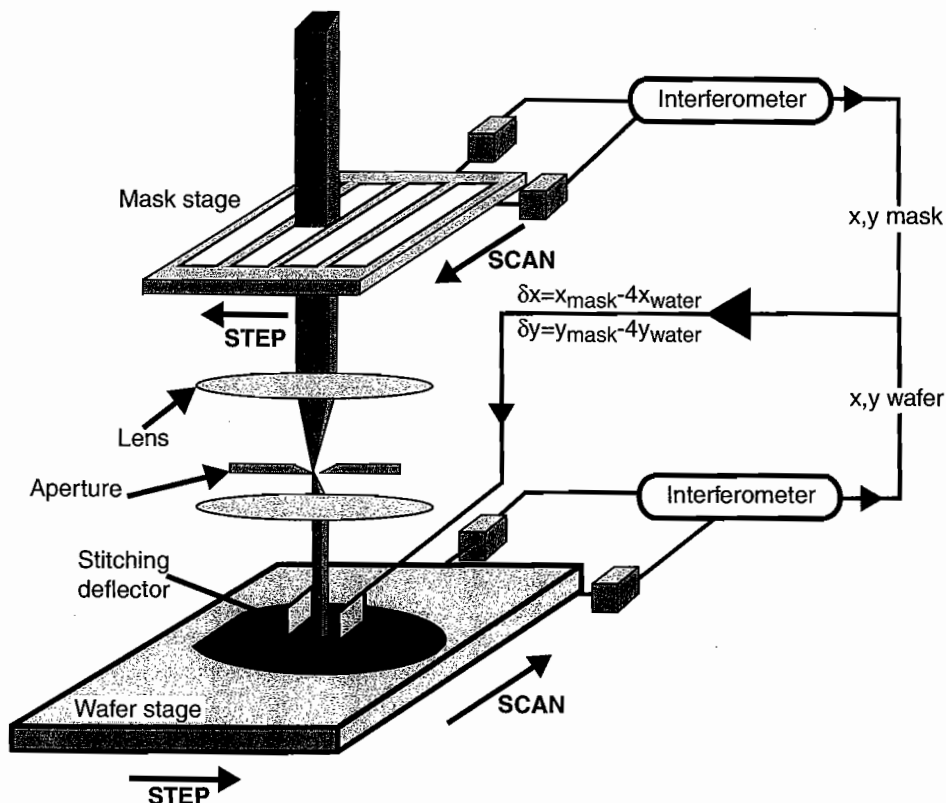s through both membrane and W pattern with minimal energy loss, leading to minimal heating effects. Electrons passing through the pattern layer are scattered to large angles, while those passing through the membrane material only suffer very little scattering. A narrow aperture blocks the scattered electrons resulting in a high-contrast image at the wafer level. The aperture rather than the mask absorbs unwanted energy. Today, image field is still limited to 1×1 mm at the mask level. The small images at the wafer are assembled (also stitched together) from the individual portions on the mask through a combination of interferometrically monitored mechanical motions of the mask and wafer stages and electronic deflections of the beam (Inset 1.25). The relatively low-cost masks in SCAL-PEL ($27,000 vs. $40,000 for a phase shift mask (PSM), $40,000 to $59,000 for an EUV mask, and $5,000 to $10,000 for standard binary masks) may make this technique a cost-effective solution for lithography at 130 nm and below.[10,106]

Ion-Beam Lithography

Introduction

In ion-beam lithography, resists are exposed to energetic ion bombardment in a vacuum. Ion-beam lithography consists of point-by-point exposures with a scanning source of liquid gallium metal and in flood exposure or ion projection lithography (IPL) with H^+, He^{2+}, or Ar^+ through thin membranes with holes (stencils).

SCALPEL writing strategy involves step-and-scan writing (see also Inset 1.12)

Interferometer

Mask stage

SCAN

x,y mask

$\delta x = x_{mask} - 4x_{water}$

STEP

$\delta y = y_{mask} - 4y_{water}$

Lens

x,y wafer

Aperture

Interferometer

Stitching deflector

Wafer stage

SCAN

STEP

Inset 1.25

Focused Ion Beam and Deep Ion-Beam Lithography

For ion-beam construction, liquid metal ion (LMI) sources are becoming the choice for producing high-current-density submicrometer ion beams. With an LMI source, liquid metal (typically gallium) migrates along a needle substrate. A jet-like protrusion of liquid metal forms at the source tip under influence of an electrical field. The gallium–gallium bonds are broken under the influence of the extraction field and are uniformly ionized without droplet or cluster formation. LMI sources hold extremely high brightness levels (106 A/cm² sr) and a very small energy spread, making them ideal for producing high-current-density submicrometer ion beams. Beam diameters of less than 50 nm and current densities up to 8 A/cm² are the norm. In addition to Ga, other pure element sources are available, such as indium and gold. By adopting alloy sources, the list expands to dopant materials such as boron, arsenic, phosphorus, silicon, and beryllium.

As in the case of e-beam systems, ion-beam lithography offers direct write and flood exposure fabrication opportunities. Compared to photons (x-rays and light) or electrons, ions chemically react with the substrate, allowing a greater variety of surface modifications such as patterned doping. The resolution of ion-beam lithography is better than for electrons, because the secondary electrons produced by an ion beam are of lower energy and have a short diffusion range so that hardly any backscattering occurs. The ion-beam spot size is the smallest possible— smaller than UV, x-ray, or electron-beam spots. The smallest focused ion beam (FIB) spot currently reached is about 8 nm, accomplished by using a two-lens microprobe system and a single-isotope gallium ion source. With this setup arrays of dots were produced in a 60 nm thick PMMA layer with dot dimensions ranging from 10 to 20 nm.[107] Ion-beam lithography experiences the same drawbacks as an electron-beam system in that it requires a serially scanned beam and a vacuum.

Focused ion beams can be used to perform maskless implantation and metal patterning with sub-micrometer dimensions. Focused ion beam also has been applied to milling in IC repair, maskless implantation, circuit fault isolation, and failure analysis (see Table 1.7). Some micromachining applications of ion-beam technology will be reviewed in Chapters 2 and 7. As a machining tool, FIB is very slow. Except for research, it may take a long time to become an accepted "micromachining tool." For additional reading on ion-beam lithography in general refer to Selinger[108]; for more specific reading on focused ion-beam-induced deposition, see Brodie.[28]

Using high-energy (2 MeV) protons deep ion-beam lithography (DIBL) in PMMA produces submicron (300 nm) walls with an aspect ratio approaching 100. Three-dimensional complex microstructures with smooth walls and corners have been produced this way. The range of 2 MeV protons in PMMA is 63 μm.[109] Multiple exposures at different ion energies (e.g., 0.6 and 2 MeV) allow production of multilayer structures in single layer resists such as SU-8.[110]

Ion Projection Lithography

Ion projection lithography (IPL) is another of the candidates for high-throughput lithography dedicated to future 50 nm and sub-50-nm IC generations. Protons are generated by a radio frequency driven filament. Efforts are especially strong in Europe in this area (e.g., at IMS, http://wwwold.ims-chips.de/). The ion flood lithography mask typically consists of a silicon membrane a few microns thin (3 μm) with pattern openings that allow protons to pass through.[111] The mask fabrication process involves silicon-on-insulator (SOI) and dry etching.[112] Current IPL approaches, like SCALPEL, are 4× technologies. Depth of focus is large and may reach up to 500 μm. The optics in IPL are all electromagnetic, and the potential for a 50 × 50 mm exposure field exist.

Comparison of Ion-Beam Lithography with E-Beam Lithography

Ion-beam lithography has at least two advantages over electron-beam lithography: (1) it has almost two orders of magnitude higher resist sensitivity, and (2) it has negligible ion scattering in the resist and very low backscattering from the substrate. A major problem is the potential for damage to sensitive electronic functions from the high energy ions.

Emerging Lithography Technologies

Introduction

In this section, we cover the emerging fields of proximal probe lithography, very thin to monolayer lithography, soft lithography, and 3D lithographies including holographic lithography, stereolithography, and lithography on nonplanar substrates using high-precision linear and rotating positioning stages. These are more futuristic lithography methods, some of which could cause a paradigm shift in the improvement of CD printing capability by perhaps a factor of 100. These would enable devices that are a few nanometers in size and have switching speeds in the terahertz range. More details and additional references on the topics covered in this section can be found in the thorough review article by Xia et al.[113]

Scanning Probe Lithography

STM/AFM Background

The scanning tunneling microscope (STM) images the surface of conducting materials with atomic-scale detail. STM was invented by Gerd Binnig and Heinrich Rohrer of IBM's Zurich Laboratory in 1985 (they received the 1986 Nobel Prize for their invention).[114] In general, STM works by bringing a small conducting probe tip up to a conducting surface. When the probe is very close to the surface (less than 10 Å) and operating voltages in the ±10-V range are applied, very small currents are produced, because the electrons in the probe and the surface have wave functions extending beyond the physical surface boundaries. To the extent that these spillover wave functions overlap, a measurable current results. The interesting part about this current is that it depends exponentially on the spacing between the two conductors (as well as the voltage). A piezoelectric transducer accomplishes the z-axis distance variation between tip and sam-

ple. By changing the distance over 1 Å, the current changes by a factor of ten. In practice, the current is kept constant through a feedback mechanism, and the probe moves up and down over the surface following the atomic contours it "sees." The images produced by the STM come from the electronic structure as well as from the geometry of the sample. Up to 100 times more powerful than SEMs, scanning tunneling microscopes measure objects in the angstrom range. Over time, many proximal probe were developed, and STM belongs now to a large new family of very local, proximal probes, such as atomic force microscopes (AFMs), scanning electrochemical microscopes (SECMs), scanning thermal microscopes, scanning capacitance microscopes, magnetic force microscopes, scanning pH probes, etc., enabling microscopy of almost any type of material and property. The common feature of these instruments is that their resolution is not determined by visible light used for the interaction with the probed object, as in conventional microscopy.[115]

Scanning Probe Lithography

Introduction

Lithography utilizing electrons from a scanning probe offers several potential advantages over writing with a traditional e-beam source.[115] One important benefit is that the low energy of electrons in SPL (<50 eV) as compared to those in EBL (300 eV to 100 keV) avoids the detrimental effects of electron back-scattering, thereby virtually eliminating proximity effects and thus producing enhanced resolution and superior pattern fidelity.[116] Due to the small tip-to-sample distance, extremely small spot sizes are achievable so that the exposure dose may be confined to a beam diameter of less than 10 nm. The method enables a wider exposure latitude than EBL, a fact demonstrated in Figure 1.45.[117] In this figure, line width is plotted vs. dose for SAL601 resist using both EBL and SPL lithography systems. From the lower slope of the SPL curve, one deduces that SPL has higher dose latitude than EBL—in other words, SPL is less

sensitive to dose variations. On the other hand, SPL is less sensitive and does require a higher dose to write the same feature size. The mechanism of electron bombardment in SPL is very different from that in EBL. In EBL, the mean free path of the bombarding electrons is long compared to the resist thickness. In contrast, low-energy electrons in SPL have a mean free path below 2 nm and reach through the resist under the influence of an electrical field, undergoing a number of scattering events before reaching the resist /substrate interface.[117] The latter makes it clear why a thinner resist will result in a better expected resolution. The influence of secondary electrons (i.e., all electrons emanating from the resist/substrate) in proximal probe lithography was investigated recently by Völkel et al.[118]

Modes of Pattern Generation in Scanning Probe Lithography

The Naval Research Laboratory (NRL) worked with proximal probes to pattern thin films of chemically amplified negative e-beam resist (SAL-601 from Shipley). Resist films of 30 to 70 nm thick were patterned with typical tip-sample voltages from –15 to –35 V, resulting in minimum feature sizes of 23 nm.[116] Using self-assembled monolayers (SAMs) as resists (see also below), it was shown that the lower the exposure threshold energy of the film, the better the lithographic resolution. There are indications that, with these low voltages (~4 V), SAM resists will eventually yield sub-10 nm CDs. In an alternative approach, surface oxides were induced on Si and GaAs by slightly increased tip voltages on samples held in a wet nitrogen atmosphere.[102] These thin oxides, although only a few monolayers thick, are sufficiently robust to act as a mask for subsequent reactive ion etching of the substrate.[116] The oxidation process is fairly general and may also be applied to Ti and Cr. At NRL, oxide features with lateral dimensions as small as 10 nm have been achieved.[116] Multilayer resist films for nanopatterning with SPL have also been developed. Sugimura et al. worked with a three-layer resist to pattern insulating substrates such as thermally grown SiO_2 with current injection from a scanning probe tip.[119] The process sequence is sketched in Figure 1.46.

The bottom layer of the three-layer resist consists of 20 nm of amorphous Si (a-Si) and is prepared by ion-beam sputtering; a second layer consists of an intermediate 2 nm thick Si oxide and is prepared by photo-oxidation of the top layer of the amorphous Si. The top resist layer is 2 nm thick and consists of an octadecylsilyl self-assembled monolayer (ODS-SAM). To pattern the resist on top of the insulating SiO_2, a bias voltage is applied between the AFM probe and the conductive a-Si layer, which is biased positively. When a scanning probe is operated in the presence of atmospheric water vapor, the probe and sample are automatically connected via a minute water column created by capillary forces of the adsorbed water. This assembly serves as a minute electrochemical cell. As a result of the electrochemical anodization reaction, the monolayer resist becomes degraded in the region where the probe has passed, and the underlying photo oxide grows thicker. A subsequent etch in 0.5 wt.% HF removes the exposed oxide, and that pattern is then transferred to the underlying a-Si by an etch of the a-Si in an aqueous solution of 25 wt.% tetramethylammonium hydroxide
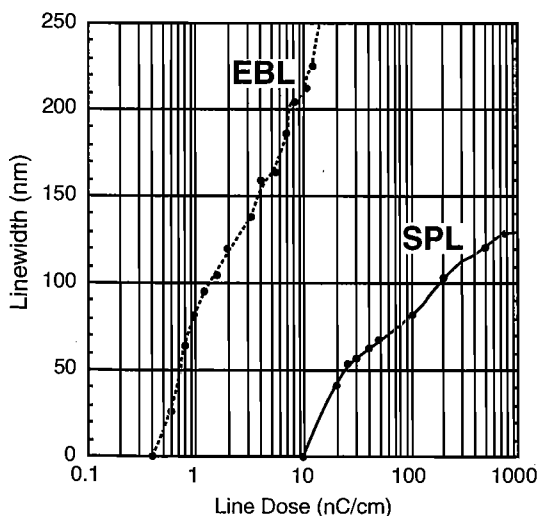

Figure 1.45 Line width vs. dose for SAL601 resist using both EBL and SPL lithography systems. (From K. Wilder et al., *J. Vac. Sci. Technol.*, B16, 3864–3873, 1998.[117] Reprinted with permission.)

A

- SAM
- Photochemical-SiO$_2$
- a-Si
- Thermal-SiO$_2$
- Si Substrate

B

Conductive AFM probe

Scan

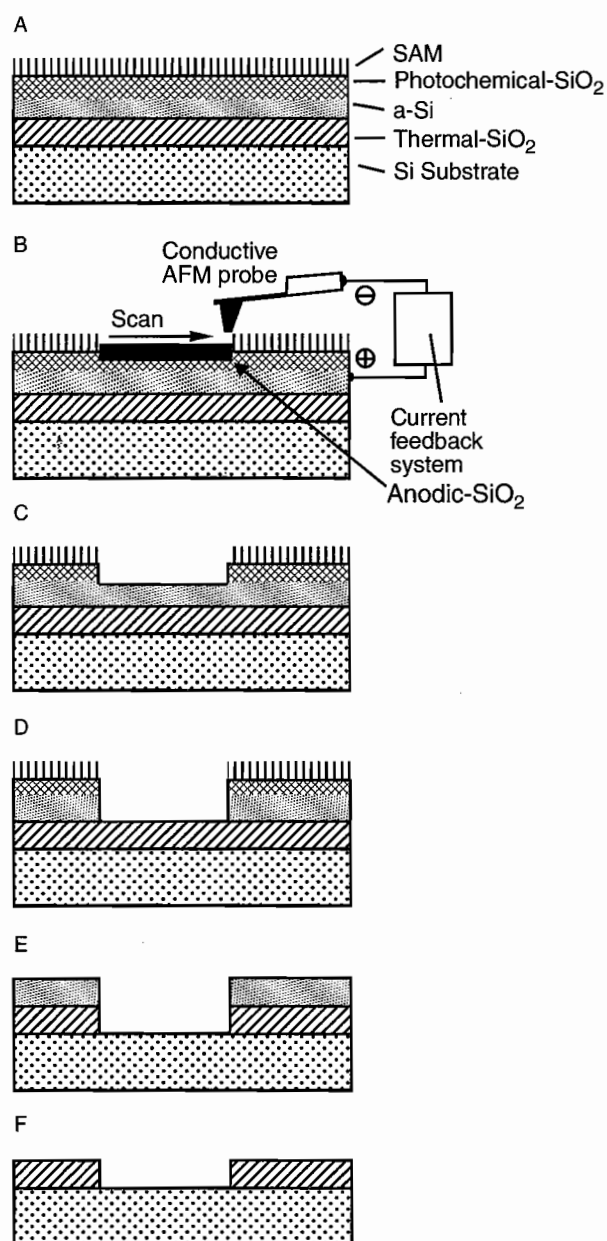Current feedback system
Anodic-SiO$_2$

C

D

E

F

Figure 1.46 Schematic of AFM lithography on SiO$_2$ using a multilayered resist system. (A) Cross section (layer thickness not to scale). (B) Exposure by drawing patterns into the ODS-SAM layer by current injection from an AFM probe. (C) First development step by HF etching to remove the SiO$_2$ just formed. (D) Second development step by TMAH etching to remove the a-Si in the exposed area. (E) Pattern transfer with HF to remove the thermal oxide in the exposed area. (F) Resist removal i.e., removal of all the remaining a-Si in TMAH etch. (Based on H. Sugimura et al., *J. Vac. Sci. Technol.*, vol. B17, 1999.[119] Reprinted with permission.)

(TMAH). The latter solution etches Si but not SiO$_2$. The amorphous Si is etched, and the etch stops when the underlying insulator is reached. The underlying thermal oxide can now be etched in HF and the resist stripped. Minimum feature sizes reached so far are about 50 nm.

Proximal probes may also be used to mechanically cut (scratch) a trench through a mask layer. A trench may be cut by simply drawing an AFM tip across a surface with enough force, or the piezomechanism driving the AFM cantilever may be modulated in the z direction to tap the tip on the surface as it is drawn over it. Cuts as narrow as 20 nm and 2 nm deep have been made in III-V semiconductor surfaces.[120] One difficulty with the scribing approach has been that probe tips often break following direct collision with a solid surface. As a possible solution, carbon nanotubes have been mounted on the tip of silicon cantilevers, enabling lines as narrow as 10 nm (Inset 1.26). Carbon nanotubes are constructed of rolled-up sheets of graphene made of benzene-type rings of carbon (Inset 1.27). The nanotubes may be single- or multiwall (concentric cylindrical shells) and have unique elastomechanical properties. They constitute the strongest material known along the axial direction and yet are highly elastic and flexible along the radial direction[121] (see also Chapter 7 Section 7.2, Table 7.12).

Writing speeds of an SPL tool of 0.1 to 100 µm/s are typical, and speeds of several thousands of micrometers per second have been demonstrated.[122] Very fast scanning will remain limited, though, as the close tip-to-substrate distance (typically 1 nm with STM) is prone to cause tip crashes. With ultrasharp tips (radii < 10 nm) in the emission mode rather than the tunneling mode, one may stay away as far as 100 nm from the surface. This approach not only enables faster writing (limited by the response of the piezo material only) without fear of crashing the tip as frequently, but the high beam currents also contribute to the possibility for exposure of large area patterns. Even at a

Nanotube mounted on the micromachined tip of a Si cantilever is used as a nanopencil for lithography of 20 to 10 nm lines

(Courtesy of Dr. M. Meyyappan, NASA Ames.) SEM picture of nanotube mounted on tip of Si cantilever (http://cnst.rice.edu/pics.html).
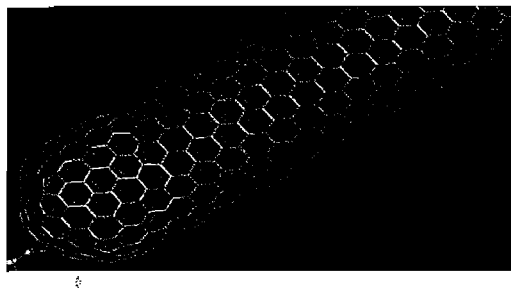
WWW.NANO.GOV

NASA Ames Research Center
Ramsey Stevens, Lance Delzeit, Cattien Nguyen
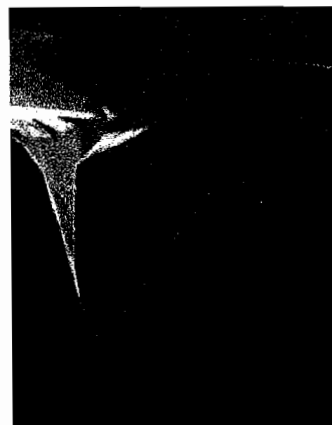
Inset 1.26

Single-walled carbon nanotube

(See also the Rice University web site at http://cnst.rice.edu/pics.html). Used with permission.

Inset 1.27

Figure 1.47 Atom manipulation for nanomatching. Top: a 60 × 48 Å STM image of four Pt adatoms assembled into a linear array on a Pt(111) surface. Pt atoms were herded four unit cells apart along a close-packed direction of the Pt(111) surface. Bottom: a 40 × 40 Å STM image of a compact array of seven Pt adatoms. (From J. A. Stroscio and D. M. Eigler, *Science*, 254, 1319–26, 1991.[125] Copyright 1991 The American Association for the Advancement of Science. Reprinted with permission.) For more pictures of STM atom manipulation, visit http://www. almaden.ibm.com:80/vis/stm/gallery.html.

distance of ~100 nm, the beam diameter remains small at ~30 nm.[123] Structures of lateral dimension of ~20 nm have been created in SAMs using ultrasharp field emitters.[124]

Atom Lithography or Mechanosynthesis

As we just learned, proximal probe equipment locally modifies surfaces[125] and is used to expose resists and to oxidize and mechanically scribe surfaces (see above and Inset 1.26). Also, a variety of direct atomic manipulations have been demonstrated with proximal probes.[125] The electric field strength in the vicinity of a probe tip is very strong and inhomogeneous (say, a field of 2 V Å^{-1} concentrated around the probe tip). This field can manipulate atoms, including sliding of atoms over surfaces and transferring atoms by pick (erase) and place (write). Drexel calls this type of machining *mechanosynthesis*.[126] These atomic manipulation processes may thus be classified as parallel processes and perpendicular processes. In parallel processes, an adsorbed atom or molecule is induced to move along the surface (sliding); in perpendicular processes, the atom or molecule is transferred from the surface to the tip of a proximal probe or vice versa. Patterning surfaces at the atomic level is demonstrated in Figure 1.47 (top) where a 60 × 48 Å image of four Pt adatoms is shown. The four atoms were herded into a linear array.[125] The bottom image in Figure 1.47 is one of seven Pt atoms compacted together. The theoretical resolution of a lithography technique based on these atomic probes is a single atom. In practice, lines of 100 Å in width have been written using STM. Even with a 100-Å resolution "only," a single memory bit could be stored say in an area that measures 100 Å on a side. This will enable bit storage of 10^{12} bits/cm^2 as compared to 10^9 bits/cm^2 with conventional technology.

One major drawback to bear in mind with atom placing or removing techniques for micromachining is the time involved in generating even the simplest of features. If one wanted to deposit a metal line 10 μm long, 1 μm wide, and 0.5 μm high, 10^{16} atoms would need to be manipulated. Even at a deposition rate of 10^9 atoms per second, this would take more than 100 days. In the late 1980s, it took an IBM team a week to spell out the IBM logo with an STM, putting Xe atoms in place on a Ni surface. In 1993, Kamerzki et al. succeeded in shortening the time considerably.[127] Using an atomic processing microscope

(APM), they imaged selected atoms on a surface, stripped them off, and replaced them with other atoms, all in a matter of minutes. The APM relies on several technical innovations such as photon biasing whereby a precise voltage and photon pulse is "tuned" to a specific atom to strip it from the surface (Inset 1.28). Optical biasing, also called *optical tweezers*, has been used to manipulate a variety of individual particles, and the technology will be reviewed in the *Micro- and Nanoassembly* section of Chapter 8. Using an STM or APM, the number of atoms manipulated is minuscule, and depositing or removing clusters of atoms and parallel processing will be essential for these approaches to become viable. Nature, working with similarly small building blocks (amino acids and proteins) to circumvent the time problem, has a great deal of redundancy and parallel processing built in.

Utilizing mechanosynthesis, Wada[128] describes an atom relay transistor (ART). The ART shown in Figure 1.48 consists of an atom wire, a switching gate, and a reset gate. The atom wire is ON when the switching atom is inside the wire and OFF when it is out of the wire. The switching speed of the ART should be on the order of several tens of terahertz, because the intrinsic vibration frequency of an atom is 100 THz. The time required to build a supercomputer based on these ART switches would be way too long, and the self-assembly of atoms on solid substrates must be achieved before the exciting prospect of an ART will be realizable.

Very Thin Resist Layers

Introduction

Radiation scattering during exposure of a resist limits resolution to no better than the resist thickness. Using ever thinner resist layers further improves the resolution and renders a low DOF—associated with high numerical aperture (NA) lenses and short wavelength lithography—less critical. It is in this context that

Atomic processing microscope

Schematic shows tunable pulsed lasers for photon biasing; catheter probe for imaging, processing, and analysis; and computer for device control and data acquisition. (Adapted from editorial in *Research and Development,* April, 17, 1993.)
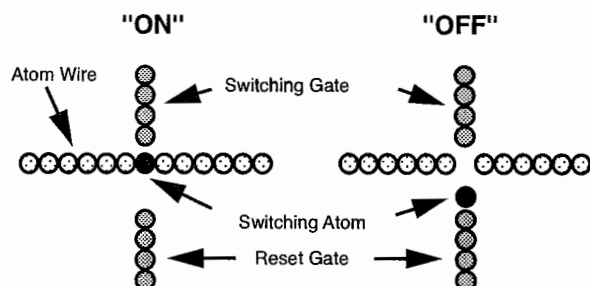
APM Permits Atom Removal and Placement

Inset 1.28

"ON"   **"OFF"**

Atom Wire   Switching Gate
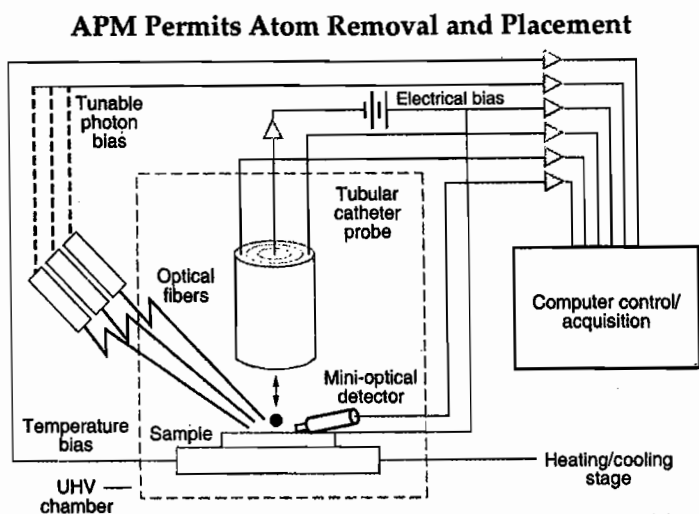
Switching Atom

Reset Gate

Figure 1.48   Schematic of the atom relay transistor (ART), consisting of an atom wire, switching atom and reset gate. (Based on Y. Wada, *Microelectron. Eng.,* 30, 1996.[128])

physisorbed Langmuir-Blodgett (LB) resists, self-assembled chemisorbed monolayer (SAM) resists, and ultrathin film (UTF) resists are explored.

Langmuir–Blodgett Resists

LB resist films were considered for some time for the finest lithography applications such as ultraresolution nanolithography. In this method, LB resist films are prepared by transferring organic monolayers floating on a water surface onto solid substrates. The procedure is explained in Chapter 3 (see Figure 3.30). Multilayer films of poly(methylmethacrylate) (PMMA) deposited by the LB technique, for example, have been investigated as a potential e-beam resist for nanolithography. The range of backscattered electrons limits the resolution of e-beam resists through the proximity effect, and nanometer thick layers minimize this effect. Despite extensive work in the early 1980s to automate LB resist deposition systems, the technology is not envisioned as a viable manufacturing option because of its low mechanical and chemical stability.[129,130] Other major problems concerning LB films need to be addressed before the method might become viable, and research in this area focuses on speeding up the coating process, simplifying surface cleaning procedures, and improving etching resistance. Since the bonding of an LB film with the substrate is based on physisorption, the

adhesion is weak, and chemisorbed self-assembled monolayers or SAMs have become much more popular (see next).

Self-Assembled Monolayers

Molecular self-assembly is a chemical process in which molecules spontaneously organize to form larger ordered structures. Self-assembly is one aspect of supramolecular chemistry, i.e., chemistry beyond the molecule or the chemistry of the intermolecular bond.[131] Self-assembly is studied extensively with an eye toward mimicking nature's spectacular use of this bottom-up, automated manufacturing approach (see Chapter 7 and Chapter 8). In one special type of self-assembly, self-assembling (SA) precursor molecules, from solution or vapor phase, react at interfaces to produce layers of monomolecular thickness that are chemically bonded to solid surfaces. Such layers belong to a class of materials known as *self-assembled monolayers.* The thermodynamically favorable bond formation involves chemisorption, which results in monolayers that are more stable than those possible with physisorbed LB films (see above); therefore, SAMs potentially make for better resist candidates. Homogeneous and densely packed molecular layers incorporate reactive groups that form bonds with the substrate on one side, while an organic group (e.g., an alkyl group, R) on the other side imparts the desired chemical functionality to the surface modified with the thin film. Changing even one atom at the end group of self-assembling molecules is sufficient to dramatically alter macroscopic properties such as wettability, biocompatibility, and adhesion. Fluorinated SAM precursors, for example, form monolayers ~1 nm thick with the same low wettability and resistance to adhesion typical of thick samples of Teflon®.[132] Many SAM films can be molecularly engineered to be patterned by various types of energetic radiation including DUV, soft x-rays, ion beam, and low-energy electrons.[116] Scratching (see above) and micro-contact printing (see below) can also be used to pattern SAMs. Structures below 20 nm have been obtained with STMs and conventional electron-beam lithography systems.[123] SAMs have successfully been deposited on metals like gold, aluminum, titanium, zirconium, silver, copper, and plat-

inum as well as on SiO_2, GaAs, and other surfaces.[132] The quality of the films depends strongly on surface pretreatment. Pinhole density on a gold surface of less than 5/mm² has been reported (see Müller[123] and references therein). Alkane-thiols and dialkyldisulfides are typical precursor materials for SAMs on gold, the most extensively studied substrate. The formation of long-chain ω-substituted dialkyldisulfides on gold was first demonstrated in 1983.[133] Films of better quality are formed by the adsorption of alkylthiols.[134] Alkanethiols and dialkyldisulfides are lipid-like organic molecules having the general formula $HS-(CH_2)_n-X$ and $X-(CH_2)-SS-(CH_2)_m-Y$, respectively, where n and m indicate alkyl chain length and X,Y the end groups ($-CH_3$, -azobenzene, -OH, etc.).[135] Organosilanes are often used to form SAMs on Si. These molecules form a Si-O-substrate siloxane bond on the Si surface and the alkyl group R is responsible for the ordered nature of the film. For n > 6 (n is number of carbon groups in the R chain), SAMs have substantial crystalline order at the air/monolayer interface.[132,135]

Micro-contact printing, discussed below, also takes advantage of SAMs, and so does DNA and protein patterning discussed in Chapter 3. Micro-contact printing can be considered a merging of top-down manufacturing techniques (i.e., traditional lithography used to make the elastomeric stamp) with bottom-up manufacturing (i.e., proteins used to ink the stamp as building blocks). Protein and DNA patterning has evolved into an important research application for lithography. It is easy to imagine arrays with several different enzymes, antibodies, or DNA-probes immobilized precisely onto a small transducer surface as a diagnostic panel for clinical applications. Different approaches for patterning organic materials are reviewed in Chapter 3.

Ultrathin Film Resist Layers

Polarity changes are at the heart of the newest ultrathin film resist strategies. UTF resists represent the next evolutionary step in surface imaging technologies (see earlier). In UTF, only the top monolayer of the resist changes, opening up the potential for further resolution improvement. Calvert et al. use a few monolayers (<10 Å) of organosilanes, which they chemisorb onto a substrate.[136] These chemisorbed films are easier to prepare and more robust than the physisorbed LB films (see above). Deep-UV irradiation of these films cleaves organo functional groups from the film and produces an extremely hydrophilic surface (water contact angle <10°) to which colloidal Pd/Sn catalyst does not adhere. Subsequent electroless deposition occurs only on those surface regions that were not irradiated. Patterns of electroless Ni, Co, and Cu are thus produced on a variety of substrates including silicon, quartz, alumina, metals, and polymers. These metal patterns serve as an efficient plasma-etching mask for pattern transfer. Fabrication of 0.3 μm line widths has been demonstrated in this way.[129,136]

Soft Lithography

Introduction

Soft lithography is another new tool in the nanofabrication arsenal.[137] It is the collective name for a set of new techniques: replica molding (REM), micro-contact printing (μCP), micro-

molding in capillaries (MIMIC), micro-transfer molding (μTM), solvent-assisted micromolding (SAMIM), and near-field conformal photolithography using an elastomeric phase-shifting mask.[113] All these methods use a patterned elastomer as a stamp, mold, or mask to generate micropatterns and microstructures instead of a rigid photomask. The method has made for a very exciting research tool and may offer advantages over conventional methods for patterning of nonplanar substrates, unusual materials, and large areas.

In soft lithography, a master mold is first made by lithographic techniques, and an elastomeric stamp (e.g., polydimethylsiloxane) is cast from this master mold. A simple example procedure for making a polydimethylsiloxane (PDMS) stamp from a photolithographically-patterned resist layer as master mold is outlined in Figure 1.49. A thin layer of SU-8 photoresist [SU-8 (50) from MicroChem, Newton, MA] is coated on a Si wafer. The resist is patterned by UV lithography. To make things really fast (<24 h) and inexpensive (<$20), the contact mask for the UV lithography may be a transparency on which, using drawing software such as Freehand, a design is printed with a high-resolution printer (>3300 dpi). After development, the photoresist is treated with (tridecafluoro-1,1,2,2-tetrahydrooctyl)-1-trichlorosilane (Hüls Chemicals, http://www.degussa.com/) vapor to facilitate PDMS removal once cured. A 10:1 ratio of a PDMS mix, PDMS oligomer and cross-linking agent (Sylgard 184, Dow Corning, http://www.dowcorning.com/) is cast on the photoresist film and cured for 1 hour at 60°C in an oven.[137]

PDMS has properties that make it suitable as a stamp material. It (a) provides a surface that has a low interfacial free energy (~21.6 dyn/cm), (b) is chemically inert, (c) is non-hygroscopic (does not swell with humidity), (d) passes gas easily, (e) has good thermal stability (~186 °C in air), (f) is optically transparent down to ~300 nm, (g) is isotropic and homogeneous,
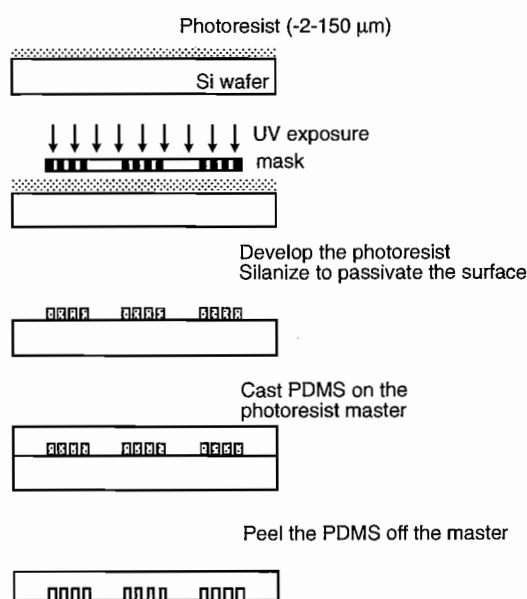

Figure 1.49 Making a PDMS mold. (Based on Y. Xia and G.M. Whitesides, *Ann. Rev. Mater. Sci.*, 28, 1998[138] and Y. Xia and G.M. Whitesides, *Angew. Chem. Int. Ed.*, 37, 550–575, 1998.[139])

(h) is durable (stamps may be used >50 times over several months without degradation), and (i) has interfacial properties that are easy to modify.[138]

Many PDMS stamps can be generated, and each one may be used many times. The rubber stamp can be used in a number of different ways. Some different soft lithography processes are outlined below. For further reading on soft lithography, consult Xia et al.[138,139] and visit the Softlithography home page at UW/WTC (http://www.engr.washington.edu/~cam/CAMsoft-lithhome.html).

Micro-Contact Printing

In micro-contact printing (μCP), the PDMS rubber stamp is coated with an ink of the molecules (say, alkylthiols) that one wants to print in selected patterns on a solid substrate. During stamping, only the raised parts of the stamp collect the "ink." The inking of the substrate consists of self-assembled monolayer formation on the solid surface by covalent chemical reactions. The inked areas are self-passivating and exhibit very low interfacial tension that repels additional molecular layers so that SAMs form only in areas of conformal contact between polymer and substrate. STM studies of the transferred SAMs reveal an achievable order indistinguishable from that found for SAMs prepared from solution.[132] The SAM pattern acts as a highly localized and efficient barrier to some wet etches. This lithographic technique—once the master is made—is not subject to diffraction or DOF limitations. The deformability of the elastomeric stamp allows it to accommodate rough surfaces. The method even works on spherical substrates (such as optical fibers or lenses) with radii of curvature of less than 10 μm. PDMS-based elastomers do not adhere to novolak or PMMA-based polymers, allowing convenient replication of masters formed by electron-beam lithography.[132] The technique has been used, for example, to build an antibody grating on a Si wafer by inking the rubber stamp with an antibody solution.[140] The antibody grating alone produces insignificant diffraction but, upon immunocapture, the optical phase change produces diffraction.

Micro-Transfer Molding

In micro-transfer molding (μTM), illustrated in Figure 1.50, the rubber mold is filled with a polymer precursor, and the rubber stamp is pushed up against a substrate (e.g., a Si wafer or another flat sheet of PDMS).[141] The polymer in the rubber stamp relief is cured and transferred to the substrate, and the stamp is peeled off. Schueller et al.[137] used this technique to make curved glassy carbon micro grids. In this case, a photoresist was used as polymer precursor and cured while sandwiched between a patterned PDMS rubber stamp and a planar PDMS sheet. To obtain the curved carbon grid, the PDMS sandwich was deformed against a curved cylinder during curing. After removing the two thin strips of PDMS, the freestanding and curved photoresist grid was carbonized at high temperature in an argon atmosphere (Inset 1.29).

Micromolding in Capillaries

In micromolding in capillaries (MIMIC), illustrated in Figure 1.51, the rubber stamp is pushed up against a substrate, and
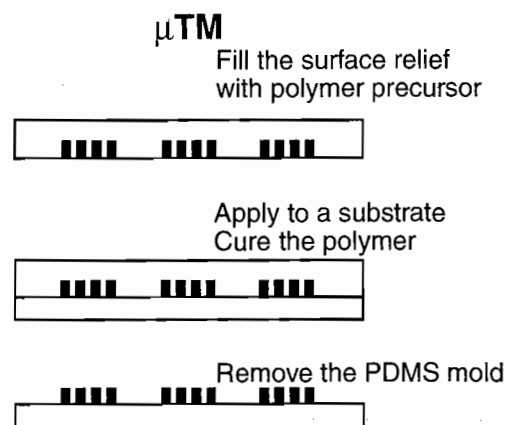
μTM

Fill the surface relief
with polymer precursor

Apply to a substrate
Cure the polymer

Remove the PDMS mold

Figure 1.50 Schematic description of micro-transfer molding or μTM. (Based on Y. Xia and G.M. Whitesides, *Ann. Rev. Mater. Sci.*, 28, 155–184, 1998[138] and Y. Xia and G.M. Whitesides, *Angew. Chem. Int. Ed.*, 37, 550–575, 1998.[139])

Curved glassy carbon structure[137]

(Courtesy of Dr. G. M. Whitesides, Harvard University.)

100 μm

Inset 1.29

liquid is applied to access holes in the mold. Sometimes this process is vacuum-assisted.[142] The liquid wicks into the cavities formed by the rubber mold against the substrate. The polymer is cured, and the stamp is removed. MIMIC has been used to fabricate all plastic field effect transistors[143] and in the fabrication of Pt-Si Schottky diodes.[144]

Micro-Replica Molding

In micro-replica molding, the master mold is replicated in PDMS by casting and curing the PDMS pre-polymer; this negative replica is then oxidized in an oxygen plasma for 1 min and exposed to fluorinated silane for 2 hr to provide a surface with low adhesion to PDMS. PDMS is then cast against this negative replica, cured, and peeled away to reveal a positive replica of the original master.[5]
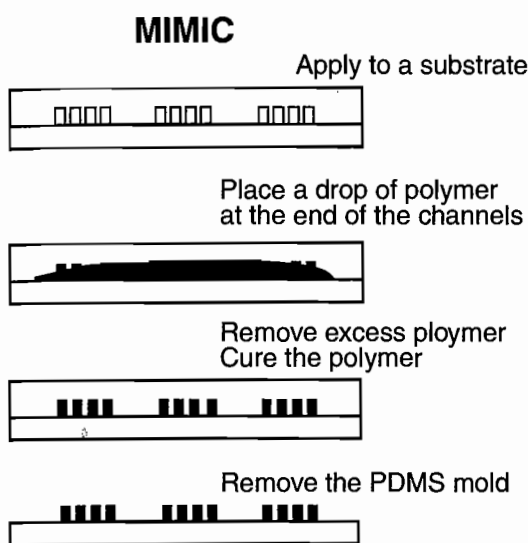
MIMIC

Apply to a substrate

Place a drop of polymer at the end of the channels

Remove excess ploymer
Cure the polymer

Remove the PDMS mold

Figure 1.51 Schematic description of micromolding in capillaries or MIMIC. (Based on Y. Xia and G.M. Whitesides, *Ann. Rev. Mater. Sci.*, 28, 155–184, 1998[138] and Y. Xia and G.M. Whitesides, *Angew. Chem. Int. Ed.*, 37, 550–575, 1998.[139])

Soft Lithography Summary

Soft lithography is capable of generating structures as small as 30 nm and is especially attractive as a potential method for fabricating devices on nonplanar substrates and devices based on conductive polymers. The major advantages of soft lithography are that it is intrinsically very fast. It is possible to go from design to production of replicated structures in less than 24 hr. The method is low in cost and, unlike photolithography, soft lithography is applicable to almost all polymers and thus to many materials (e.g., carbon, glasses) that can be prepared from polymeric precursors.

Some of the disadvantages of soft lithography based on PDMS include (a) shrinkage during curing (~1%), (b) swelling by nonpolar solvents such as toluene and hexane, (c) thermal expansion, (d) level of defects in printing SAMs, (e) softness of the material limits the achievable aspect ratio through sagging, and (f) deformation of the soft elastomeric stamps. These factors limit the accuracy in registration across a large area and may limit the practical utility in nanofabrication and multilayer devices.[138]

Despite these challenges, soft lithography technology is inspiring new approaches to fabricate nano devices. Krauss et al., for example, used nano-imprint lithography (NIL) to fabricate a 400 Gbit/in² storage device containing sub-10 nm minimum features.[140] In this case, the master mold is an electron beam and RIE patterned SiO_2 layer (e.g., 10-nm wide features with 40-nm period and 75-nm height). This mold is then imprinted into a 90-nm thick PMMA film on a Si disc. During the hot embossing, the mold and resist-coated disk are heated to 175°C and pressed together at 4.4 MPa for 10 min. After cooling to room temperature, the mold is separated from the disk, resulting in replication of the nano-CD pattern in the PMMA film. Surface Logix, Inc. (50 Soldiers Field Place, 2nd Floor, Brighton, MA 02135) is the first commercial entity based entirely on exploitation of soft lithography.

3D Lithography Methods

Introduction

Traditional lithography techniques result in projected shapes and, although these may be high in the z direction, they are not truly 3D, as the method does not lend itself to making curved surfaces. In miniaturization science, true 3D lithography is the subject of increased research. The first example of 3D lithography we encountered is the above-reviewed soft lithography. Here, we briefly summarize additional means to achieve truly 3D shapes.

Holographic Lithography

In holographic lithography[145,146] (Figure 1.52), a holographically constructed photomask replaces the standard photomask. In the holographic recording or construction phase of the hologram, one uses the interference of two mutually coherent beams. A well collimated flood laser (object beam) passes through the photomask and is diffracted by the mask features. This signal beam contains the amplitude and phase information of the photomask. When this diffracted object beam passes through the holographic recording layer, it interacts with the reference beam to create the interference pattern. The reference beam converts the amplitude and phase information into intensity information that is stored in the photosensitive holographic medium. The light-sensitive recording layer stores the holographic image data as variations in the refractive index of the
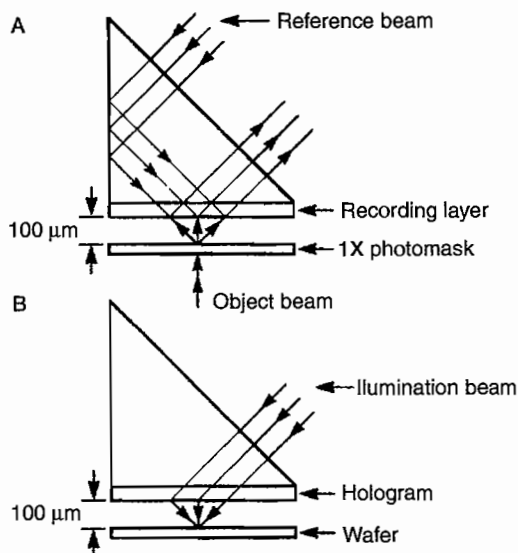

Figure 1.52 Basic arrangement for total internal reflection holographic lithography. (A) The photomask pattern forms a hologram in the polymer recording layer. (B) Using the illumination beam, the high-resolution holographic mask image is reconstructed into a printable masking layer in the resist-coated substrate. (Adapted from J. Brook and R. Dandliker, *Solid State Technol.*, 32, 91–94, 1989[145] and B. Omar et al., *Solid State Technol.*, Sept., 89–94, 1991.[14])

photopolymer. Pattern printing or image reconstruction is accomplished by scanning a collimated laser illumination beam—the phase conjugate of the reference beam—to create the hologram. By interaction with the recorded hologram, the latter beam generates an image of the original photomask at precisely its original position in space. When a photoresist-coated substrate resides in this plane, a copy of the original mask can be printed. The image is only diffraction limited and, because the holographic mask and the wafer can lie very close, a high NA and thus a very good resolution is possible (0.3 μm has been reported). Only the size of the photomask itself restricts the image field size. The holographic image of the original mask exactly overlays the wafer surface with a full field.

Stereolithography/Micro-Photoforming Process

In stereolithography, light exposure solidifies a special liquid resin into a desired 3D shape. Besides producing industrial 3D mock-ups, micromachinists are exploring the same technology to produce micromachines. Some concepts illustrating the technology are shown in Figure 1.53. A liquid resin is kept either in the free surface mode (Figure 1.53A) or in the fixed surface mode (Figure 1.53B). The latter has a resin container with a transparent window plate for exposure. The solidification always happens at the stable window/resin interface. An elevator is pulled up over the thickness of one additional layer above the window for each new exposure (Figure 1.53C). In the case of the surface mode, solidification occurs at the resin/air interface, and more care
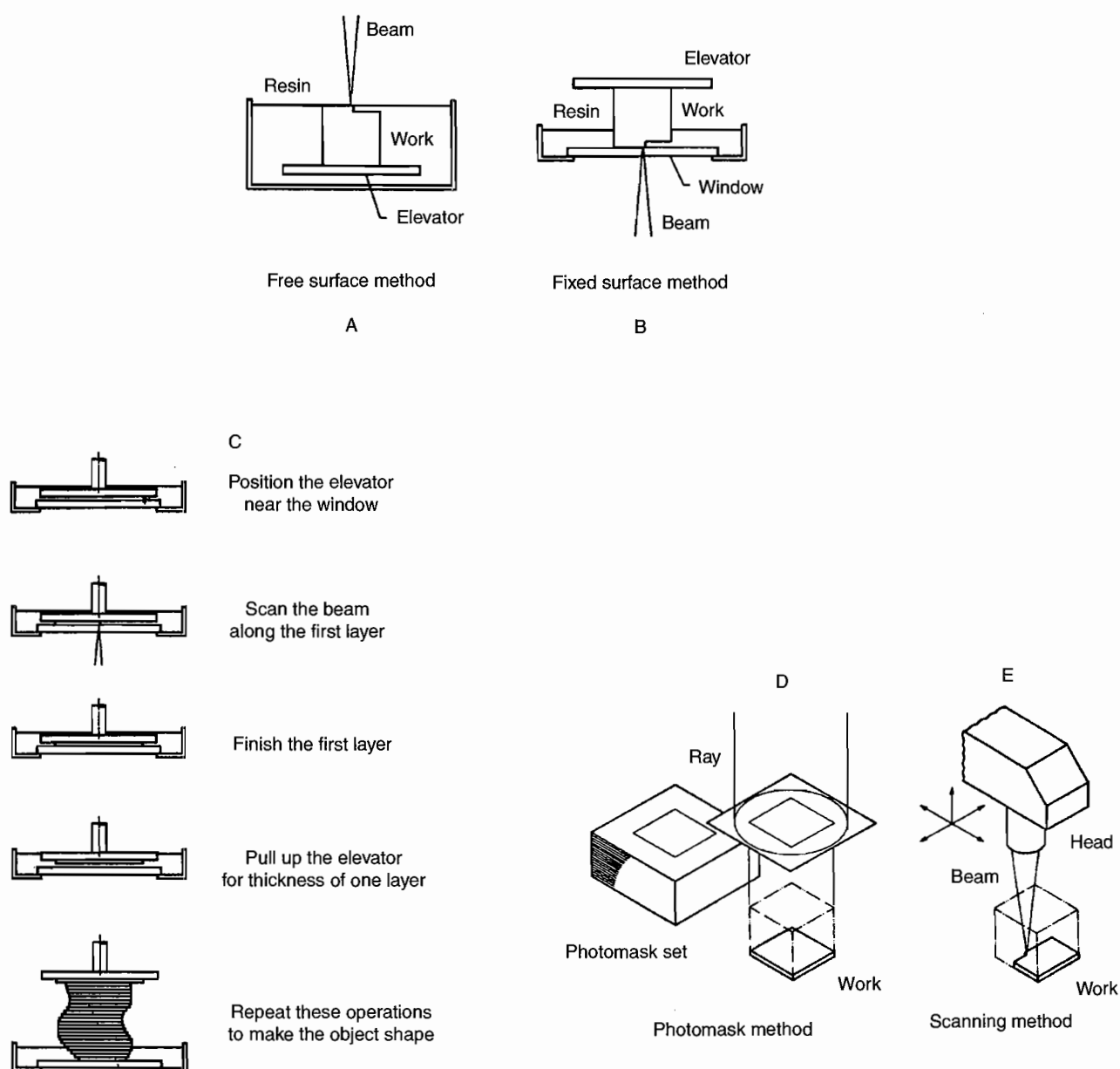

Figure 1.53 Stereolithography or photoforming: (A) free surface method, (B) fixed surface method, (C) forming process with the fixed surface method, (D) exposure with photomask set, and (E) exposure with a scanning beam. (Adapted from K. Ikuta and K. Hirowatari, *Proc. IEEE MEMS '93*, pp. 42–47, 1993[147] and K. Ikuta, et al., IEEE International Workshop on MEMS '94, pp. 1–6, 1994.[148])

needs to be taken to avoid waves or a slant of the liquid surface. A truly 3D structure can be made of a UV polymer by exposing the polymer with a set of two-dimensional cross-sectional shapes (masks) of the final structure. These two-dimensional shapes are a set of photomasks used to subsequently expose the work (photomask method, Figure 1.53D), or the sliced shapes can be written directly from a computerized design of the cross-sectional shapes by a beam in the liquid (Figure 1.53E).

The scanning method has the advantage of point-by-point controllability, thereby avoiding unevenness of solidification, which leads to nonuniform shrinking of the workpiece. The photomask approach, which solidifies a whole layer at the time, is significantly faster than the point-by-point technique. When applying the scanning technique, a laser beam (e.g., a He-Cd laser) is used to solidify one microscopic polymer area at a time to arrive at complicated 3D shapes by stacking thin films of hardened polymer layer upon layer. Process control in this case simply is directed from a CAD system containing the "slice data." The laser beam is focused down to 5 μm spot size, and typical machining time ranges from 30 min to 1 hr. The position accuracy for the laser beam spot is 1 μm in the z-axis, and 0.25 μm in the x and y directions. Takagi et al.[149] obtained an 8-μm resolution with their photoforming setup, and Ikuta et al.[147, 148] report a minimum solidification unit size of $5 \times 5 \times 3$ μm and a maximum size of fabricated structures of $10 \times 10 \times 10$ μm. No physical contact occurs between tools and works, and a very large number of layers can be achieved (e.g., 1000 layers). Very complex shapes, including curved surfaces, can be made with this type of desktop microfabrication method. The objects realized this way include liquid chromatography systems and electrostatic microactuators. Takagi et al.[149, 150] introduced the combination of this type of plastic micromachining with more traditional Si micromachined substrates. In Figure 1.54, a schematic for a photoformed plastic clamp anchored to a Si substrate is represented. The fabricated plastic clamps measure about 2 mm long and 2 mm high and are 250 μm thick.
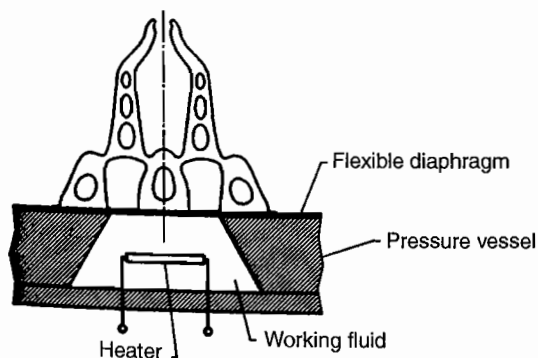

Figure 1.54 Thermally driven micro-clamping tool. The clamping tool is made by photoforming, and the Si substrate by Si bulk micromachining (see text for dimensions). (From T. Takagi and N. Nakajima, MEMS '94, pp. 211–16, Oiso, Japan, 1994.[150] Copyright 1994 IEEE. Reprinted with permission.)

Bertsch et al.,[151] at the Swiss Federal Institute of Technology (EPLF), optimized an interesting variation on the photomask approach, which, given the many layers involved, is impractical. This research group uses a computer-controlled liquid crystal display (LCD) as a dynamic pattern generator. A light beam passes through the LCD, and a beam reducer focuses it on the surface of a polymerizable medium. Selective polymerization takes place in the irradiated areas corresponding to the transparent pixels of the LCD. Between the irridiation steps, a shutter blocks the beam, and a new layer of fresh resin (about 5 μm thick) spreads over the object under construction. Complex objects, with a resolution better than 5 μm in the x, y, and z directions have been made this way, as illustrated in Figure 1.55.[152] The same research group pioneered the combination of planar UV/LIGA lithography (e.g, using SU-8) with micro stereolithography—to add nonvertical structures with curved and conical surfaces in a post-processing step onto planar structures without the need of microassembly.[57] The surfaces of the parts made by microstereolithography are smooth enough to enable conformal electroplating to make a metal mold from the obtained 3D micro object. Polymer replication, of course, remains limited to convex and nonreentrant structures (see Chapter 6).

A comparison of the resolution between conventional stereolithography, small-spot and layer-by-layer microstereolithography was recently presented by Bertsch et al. in *Rapid Prototyping Journal*.[153] To further improve the lithography (3D photoforming, in this case), it is necessary to better understand the shape of a "solidified cell," which depends on both the characteristics of the beam and the resin.

Summarizing, micro-photoforming offers several advantages over more classical photolithography-based micromachining processes:

1. The turnaround from CAD to prototype takes only an hour or less.
2. Photoforming is an additive process, accommodating virtually any shape.
3. It is a fully automatic process.
4. It requires a small capital investment (less than $30,000).
5. There is no need for a clean room.

This direct write lithography technique might well represent an alternative to LIGA in cases where 3D-shape versatility outweighs accuracy. As with the LIGA technique, the plastic shapes made by stereolithography may be used as a cast for electroplating metals or for other materials that can be molded into the polymer structures.

Lithography on Nonplanar Substrates

Jacobsen et al. used a numerically controlled e-beam for nonplanar lithography.[154] They equipped an SEM with high-precision linear and rotary positioning stages inside the vacuum chamber and numerically controlled beam direction and stage position. Cylindrically shaped metals were patterned in this setup. As detailed above, soft lithography also enables lithography on nonplanar substrates (see, for example, Inset 1.29).
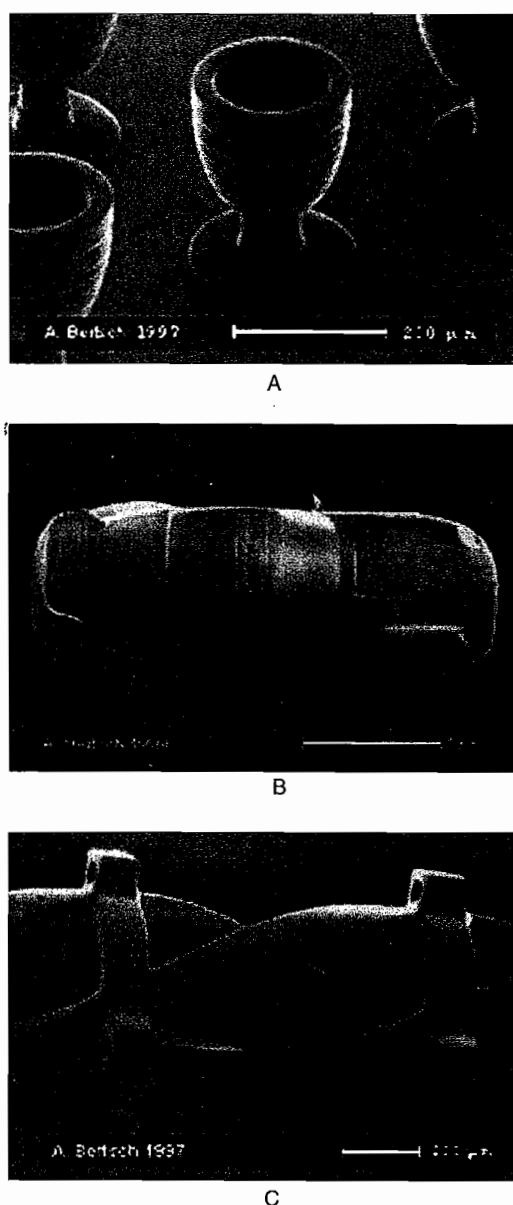
A

B

C

Figure 1.55 Examples of objects made by microstereolithography: SEM photos of (A) micro cups (80 layers × 5 μm, external diameter 200 μm), (B) Porsche (673 layers × 5 μm) (the diameter of a wheel is 400 μm), and (C) micro spring (1000 layers × 5 μm, external diameter 500 μm).

Jackman et al. use PDMS as an elastomeric mask in place of a patterned layer of photoresist.[155] The elastomeric membrane mask is made by spin coating a liquid PDMS prepolymer onto a "master" consisting of features of photoresist on a silicon wafer. The PDMS "stencil" is then peeled off and, because it conforms and seals reversibly against the contours of a surface, it can function both as a dry resist and as a mask for dry lift-off. The method, although limited to features that are >1 μm, is of potential utility in patterning the interior and exterior of curved substrates. The technique is somewhat akin to screen printing in that it uses a stencil and allows patterning of a wide variety of materials, but it achieves higher resolution (5 μm vs. 75 μm) and can be extended to nonplanar substrates. Of particular interest is its capability of patterning hydrogels.

Examples

1.1 Protein Patterning

Microlithographic techniques can be applied successfully to the field of protein patterning. In this first example, we show how clever use of lithography may help solve the problem of non-specific protein adsorption competing for detection sites in immunosensors. Proteins bind with considerable avidity to a wide range of surfaces, and a reference surface in an immunosensor should adsorb all the same proteins except for the protein of interest. The best way to avoid nonspecific binding effects in immunosensors is to make the reference surface as similar to the sensing surface as possible, i.e., a reference substrate, subject to all the same nonspecific protein binding phenomena as the sensor surface itself, except for the antigen-antibody of interest. This way, measuring differentially enables the best possible correction for nonspecific binding. An interesting way of implementing this idea is to create an optical grating pattern, with the antigen-antibody coupling providing the grating structure. This can be realized, for example, based upon the loss of antigenicity upon ultraviolet radiation of thin layers of antigen as observed by Panitz and Giaver.[156] Panitz and Giaver found that the antigenic sites on proteins display sensitivity to ultraviolet radiation in air and that a small dose of ultraviolet radiation can destroy the antigenicity of antigens (Figure 1.56). One presumes that the ultraviolet light breaks up chemical bonds in the adsorbed protein layer and simultaneously produces ozone. The broken bonds are targets for the highly reactive ozone, resulting in the progressive oxidation of the protein layer. The protein can actually be completely removed from the surface by too long an irradiation (say, 10 min). With a short radiation time, the optical density does not change, but antibodies will no longer bind to the antigenetic sites. This scheme was put to use in the fabrication of an elegant biograting for immunosensing as
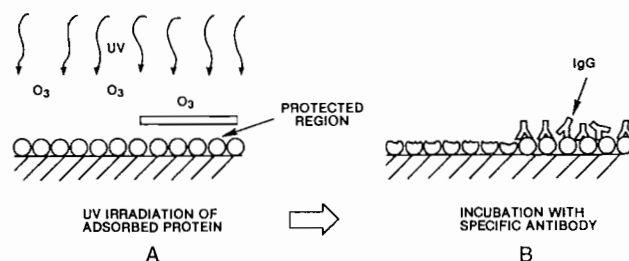

A

B

Figure 1.56 Ultraviolet radiation of antigens leads to bands of "live/dead" protein. (A) A preadsorbed layer of antigen is briefly exposed, in air, to an intense ultraviolet source, resulting in the production of ozone (O_3). The combination of ultraviolet damage and ozone results in a partial oxidation of the adsorbed protein. The area of protein shielded from the irradiation is not oxidized. (B) Following this treatment, the protein-coated substrate is incubated with antiserum specific to the adsorbed antigen. The partially oxidized protein layer is no longer antigenic, and IgG molecules are able to bind only to the previously shielded portion of the antigen layer. (From E. Clementi et al., *Structure and Motion: Membranes, Nucleic Acids, and Proteins*, Adenine Press, Schenectady, NY, 1985.[156] Reprinted with permission.)

shown in Figure 1.57.[157] By using a photomask grating and inactivating alternating bands of antibodies, a biological diffraction grating is created. In this device, the antigen-antibody coupling itself constitutes the grating structure, causing diffraction only if the target antigen is present. A CD-type He-Ne laser beam diffracts from the grating with an intensity related to the antigen concentration. The above grating device comes close to embodying an ideal relative reference, as the radiated protein bands are almost identical to the active surface except for being incapable of reacting with the target complementary protein.[158]

It is easy to imagine arrays with several different enzymes, antibodies, or DNA-probes immobilized precisely onto a transducer surface as a diagnostic panel for clinical diagnosis, high-throughput drug screening, or environmental monitoring.[159] Various approaches to making such panels are explored in Chapter 3.

1.2 Inclined LIGA Walls

Using LIGA, microstructures as high as hundreds or even thousands of micrometers can be achieved with an aspect ratio of 200, lateral dimensions smaller than 0.5 μm, accuracy of 0.1 μm, and wall perpendicularity of 90°. For many applications, inclined side walls would be a very useful addition to the LIGA tool box; for instance, for making lenses or for easier release of molded parts. This second lithography example concerns both of these important applications. When using LIGA for plastic molding, such as through hot embossing, it is important to be able to control the resist wall inclination to facilitate the release of molded devices (see also Chapter 6). To achieve this, one could make an inclined absorber or, following Tabata et al., one can also move the mask during exposure in moving mask deep x-ray lithography (M²DXL).[160] The movement of the mask is realized by a precision x-y stage with a 10-nm step resolution. The principle of the technique is illustrated in Figure 1.58. Here, the x-ray mask with a circular hole of 30 μm is moved in a circle with a 15 μm diameter. The x-ray dose distributes according to the movement of the mask during exposure. As a result, the energy delivered to the PMMA photoresist shows a high constant value within the inner circle of the trajectory of the hole's movement and decreases with increasing trajectory diameter, becoming zero at the outside region. A truncated conical PMMA structure results with an inclination of the wall controlled by the diameters of the mask hole and its circular movement and the total x-ray dose.

For making inclined microfluidic channels whose depth is 40 μm, a mask with a channel width of 50 μm was used, and a mask movement of 10 μm. A good inclination for the sidewalls of microfluidic devices for easy release of molded parts is 80°. Tabata et al. demonstrated that their technique may also be used to make micro-lens arrays. In the latter case the mask hole diameter was 210 μm, and the mask movement was 100 μm.

1.3 PDMS-Based CDs

At Tecan Boston (formerly Gamera Biosciences) (http://www.tecan-us.com/) and in our laboratories at Ohio State University (OSU), drawing programs such as Canvas and Freehand are used to design masks for the lithographic manufacture of fluidic structures on a compact disc (CD) fluidic platform. The mask design is saved as a PostScript file and printed with a high-resolution printer (say, 4000 dpi) on a regular transparency. This fast, inexpensive, and simple prototyping method was pioneered in Whitsides' laboratory at Harvard.[139] An example design for a CD-based fluidic platform by Duffy et al. performs 48 enzymatic assays and is shown in Figure 1.59.[161] Each individual fluidic
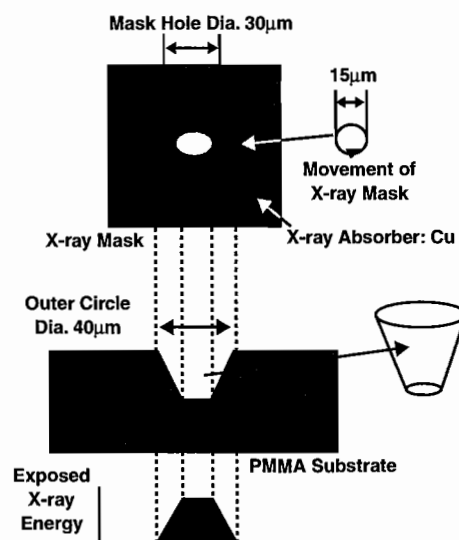

Figure 1.58 Principle of M²DXL. (Based on Tabata et al., in *Micro Total Analysis Systems 2000: Proceedings of the μTAS 2000 Symposium*, Kluwer Academic, 2000.[160] Reprinted with permission.)
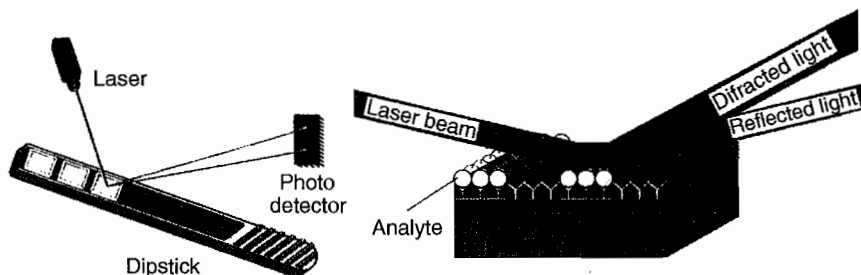

Figure 1.57 Idetek, Inc. immunosensor based on biological grating. (This figure also appears in the color plate section following page 394.) (Courtesy of Mark Platshon, Idetek, Inc., Sunnyvale, California.)[157]

structure allows mixing of an enzyme and inhibitor, followed by mixing with a substrate, and detection of the product (Figure 1.59A). The physics behind the operation of the CD fluidic platform will be explained in Chapter 9; here, we are interested in the fabrication process only. The transparency photomask (Figure 1.59B) is clamped between a 5-in Si wafer coated with SU-8 photoresist (MicroChem, Newton, MA) and a blank glass plate. After exposure and development, a positive relief of photoresist on the Si wafer results. The mold is then passivated by exposing it to a vapor of tridecafluoro-1,1,22-tetrahydrooctyl-1-tricholorsilane (United Chemical Technologies, Bristol, PA, http://www.unitedchem.com/) for 2 h. A 10:1 mixture of PDMS

oligomer and cross linking agent (Sylgard 184, Dow Corning, Midland, MI) prepolymer of PDMS, degassed under vacuum, is cast against this mold, cured for 1 hr at 65 °C, and removed from the mold. This process yields a transparent polymer part that contains channels and chambers that correspond to the positive relief of the photoresist. In a final step, the PDMS CD is sealed to a reservoir layer, a piece of acrylic that contains macroscopic reservoirs (typical volumes of up to 100μL) for liquid input (Figure 1.59C).

The advantage of this rapid prototyping method is that one can go from idea to device in 24 hr with a $20 mask and a $10 mold. However, the mask has a short lifespan compared to a conventional Cr mask and cannot stand high exposure doses.
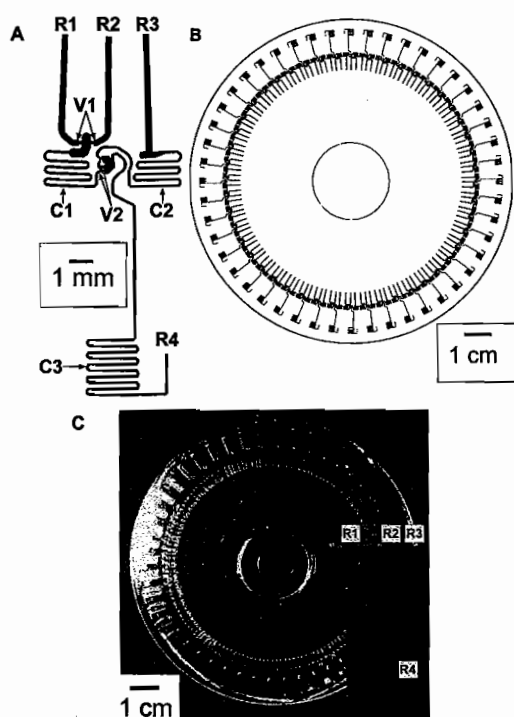

Figure 1.59 (A) Design of a fluidic structure used to perform an assay composed of mixing an enzyme with an inhibitor, followed by mixing with a substrate, and detection. The solutions of enzyme, inhibitor, and substrate were loaded in reservoirs that were connected to channels labeled R1, R2, and R3, respectively. Enzyme and inhibitor combined after being released by capillary burst valves, V1, and mixed in a meandering 100-μm-wide channel, C1. The enzyme-inhibitor mixture was combined with the substrate in a chamber after being released by capillary valves, V2. These solutions mixed in a meandering channel, C3, and emptied into a cuvette (not shown) from a section of channel labeled R4. (B) Design for the photomask used to create PDMS replicas for carrying out 48 enzymatic assays simultaneously. The design was created by rotating the fluidic structure shown in (A) about the center of a 120-mm diameter circle by 7.5° 48 times. (C) Photograph of the disk composed of a PDMS replica that was fabricated by rapid prototyping using a photomask generated from the design in (B) and sealed to a layer of PMMA with reservoir layers machined into it. The inset shows a magnified detail of one of the fluidic structures of the disk. The labels indicate the various reservoirs in the PMMA layer. (From D.C. Duffy et al., *Anal. Chem.*, vol. 71, 1999.[161] Copyright 1999, American Chemical Society. Reprinted with permission.)

Problems

1.1 Analyze each of the expressions for photolithography resolution and explain how to maximize resolution in every case. What are the advantages and disadvantages of using e-beam lithography compared to typical photolithography using UV radiation? What is the most likely next generation lithography?

1.2 An exposure is performed with coherent light using a step-and-repeat projection printing system. The light source has a wavelength of 365 nm (I-line of a mercury arc lamp). The pattern is a grating with a line-to-line spacing of 1 μm.

    (i) Calculate the minimum value of the numerical aperture (NA) that will provide contrast at the image plane (the plane of the resist).

    (ii) What is the maximum value of the numerical aperture, above which there will be no improvement in image quality?

    (iii) Calculate the depth of field of the image for cases *(i)* and *(ii)*.[*]

1.3 Why is it easier to obtain a lift-off profile with a negative resist than with a positive resist?

1.4 How would you make a conical PMMA structure that is 150 μm tall and features a top angle of 45°?

1.5 Design a miniaturized device incorporating both a lift-off process and a self-aligned mask step in its manufacture.

1.6 Which statements are *not* correct?

❑ Short exposure wavelengths can create standing waves in a layer of photoresist. Regions of constructive interference create increased exposure.

❑ Standing waves can impair the structure of the resist, but they can be eliminated by use of multiple wavelength sources or postbaking.

❑ Standing wave effects are most noticeable at the center of the resist.

❑ The primary components of a positive photoresist are
    a. Nonphotosensitive base phenolic resin
    b. Photosensitive dissolution inhibitor
    c. Casting solvent