ULSI Technology

EDITED BY

C. Y. Chang

Chair Professor, College of Electrical Engineering and Computer Science
National Chiao Tung University
Director, National Nano Device Laboratories
Hsinchu, Taiwan, ROC

S. M. Sze

UMC Chair Professor
Department of Electronics Engineering
Director, Microelectronics and Information Systems Research Center
National Chiao Tung University
Hsinchu, Taiwan, ROC

THE McGRAW-HILL COMPANIES, INC.

New York St. Louis San Francisco Auckland Bogotá Caracas Lisbon
London Madrid Mexico City Milan Montreal New Delhi
San Juan Singapore Sydney Tokyo Toronto

McGraw-Hill

A Division of The McGraw-Hill Companies

ULSI TECHNOLOGY

Acknowledgments begin on page 724 and appear on this page by reference.

This book is printed on acid-free paper.

2 3 4 5 6 7 8 9 0 DOC DOC 9 0 9 8 7 6

ISBN 0-07-063062-3

This book was set in Times Roman by Publication Services, Inc.
The editors were Lynn Cox and John M. Morriss;
the production supervisor was Denise L. Puryear.
The cover was designed by Farenga Design Group.
Project supervision was done by Publication Services, Inc.
R. R. Donnelley & Sons Company was printer and binder.

Cover photo: Electron micrograph of contact holes filled with CVD tungsten plugs (see Chapter 8). The diameter is 0.25 micron. *Courtesy of the National Nano Device Laboratories, National Science Council, R.O.C.*

Library of Congress Catalog Card Number: 95-81366

CONTENTS

GROWTH OF THE INDUSTRY

The United States has the largest electronics industry in the world, with a global market share of over 40%. Since 1958, the beginning of the integrated-circuit (IC) era, the factory sales of electronic products have increased by about thirty times [see Fig. 1, curve (a)[1,2]]. Electronics sales, which were $303 billion in 1993, are projected to increase at an average annual rate of 8.5% and reach a half-trillion-dollar level by the year 2000. In the same period, the IC market itself has increased at an even higher rate [see Fig. 1, curve (b)[1,2]].* IC sales in the United States were $28 billion in 1993 and are expected to grow by 13% annually, reaching $65 billion by the year 2000. The main impetuses for such phenomenal market growth are the intrinsic pervasiveness of electronic products and the continued technological breakthroughs in integrated circuits.

The world markets of electronics and semiconductor industries will grow at comparable rates. Figure 2 shows the 1993 world electronics industry with a global sales volume of $679.7 billion. Also shown are the market shares of the six major electronics applications: computer and peripherals equipment at 32.3%, consumer electronics at 21.2%, telecommunication equipment at 16.5%, industrial electronics at 14.3%, defense and space at 11.5%, and transportation at 4.2%. By the year 2000, the world electronics industry is projected to reach $1200 billion, which will surpass the automobile, chemical, and steel industries in sales volume.

Figure 3 shows the 1993 world semiconductor industry, with total sales of $85.6 billion. Only 14% is related to optoelectronics and discrete semiconductor devices. IC sales constitute 86% of the total volume, with the largest segment being memory ICs, followed by microprocessor and microcontroller units, logic ICs, and analog ICs. In 2000, the semiconductor industry is projected to reach $200 billion, with over $170 billion in integrated circuits.

Figure 4 shows the market shares of the three major IC groups: MOSFET, bipolar transistor, and ICs made from III–V compound semiconductors.[3] At the beginning of the IC era, the IC market was broadly based on bipolar transistors. However, because of the advantages in device miniaturization, low power consumption, and high yield, sales volume of MOS-based ICs has increased steadily and in 1993 amounted to 75% of the total IC market. By the year 2000, MOS ICs will capture the largest market share (88%) of all ICs sold. This book, therefore, emphasizes MOS-related ULSI technology.

*There were only two years in which the growths were negative: in 1974, due to the Middle East oil embargo, and in 1985, due to overproduction of personal computers.

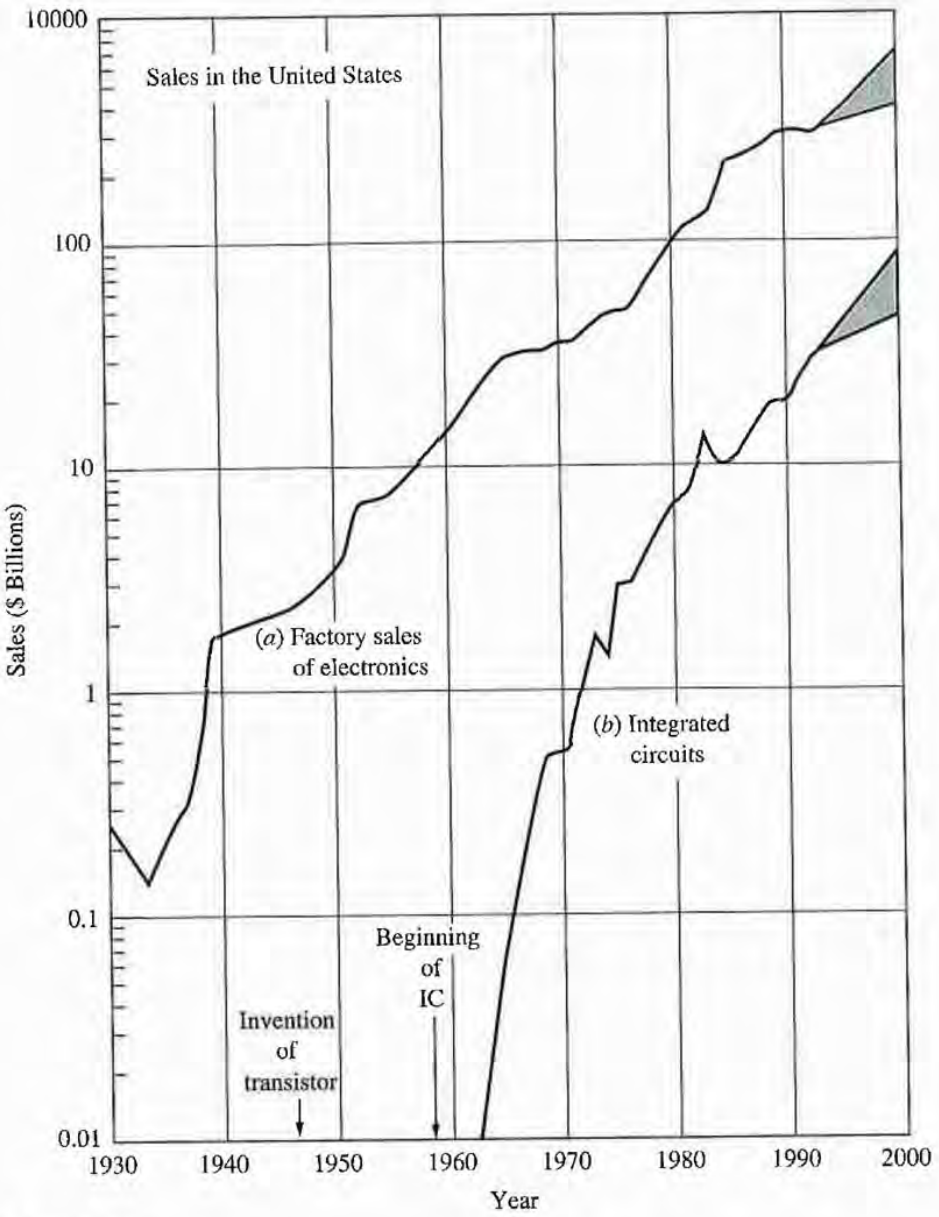


FIGURE 1
(*a*) Factory sales of electronics in the United States for the 64 years between 1930 and 1993 and projected to 2000. (*b*) Integrated circuit market in the United States for 32 years between 1962 and 1993 and projected to 2000. *(After Refs. 1 and 2.)*

DEVICE MINIATURIZATION

Figure 5, curve (*a*), shows the rapid growth in the number of components per MOS memory chip.[4,5] Note that the MOS IC complexity has advanced from small-scale integration (SSI), to medium-scale integration (MSI), to large-scale integration (LSI), to very-large-scale integration (VLSI), and finally to ultralarge-scale integration (ULSI), which has 10^7 or more components per chip. We note that since 1975 the growth has been maintained at a rate of about 40% annually; in other words, the

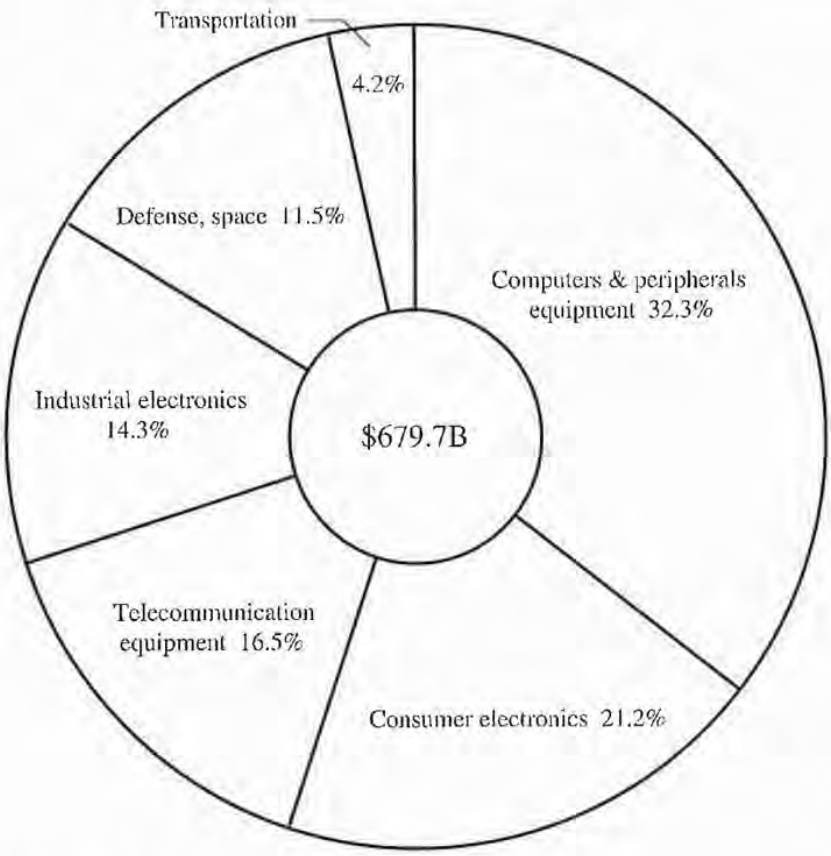


FIGURE 2
1993 world electronics industry. *(After Dataquest, 1994.)*

number of components has doubled every two years. At this rate, over 100 million components per chip will be available before the year 2000; in the early 21st century we will move into the gigabit range, with IC chips having more than one billion components.[6,7] Also shown in Fig. 5 is the growth of the number of components for bipolar, MESFET, and MODFET ICs. They are about two orders of magnitude lower in complexity compared with MOS-based ICs.

The most important factor in achieving the ULSI complexity is the continued reduction of the minimum device-feature length [see Fig. 6, curve (*a*)]. Since 1960, the annual rate of reduction has been 13%, which corresponds to a reduction by a factor of two every six years. At this rate, the minimum feature length will shrink from its present length of 0.5 μm to 0.2 μm in the year 2000. The junction depth of the source and drain junctions, and the gate oxide thickness are also being reduced at a similar rate as shown in curves (*b*) and (*c*) of Fig. 6, respectively.

The reduction of the device feature length and related dimensions has resulted in reduced overall device size and unit price per function. Figure 7 shows the relative price and size reductions.[8] In the past fifty years, prices have gone down by 100 million times, and the size has been reduced by a factor of one billion. By 2000 the price per bit is expected to be less than 0.1 millicent for a 64-megabit memory chip. Similar price reductions are expected for logic ICs. Additional benefits from device

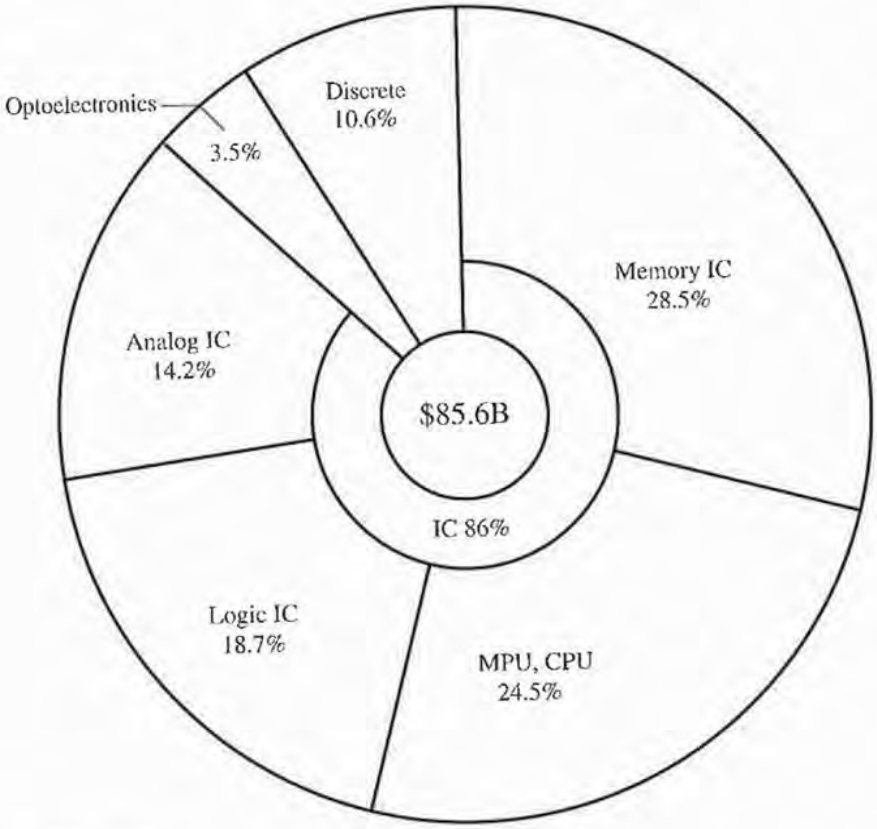


FIGURE 3
1993 world semiconductor industry. *(After Dataquest,1994.)*

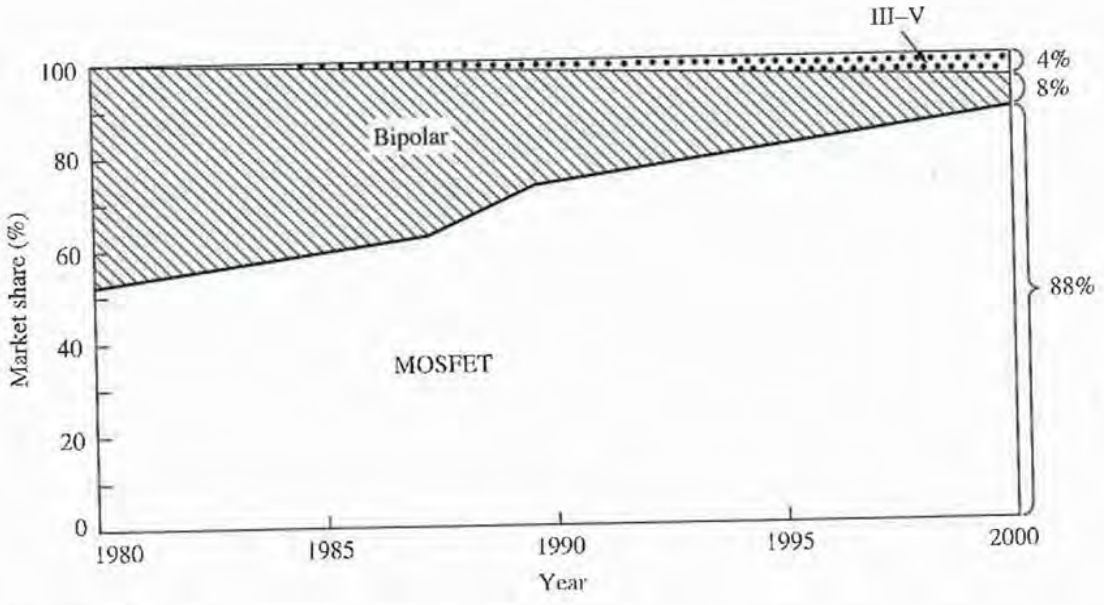


FIGURE 4
World IC market (1980–2000). *(After Zdebel, Ref.3.)*

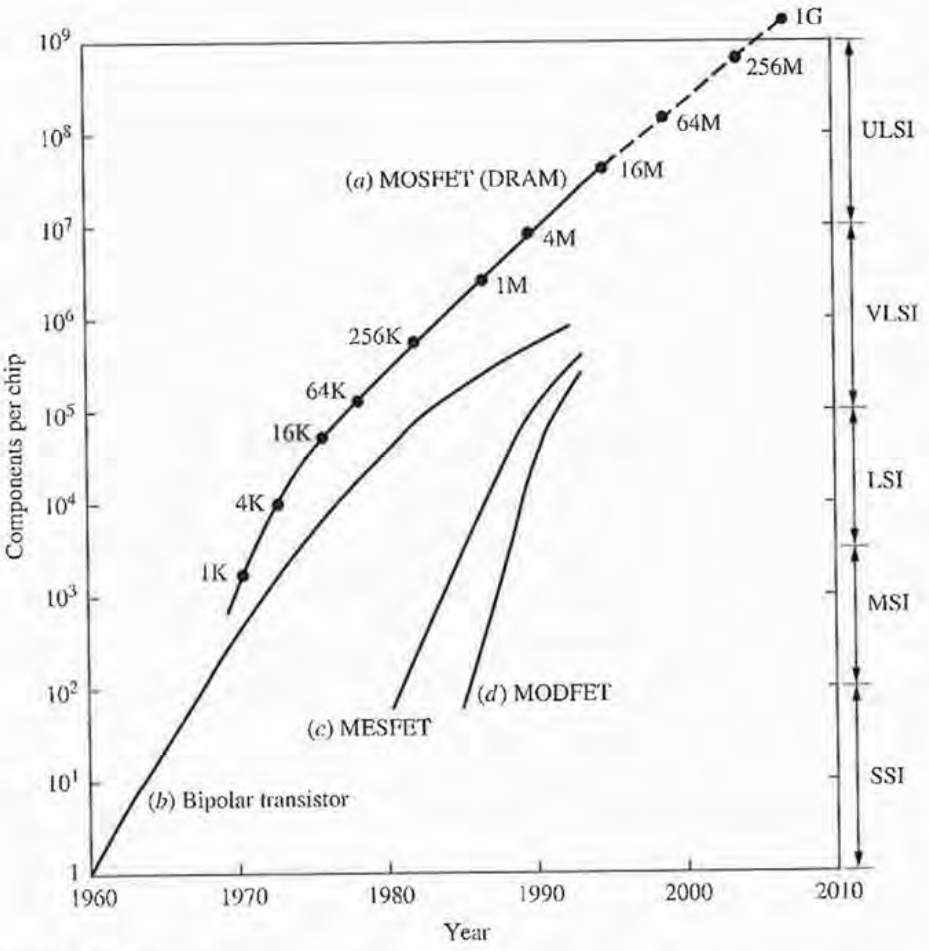


FIGURE 5
(*a*) Exponential growth of the number of components per MOS IC chip. *(After Moore, Ref. 4, and Myers, Ref. 5.)* (*b*), (*c*), and (*d*) Components per chip versus year for bipolar, MESFET, and MODFET ICs, respectively.

miniaturization include improvement of device speed (which varies inversely with the device feature length) and reduction of power consumption (which varies approximately with the square of the feature length). Higher speeds lead to expanded IC functional throughput rates, so that future ICs can perform data processing, numerical computation, and signal conditioning at 100 and higher gigabit-per-second rates.[9] Reduced power consumption results in lowering the energy required for each switching operation. Since 1960 the required energy, called the *power-delay product*, has decreased by six orders of magnitude.[10]

ORGANIZATION OF THE BOOK

Figure 8 shows how the 12 chapters of this book are organized. Chapter 1 considers cleanroom technology. The continued miniaturization in ULSI devices implies more

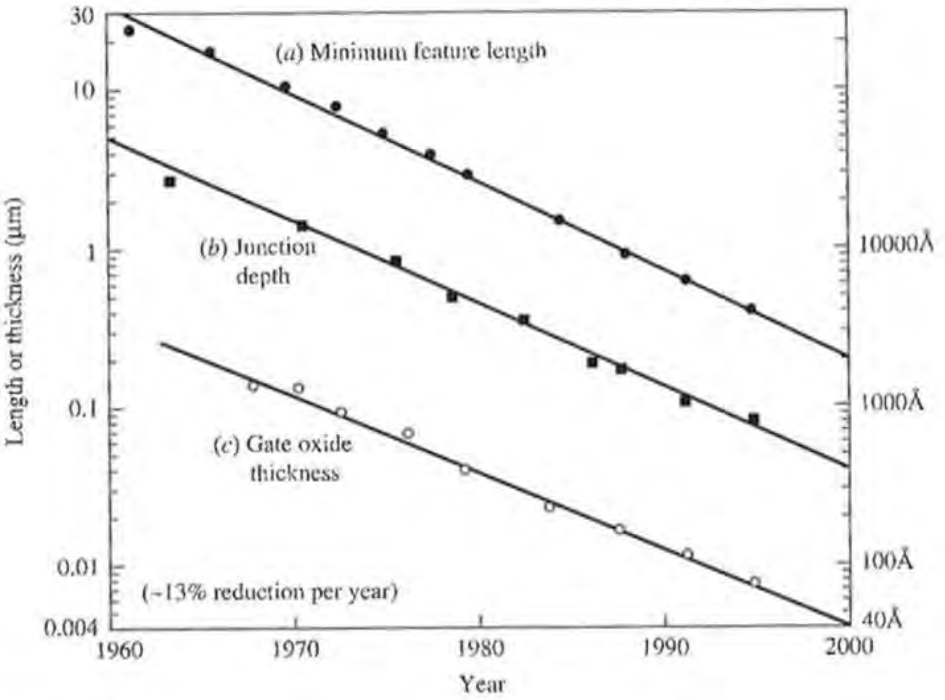


FIGURE 6
Exponential decrease of (*a*) minimum feature length, (*b*) junction depth, and (*c*) gate oxide thickness of MOSFET.

stringent requirements with respect to contamination control. Without an ultraclean processing environment, ULSI circuits simply cannot be realized.[11]

ULSI technology is synonymous with *silicon* ULSI technology. The unique combination of silicon's adequate bandgap, stable oxide, and abundance in nature ensures that in the foreseeable future no other semiconductor will seriously challenge its preeminent position in ULSI applications. Some important properties of silicon are listed in Appendix A.

Once the silicon wafers are in the cleanroom, we enter into the wafer-processing sequence, described in Chapters 2 through 8 and depicted in the wafer-shaped central circle of Fig. 8. Each of these chapters considers a specific process step. Of course, many processing steps are repeated many times in IC fabrication; for example, lithography and etching steps may be repeated 10 to 20 times. In ULSI technology the wafer-cleaning technology is as important as the cleanroom technology. Without a contamination-free wafer surface, the ICs will suffer from low yield and poor reliability. Because of limitations on the total length of the book, many classic topics, such as crystal growth, oxidation, diffusion, and ion implantation, are only briefly mentioned. The reader may consult textbooks on VLSI technology for details.[12]

The individual processing steps described in Chapters 2 through 8 are combined in Chapter 9 to form devices and integrated circuits. Chapter 9 considers the fundamental building process modules and four important IC families: CMOS (complementary MOSFET), bipolar ICs, BiCMOS (a combination of bipolar and CMOS),

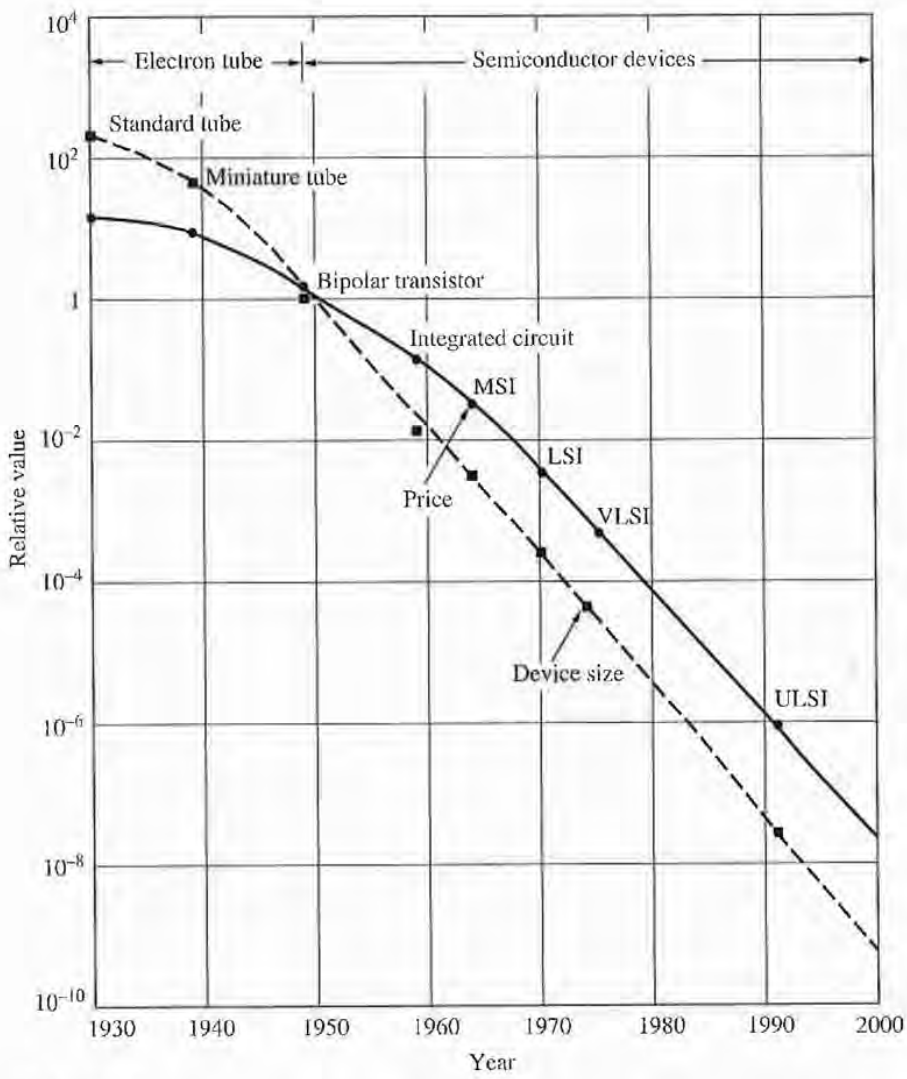


FIGURE 7
Price and size reduction of active electronic compents. *(After Shoda, Ref. 8.)*

and MOS memory ICs. After the completely processed wafers are tested, those chips that pass the tests are ready to be packaged. Chapter 10 describes the assembly and packaging of ULSI chips. Chapter 11 considers the manufacturing technology, that is, the strategy and logistics to implement various technologies to produce ULSI chips that meet customers' specifications in a timely fashion and to generate adequate return on investment for the IC manufacturer. Chapter 12 describes a multitude of reliability issues related to ULSI processes. As device dimensions move to the sub-half-micron and sub-quarter-micron regime, ULSI processing becomes more automated, resulting in tighter control of all processing parameters. At every step of production, from wafer cleaning to device packaging, numerous requirements are being imposed to improve the device performance and reliability.

To keep the notation simple in this book, we sometimes found it necessary to use a symbol more than once, with different meanings. However, within each chapter a

FIGURE 8
Organization of this book.

symbol has only one meaning and is defined the first time it appears. Many symbols do have the same or similar meanings consistently throughout this book; they are summarized in Appendix B.*

ULSI technology is presently moving at a rapid pace. The number of ULSI publications has doubled every year since 1990, the beginning of the ULSI era. Many topics, such as lithography, rapid thermal processing, and metallization, are still under intensive study. Their ultimate capabilities are not fully understood. The material presented in this book is intended to serve as a foundation. The references listed at the end of each chapter can supply more information.

REFERENCES

1. *1994 Electronic Market Data Book,* Electronic Industries Association, Washington, D.C., 1994.
2. *1994 Annual Report of Semiconductor Industry,* Industrial Technology Research Institute, Hsinchu, Taiwan, ROC, 1994.
3. P. J. Zdebel, "Current Status of High Performance Silicon Bipolar Technology," *14th Annual IEEE GaAs IC Symp. Tech. Digest,* 15 (1992).
4. G. Moore, "VLSI, What Does the Future Hold," *Electron Aust.*,**42,** 14 (1980).
5. W. Myers, "The Drive to the Year 2000," *IEEE Micro,***11,** 10 (1991).
6. P. K. Chatterjee and G. B. Larrabee, "Gigabit Age Microelectronics and Their Manufacture," *IEEE Trans. VLSI Syst.***1,** 7 (1993).
7. K. Mori, H. Yamada, and S. Takizawa, "System on Chip Age," *Proceedings of the International Symposium on VLSI Technology, Systems, and Applications,* k15 (1993).
8. K. Shoda, "Home Electronics in the 1990s," *Proceedings of the International Symposium on VLSI Technology, Systems, and Applications,* 1(1991).
9. H. Komiya, M. Yoshimoto, and H. Ishikura, "Future Technological and Economic Prospects for VLSI," *IEICE Trans. Electron.* **E76-C,** 1555 (1993).
10. R. W. Keyes, "Limitations of Small Devices and Large Systems," in N. G. Einspruch, Ed., *VLSI Electronics,* Academic, New York, 1981, Vol. 1, p. 186.
11. T. Ohmi, "ULSI Reliability through Ultraclean Processing," *Proc. IEEE,* **81,** 716 (1993).
12. For example, S. M. Sze, Ed., *VLSI Technology, 2nd Ed.,* McGraw-Hill, New York, 1988.

*Also included are the International System of Units (Appendix C) and Physical Constants (Appendix D).

CHAPTER 6

Lithography

K. Nakamura

6.1 INTRODUCTION

Lithography is a kind of art made by impressing, in turn, several flat embossed slabs, each covered with greasy ink of a particular color, onto a piece of paper. The various colors or levels must be accurately aligned with respect to one another within some registration tolerance. Many "originals" can be made from the same slabs as long as the quality remains adequately high.

Several methods can be used to make ULSI circuit patterns on wafers, as shown in Fig. 1*a*. The most common process is to make the master photomask using an electron beam exposure system and replicating its image by optical printers, as shown in Fig. 1*b*. The exposing radiation is transmitted through the "clear" part of a mask. The opaque part of the circuit pattern blocks some of the radiation. The resist, which is sensitive to the radiation and has resistance to the etching, is coated on the wafer surface. The mask is aligned within the required tolerance on the wafer; then radiation is applied through the mask, the resist image is developed, and the layer underneath the resist is etched.

Therefore, lithography for integrated circuit manufacturing is analogous to the lithography of the art world. The slabs correspond to masks for the various circuit levels. The press corresponds to the exposure system, which not only exposes each level but also aligns it to a completed level. The ink may be compared to either the exposing radiation or the radiation-sensitive resist; the paper can represent the wafer into which the pattern will be etched, using the resist as a stencil.

Lithography is the key technology in semiconductor manufacturing, because it is used repeatedly in a process sequence that depends on the device design. It determines the device dimensions, which affect not only the device's quality but also its product amount and manufacturing cost.

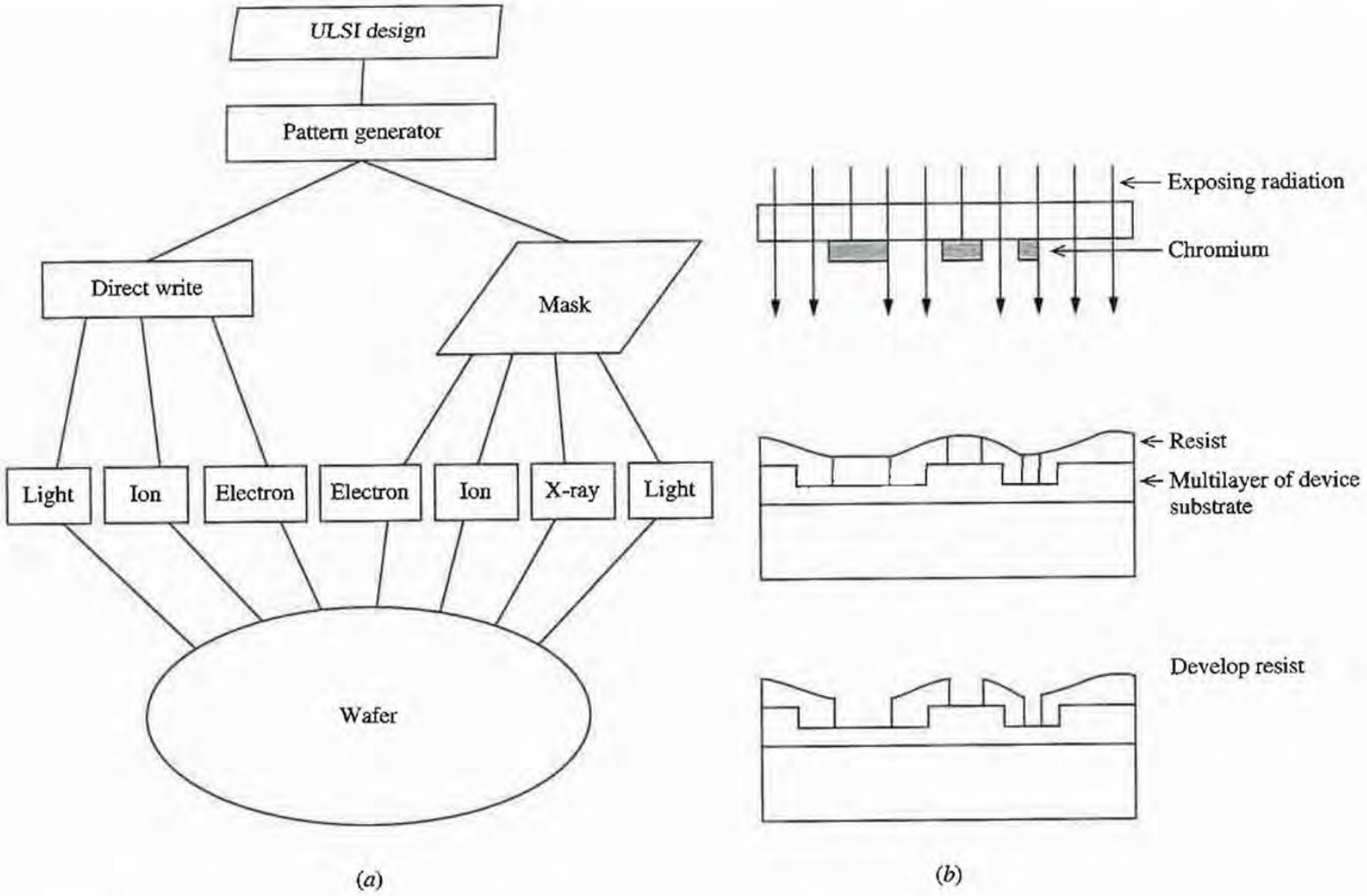


FIGURE 1
Device lithography generalization. (*a*) Lithographic process for ULSI. (*b*) Optical replication process.

6.2 OPTICAL LITHOGRAPHY

Optical lithography comprises the formation of images with visible or ultraviolet radiation in a photoresist using proximity or projection printing. Two methods are available to make masks for optical lithography: electron beam exposure and laser beam scanning. These are described in the next section. In the 1970s, the major technology was the combination of a negative resist and proximity printing. At present the most common method is the combination of a positive resist and a stepper. However, proximity printing is still used because of its convenience and low cost.

Table 1 lists examples of commercially available optical printers[1] designed to manufacture ULSI circuits. These machines are classified into three groups: proximity, reflective projection, and refractive projection. The key parameters such as numerical aperture (NA), depth of focus (DOF), resolution (usable linewidth), overlay accuracy, and throughput are also listed.

The most advanced ULSI has a minimum feature size of 0.3 μm to 0.4 μm, which has been reported by several organizations in 64-Mbit dynamic random access memories (DRAMs). It is believed that the linewidth limit of optical lithography lies near 0.2 μm using phase-shifting masks combined with a high numerical aperture projector and a short-wavelength light source. The performances of the machines listed in Table 1 are very close to the resolution limit of optical lithography.

6.2.1 Contact and Proximity Printing

Contact and proximity printings are relatively simple because they do not have any means of image formation between masks and wafers. A typical contact or proximity mask aligner consists of a light source, a condenser, a filter, a mirror, a shutter, the wafer stage, and the alignment microscope. In contact printing, a photomask is pressed against the resist-covered wafer, with pressures typically in the range of 0.05 atm to 0.3 atm, and is exposed by light with a wavelength near 400 nm. Very high resolution of less than 0.5 μm linewidth is possible, but because of spatial nonuniformity of the contact, resolution may vary considerably across the wafer. To provide better contact over the whole wafer, a thin (0.2 mm) flexible mask has been used; 0.2-μm space patterns have been formed by using 3-μm-thick PMMA (poly-methyl methacrylate) resist and 200 to 260 nm radiation.[2] Quartz or Al_2O_3 mask substrates must be used to pass these shorter wavelengths, since the usual borosilicate glass strongly absorbs wavelengths less than 300 nm.

Contact printing produces defects in both the mask and the wafer while the two are in contact and as they are separated from each other, so that the mask, whether thick or thin, may have to be replaced after a short period of use. Nevertheless, contact printing is still widely used. It is the most convenient way to get high resolution. Strictly speaking, *contact* printing is never actually practiced, because of the difficulty of uniform contact.[3] When a typical mask and wafer are brought into hard contact, the nonuniform gap can be as large as 15 μm because of unevenness of the surfaces. Besides that, the thickness of the resist between the mask and the wafer cannot be neglected for features of less than 1 μm.

TABLE 1
Commercially available optical printers[1]

Category	Model	Maker	NA†	DOF† (μm)	Field size (mm)	Usable linewidth (μm)	Overlay accuracy (μm)	Throughput/6″-hr
Contact/proximity	PLA600	Canon			160	1	0.5	120
	3HRP	JBA				0.25	0.5	45/10″
	Q-6000	Quintel				0.25		60
	MA150M	Karl Suss				0.5	0.5	60
Catadioptric/4×	Micrascan IIi	SVG	0.5	1.2	22 × 32.5	0.5	0.07	51/8″
/1×	Micralign 700	SVG	0.167	6	150 × 135	1.0–1.2	0.25	100
/1×	2000gh	Ultratech	0.24–0.40	2–3	18 × 18–39 × 11	0.8–1.0		78
I-line stepper	PAS5500/100	ASM	0.6	1	22 × 27.6	0.4	0.07	88
	XLS7500/29	GCA	0.55	1.1	21 × 21	0.5	0.09	70
	NSR2005i9C	Nikon			21.2 × 22.8/18 × 25.2	0.45	0.1	47/8″
	FPA2500i3	Canon	0.6	1	20 × 20	0.4	0.12	62
KrF stepper	PAS5500/90	ASM	0.5	1	21 × 27.7	0.35	0.07	80
	XLS7800/31	GCA	0.53	1.2	22 × 22	0.35	0.07	72
	ISR2005EXBA	Nikon			21.2 × 21.2/16 × 25.2	0.4	0.1	34/8″
	FPA3000EX1	Canon	0.45	1	20 × 20	0.3	0.1	66

†NA - numerical aperture
DOF - depth of focus

Masks used for proximity printing have the advantage of longer life because there is no contact between the mask and the wafer. Typical separations between mask and wafer are in the range of 20 to 50 μm. Resolution is not so good as in contact printing or projection printing. Commercial machines can select the mode automatically and switch to either contact or proximity printing.

Figure 2*a* shows proximity printing schematically with a slit of width W, illuminated by a monochromatic source, separated from a parallel image plane (wafer) by a gap g. We assume that g and W are larger than the wavelength λ of the imaging light and that $\lambda \ll g < W^2/\lambda$—the region of Fresnel diffraction. In this case, the diffraction that forms the image of the slit is a function only of the particular combination of λ, W, and g, which we call the parameter Q, where[4]

$$Q = W\sqrt{2/g\lambda} \tag{6.1}$$

This parameter can be considered as a normalized slit width. The lower limit of the nearfield diffraction region is at $Q = \sqrt{2}$ from Eq. (6.1), or $g = (W^2/\lambda)(2/Q^2) < W^2/\lambda$. Thus, the resolution, W in Eq. (6.1), becomes better at smaller gaps and shorter wavelengths. The resist image heights, which for practical purposes are converted to resist thickness and mask-wafer gap, have been measured as a function of linewidth, wavelength, and intensity tolerance.[3] The optimum image heights of 0.25-μm and 0.5-μm patterns are 0.9 μm and 3.2 μm, respectively, in the case of a deep UV (wavelength less than about 200 nm) light source with ±5% intensity tolerance.

Another parameter is the divergence of the illuminating beam as a result of the size of the light source. One effect is to smooth out the undulations in the image intensity profile, and the other is to produce a greater linewidth variation as the mask-to-wafer distance varies. The apparent source size of the illumination system must

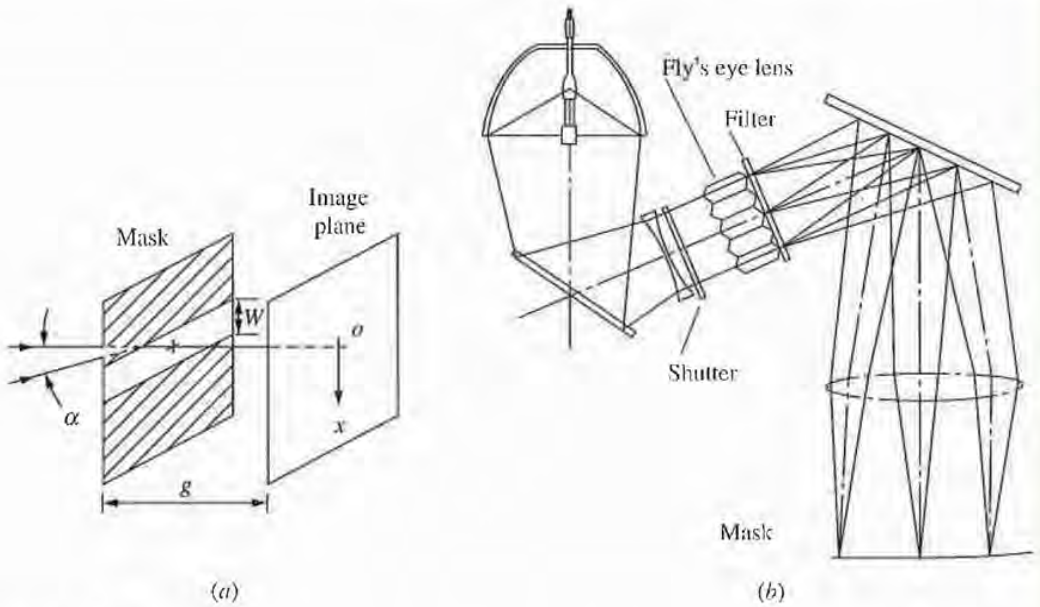


FIGURE 2
Schematic illustration of a proximity printer and illumination system. (*a*) Proximity printing schematic. (*b*) Illumination system of the Canon PLA600.

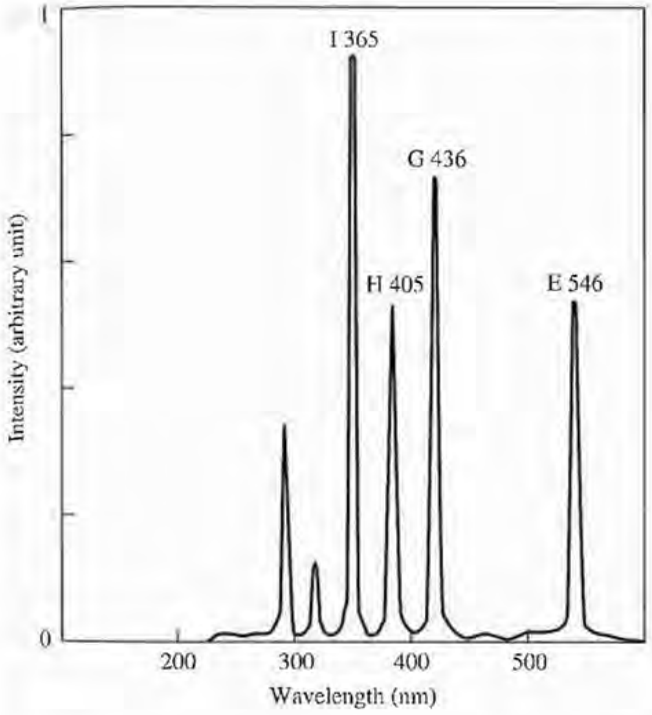


FIGURE 3
Spectrum of a high-pressure mercury lamp.

be large enough to give a value of α (Fig. 2) that allows the smallest features to be printed, typically, a few degrees. The mercury arc lamp used as the source[5] is too small to yield the required α. The illumination is telecentric or normally incident at the mask to prevent runout (magnification) error similar to that in x-ray lithography. The optical system must also minimize nonuniformity of the intensity across the field. The illumination system illustrated in Fig. 2*b* attains ±3% uniformity across a 6-inch wafer, as well as a large enough α and normal incidence. With a Hg arc source, the strong lines at 436 nm, 405 nm, and 365 nm provide the exposure flux shown in Fig. 3. The same printer is available with a Xe-Hg source for enhanced output in the 200–300 nm spectral region.

Commercially available machines are listed in the first group in Table 1.

6.2.2 Projection Printing

Projection printing offers higher resolution than proximity printing. It has a larger separation between the mask and the wafer because of its image formation system. According to Rayleigh's criterion, the resolution W of the optical systems and the depth of focus (DOF) are given in the following equations, where the numerical aperture NA = $n \sin \alpha$, n is the refractive index = unity for air, 2α is the solid angle of the ray reaching the imaging point from the objective lens, and λ is the wavelength:[6]

$$W = 0.6\lambda/\text{NA} \tag{6.2}$$

$$\text{DOF} = \pm\lambda/2(\text{NA})^2 \tag{6.3}$$

Modern projection printers employ diffraction-limited optics, which means that the design and fabrication of the optical elements are not a source of optical problems. Accordingly, the image characteristics are dominated by diffraction effects associated with the finite apertures in the condenser and with projection optics rather than by aberrations. When monochromatic light from a very small point is imaged by a diffraction-limited lens, the image consists of diffraction rings of light surrounding a central bright spot called the *Airy disc*.[6] The diameter of the pattern is $1.2\lambda/\text{NA}$.

It is very useful to apply a modulation transfer function (MTF) to characterize the resolution capability of projection printers and proximity printers as well.[7] We consider the mask pattern as consisting of a periodic grid (line and space) with equal linewidths b. Light intensities in the center of dark lines and of bright lines are assigned as $I_{\min}$ and $I_{\max}$, respectively. Modulation is defined as

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{6.4}$$

The MTF is the ratio of the modulation in the image plane (wafer) to that in the object plane (mask). It is a function of the spatial frequency ($\nu = 1/2b$) of the mask and the NA. The shape of the MTF, related to the spatial frequency curve, depends on the coherence of the illumination system.[7] The degree of coherence is defined by $\sigma = (\text{NA})_c/(\text{NA})_o$, where $(\text{NA})_c$ is the numerical aperture of the condenser and $(\text{NA})_o$ is the numerical aperture of the objective as shown in Fig. 4. In the coherent system, Fig. 4*a*, an object at point A is illuminated by only a narrow angle. Hence, all the light diffracted by A is coherent in amplitude at the image plane. In Fig. 4*b*, A is illuminated by rays from all portions of the extended incoherent source. Each ray is diffracted by the object A and forms an image in the image plane.

In Fig. 5, the solid lines show a plane wave front [$\sigma = (\text{NA})_c/(\text{NA})_o \to 0$] incident normal to the mask, which contains a grating pattern with b width lines and spaces (spatial frequency $\nu = 1/2b$).[8] The undiffracted component of light emerging from the mask contains no information about ν. This information is contained only in the diffracted light. The direction of the first diffraction peak is given by the grating formula $2b \sin\theta = \lambda$, so that $\nu = \sin\theta/\lambda$. If the light diffracted to direction θ is to reach the image plane, $\theta \approx \alpha$ where $\text{NA} = \sin\alpha$. Therefore, the highest grating frequency that can be imaged by an optical system with coherent illumination becomes the result of Eq. (6.5).

For incoherent illumination (shown by dotted lines in Fig. 5), incident at a general angle i, the direction of the first maximum is given by $2b(\sin i + \sin\theta) = \lambda$. For both the undiffracted beam and the first diffraction peak to reach the image plane, i and θ both $\leq \alpha$. Therefore, $2b \geq \lambda/2\sin\alpha$ and $\nu_{\max} = 2b = 2\text{NA}/\lambda$. Consequently, $\nu_{\max}$ (incoherent) $= 2\nu_{\max}$ (coherent). In modern printers the illumination is intermediate between the coherent and the incoherent limit.

The curve for MTF ($H(\nu)$) vs. spatial frequency[9] is shown in Fig. 6 and is described as[5]

$$\nu_{\max} = \text{NA}/\lambda \qquad \text{(coherent)} \tag{6.5}$$

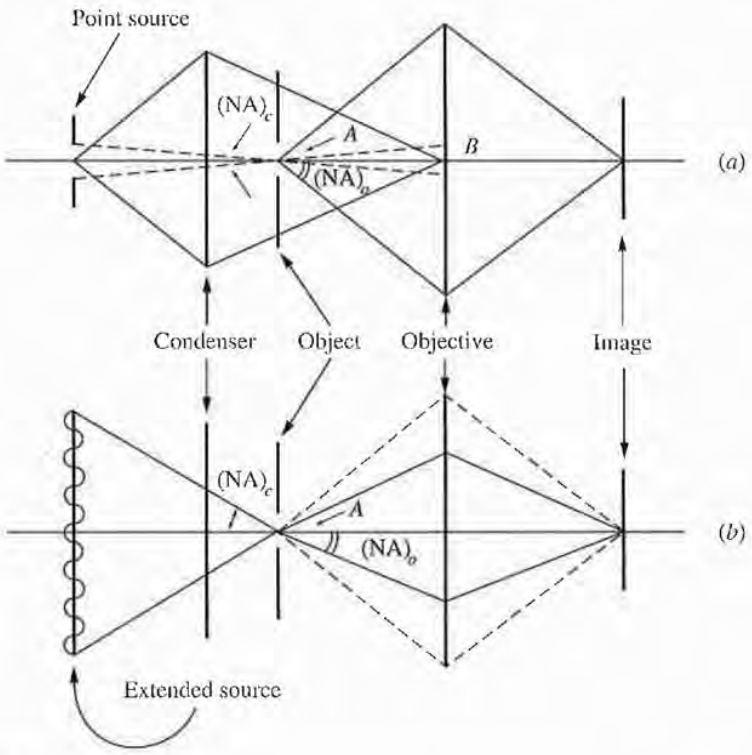


FIGURE 4
Coherent and incoherent illumination of projection printers. (*After King and Goldrick, Ref. 7.*) (*a*) Coherent system. (*b*) Incoherent system.

$$H(\nu) = \frac{2}{\pi}[\cos^{-1}(\nu/\nu_{\max}) - \nu/\nu_{\max}\sqrt{1-(\nu/\nu_{\max})^2}] \qquad \text{(incoherent)} \quad (6.6)$$

where $\nu_{\max} = 2\text{NA}/\lambda$

Figure 6 shows the modulation of the image intensity versus the spatial frequency for different values of coherence σ with a 0.3 NA lens. Note that the MTF of a fully coherent optical system has a spatial frequency cutoff that is half that of an incoherent system as explained in Eq. (6.5) and Eq. (6.6). Partially coherent illumination has a higher MTF in the region of ν larger than $\frac{1}{2}\nu_0$. The useful range, MTF > 0.6, is extended to higher spatial frequencies; edge gradients in the image become steeper, and the image is somewhat less sensitive to focusing.

The focus error is a simple but important aberration. The error is a displacement of best focus away from its intended position. In Fig. 6, the dashed curve shows the effect of a displacement of the image plane or wafer from the focal plane by one Rayleigh unit $w = \lambda/2(\text{NA})^2$, corresponding to a phase error of $\pi/2$ at the edge of the pupil.

It was once believed that a high NA is always better. However, in the submicrometer region there is an optimum NA if a resolution requirement and the imaging wavelength are given.[10] If the NA is too low, the resolution cannot be achieved, but if the NA is too high, the DOF, which is inversely proportional to $(\text{NA})^2$, becomes

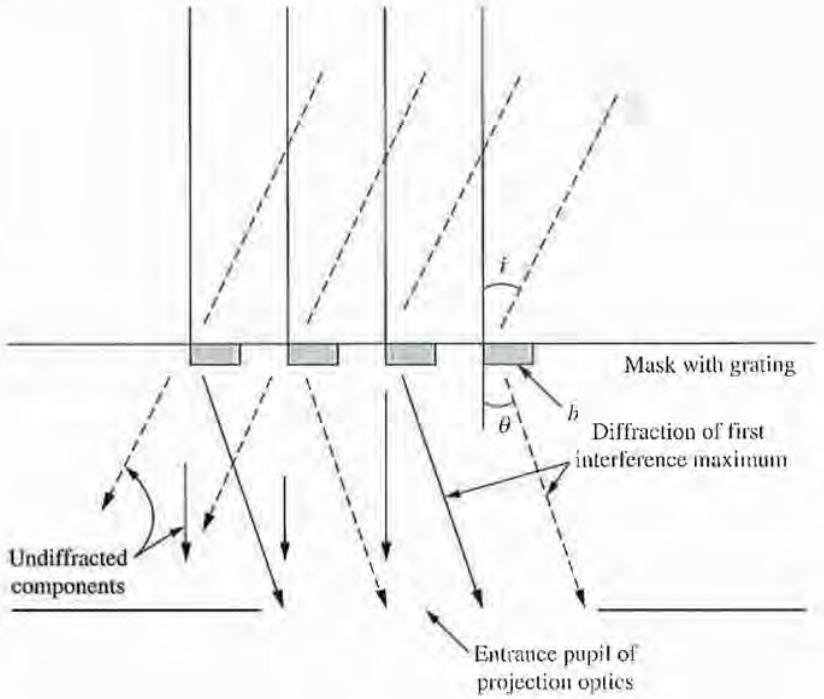


FIGURE 5
Diffraction of coherent and incoherent light by a grating pattern. (*After Cuthbert, Ref. 8.*)

unacceptable. There is an optimum NA at which the DOF is maximum. The normalization of the resolution W is given by

$$k_1 = W\frac{\text{NA}}{\lambda} \tag{6.7}$$

where k_1 is usually known as Rayleigh's coefficient for resolution but is now redefined as the normalized resolution. The normalized DOF is given by

$$k_2 = \Delta Z\frac{(\text{NA})^2}{\lambda} \tag{6.8}$$

where ΔZ is the physical axial displacement from the focal plane, and k_2, which is usually known as Rayleigh's coefficient for DOF, is now redefined as the normalized DOF.

With normalized resolution and defocus, the lithographic imaging behavior of a given feature and illumination can be universally plotted in the exposure-defocus (E-D) space in the form of constant linewidth contours. The exposure dosage required to keep the image linewidth constant at each defocal plane is evaluated to form contours. This set of contours, which is the E-D tree of the particular feature and illumination condition, is universally applicable for all W-λ-NA combinations leading to an identical k_1. The DOF is defined as the total amount of defocus allowed without violating a given linewidth tolerance. Typically, a $\pm 10\%$ tolerance is

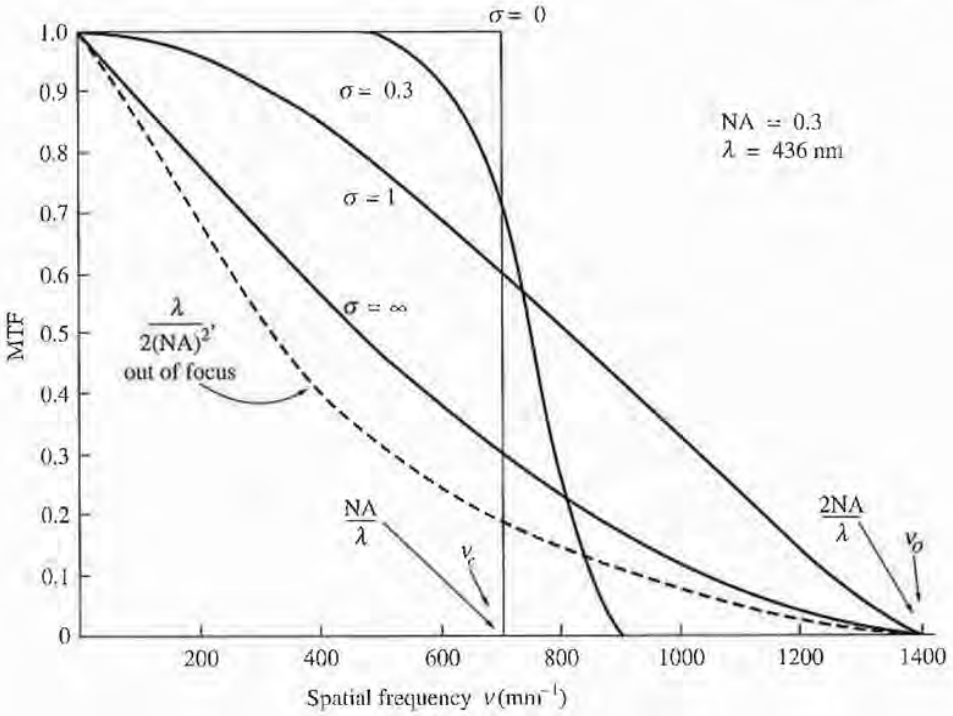


FIGURE 6
Comparison of MTFs for different optical conditions. (*After Lacombat et al., Ref. 9.*)

adopted. The DOF is the height of the largest window whose width is the exposure budget (process margin) that accommodates the tolerances that can affect the actual exposure dosage. Figure 7*a* shows the DOF determined by 10% and 30% exposure budgets, respectively. Figure 7*b* shows two E-D windows. One is bound by the common zone of three kinds of feature (line and spaces, isolated line openings, isolated spaces), and the other by five kinds of feature (including holes and islands).

The DOF equation is produced from Eq. (6.7) and Eq. (6.8) as

$$\text{DOF} = \frac{k_2}{k_1^2}\frac{W^2}{\lambda} \tag{6.9}$$

where λ/W = NA coefficient
W^2/λ = DOF coefficient
k_2/k_1^2 = DOF, normalized to W and λ

Figure 8*a* and 8*b* shows k_2 and k_2/k_1^2, respectively, as a function of k_1, and Fig. 8*c* shows W^2/λ and λ/W as a function of W. The normalized DOF and k_1 can be converted to physical DOF and NA by multiplying by the DOF coefficient and the NA coefficient, respectively, which are plotted in Fig. 8*c*.

The optimum NA for individual feature shapes is much lower than expected, as a result of Fig. 8*c*. The results lead to the problems that occur in reducing the usable k_1 factor for manufacturing. The optimum k_1 for the single-layer resist process (10% linewidth budget estimated) ranges from 0.57 to 0.87 depending on the feature shape.

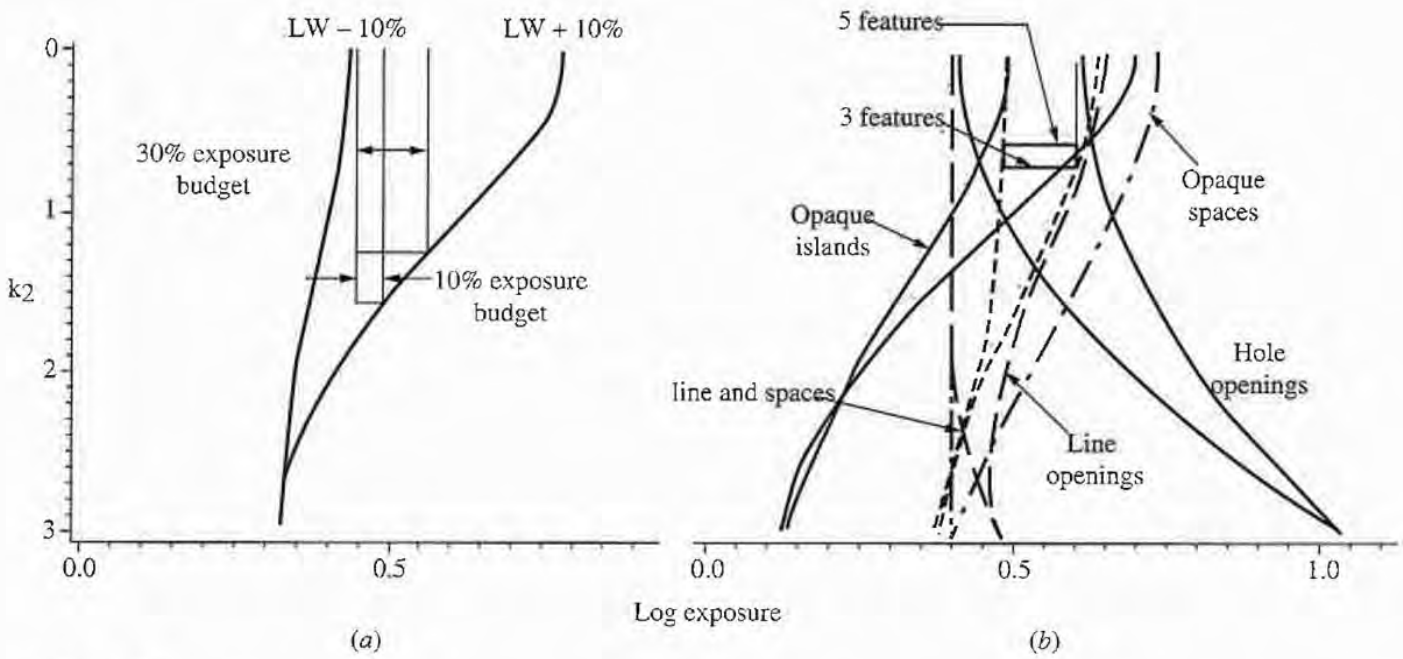
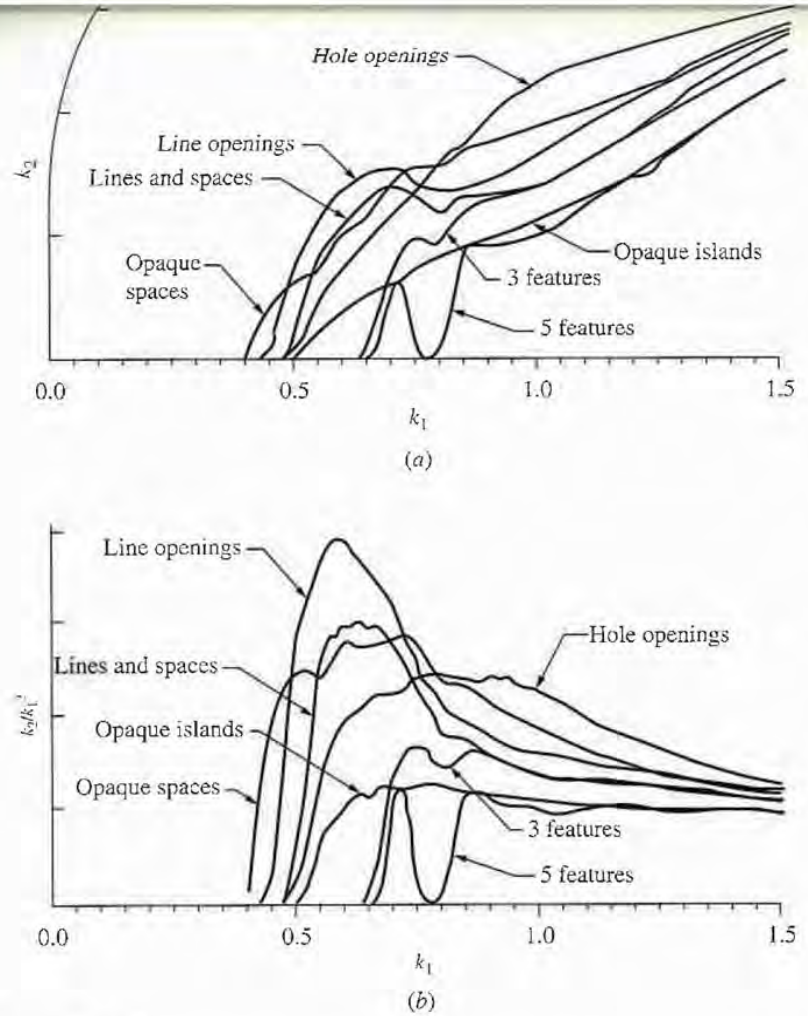


FIGURE 7
E-D windows. (*After Lin, Ref. 10.*) (*a*) The 10% and 30% E-D window bounded by ±10% E-D branches of line and space pairs at $k_1 = 0.8$, $\sigma = 0.4$. (*b*) The 3-feature and 5-feature E-D windows at $k_1 = 0.7$, $\sigma = 0.52$.



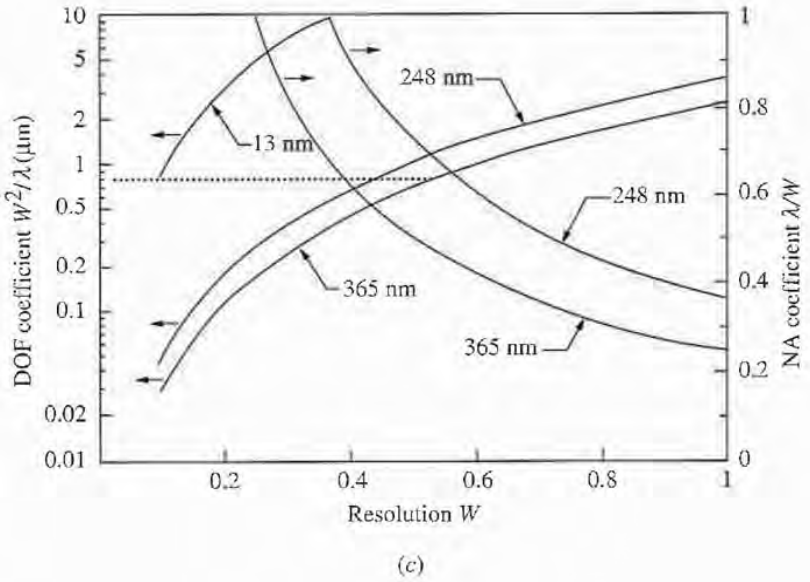


FIGURE 8
NA optimization parameters for optical projection printers. (*After Lin, Ref. 10.*) (*a*) k_2 as a function of k_1 at 10% budget. (*b*) k_2/k_1^2 as a function of k_1 at 10% budget. (*c*) The DOF coefficient and NA coefficient as a function of W. The dotted line shows the similar difficulty of achieving 0.1μm by X-ray and 0.5μm by I-line.

For a multilayer resist process (30% budget), k_1 ranges from 0.42 to 0.7, although it is generally defined as 0.6 as shown in Eq. (6.2).

Reflection and catadioptric projection

A significant advantage of the reflective or the catadioptric system[11] (combining reflecting and refracting components) is the larger spectral bandwidth. Reflection projection systems generally have higher throughput with less standing-wave interference than refraction systems because of their polychromatic characteristics. Catadioptric systems usually require more than just one optical axis and thus can lead to great difficulty in aligning the optical elements. Several systems of this type[12] have been developed as shown in the second category of Table 1 and in Fig. 9.

The Micralign®, made by Perkin Elmer, consists of only three optical components, as shown in Fig. 9*a*. It has a bandwidth of 400 nm. The aberrations vary as a function of the distance from the center and can be zero for a narrow zone about the center. Only a small zone of the spherical mirror is used, limited by the slit, providing nearly diffraction-limited imaging. The mask and the wafer are swept in unison through an arc to form an image of the whole mask. Figure 9*b* is the catadioptric system, also by Perkin Elmer, which is still widely used as a 1× full-scan system. The throughput is very high, as shown in Table 1, but the disadvantages of these systems are the low resolution and the difficulty of alignment during scanning; hence, the overlay accuracy is rather poor, as listed in Table 1.

Figure 9*c* is a catadioptric 1× stepper made by Ultratech (Table 1). Only the wafer has a step-and-repeat movement, and a die-by-die alignment is done. This system also has a very simple optical configuration, featuring a variable NA that depends on object field size—a low NA corresponds to a large field size. The practical resolution is 0.7 μm.

Figure 9*d* shows a 4× catadioptric reduction projector (Table 1) developed by Perkin Elmer that has 0.5-μm resolution. This machine also has a *ring field* similar to Fig. 9*a* and 9*b* with field-by-field alignment and a focus adjustment. The image field is larger than in I-line steppers.

Refraction projection

In refraction projection, the image of the mask is projected on the wafer through a high-resolution lens. A refraction projection system consists of a light source, a condenser, a heat-removing filter, beam-orienting mirrors, a shutter, and the mask and wafer stages. The refraction projection system is very similar to the proximity printers except for the imaging lens between the mask and the wafer. The optical stepper has become the most common lithography tool for ULSI production. The resolution is 0.4 μm using the I-line of the Hg lamp with the Kohler method, which images the exposure source through a condenser lens in an entrance pupil of the projection lens. However, the DOF is reduced to 1 μm because of the stepper's large NA and short wavelength. The features of the stepper are high resolution, and overlay accuracy as shown in Table 1.

The last group in Table 1 is an excimer laser stepper[13] whose wavelength is 248 nm. It can obtain the highest resolution of the equipment listed in Table 1 based on the excimer's shortest wavelength. The system configuration is almost the same as an I-line stepper, but the light source and the lens system are different. In the deep

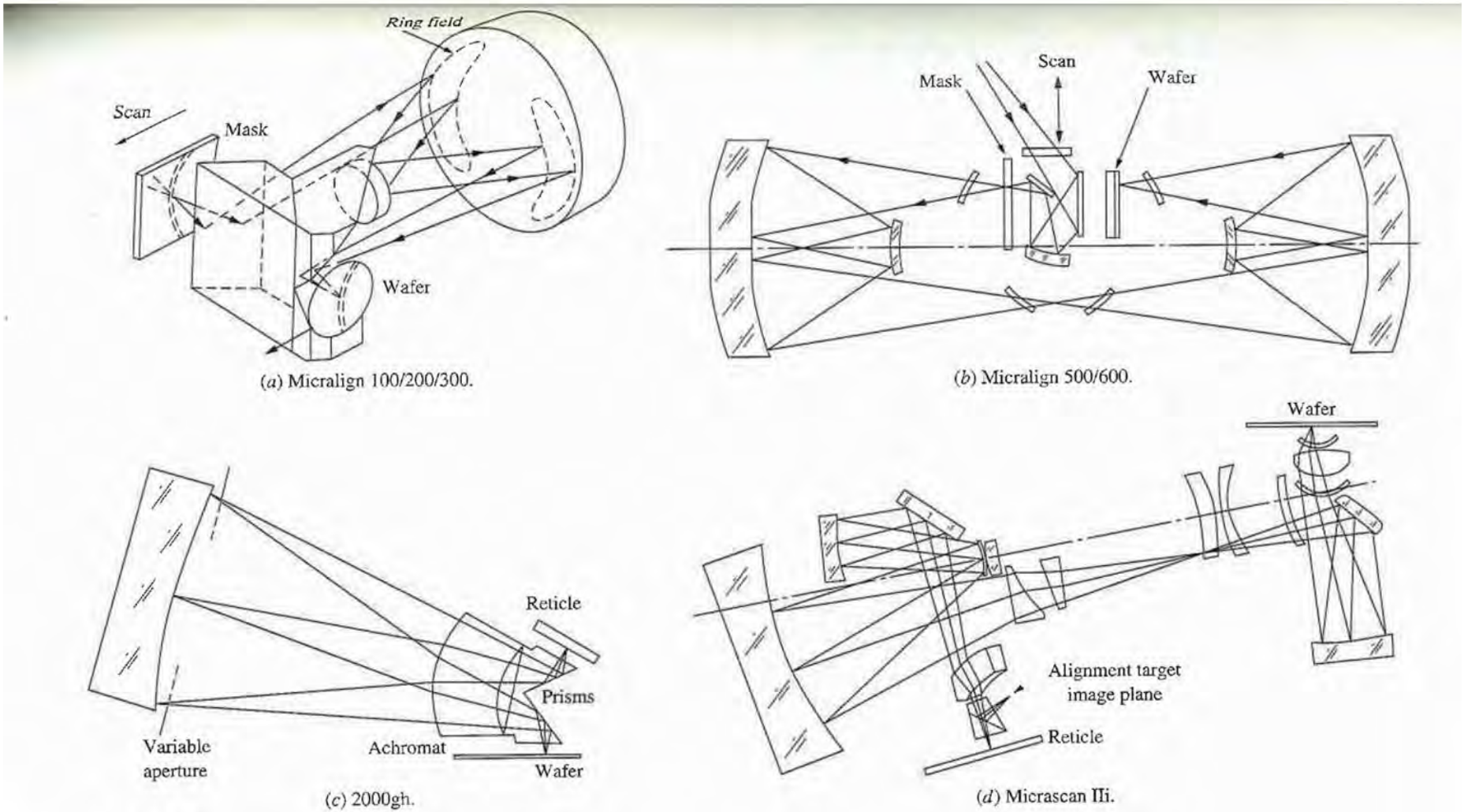


FIGURE 9
Examples of reflection projection printers. (*After Lin, Ref. 12.*)

UV region, the materials of the optical components are limited to fused silica and several crystalline fluorides because of UV transmittance characteristics. The lenses are made entirely of fused silica and therefore are chromatic. The source bandwidth must be narrow to achieve sufficient resolution. The focus and the magnification differences due to wavelength difference are 0.15 μm/pm and 0.3 ppm/pm, respectively. The current performance of the KrF laser source, such as a 2.2-pm bandwidth and a ±0.25-pm wavelength stability, is considered adequate to meet the requirement for practical use. A 0.35-μm resolution within the 0.03-μm critical dimension spread is reported.[13] The use of the excimer laser is another possible way to fabricate ULSI patterns without resists by using the damaging effect of a short-wavelength light.

6.2.3 Enhancement

Many efforts have been made to solve optical lithography problems arising from wavelength or DOF. Among these, the phase-shifting mask and the off-axis illumination technique seem to be useful. The basic concept of the phase-shifting mask[14] is shown in Fig. 10. At the transmission mask, the electric field $\mathscr{E}$ has the same phase at every aperture (clear area) in Fig. 10*a*. Diffraction and the limited resolution of the optical system spread the electric field $\mathscr{E}$ at the wafer, as shown by the dotted

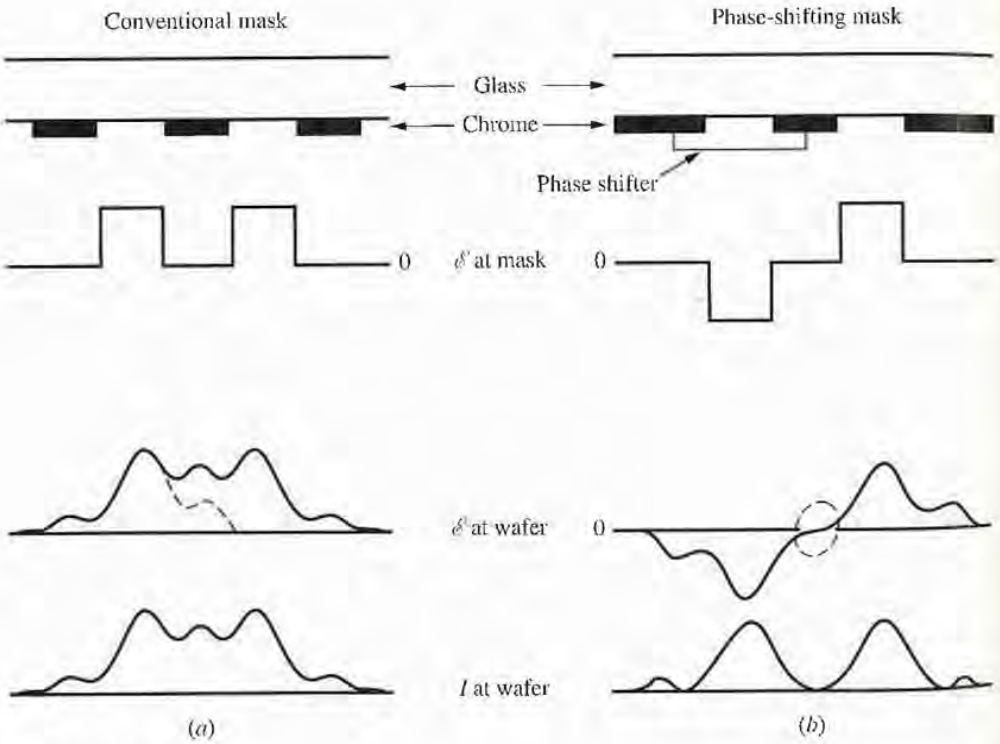


FIGURE 10
The principle of phase-shift technology. (*After Levenson et al., Ref. 14.* © 1992 IEEE.) (*a*) Conventional technology. (*b*) Phase-shifting technology.

line. Interference between waves diffracted by the adjacent apertures enhances the field between them. The intensity I is proportional to the square of the electric field.

The phase-shifting layer that covers adjacent apertures reverses the sign of the electric field as shown in Fig. 10*b*. The intensity at the mask is unchanged. The electric field of these images at the wafer, shown by the dotted line, can be canceled. Consequently, images that are projected close to one another can be separated completely. A 180° phase change occurs when a transparent layer of thickness $d = \lambda/2(n-1)$, where n is the refraction index and λ is the wavelength, covers one aperture as shown in Fig. 10*b*.

The Fourier analysis[14] of phase-shifting and conventional transmission masks is explained in Fig. 11. The conventional resolution power of a projection system for transmission objects is the critical frequency $\nu_c = \mathrm{NA}/\lambda$, which is related to the wavelength λ and numerical aperture NA and assumes coherent illumination. With incoherent illumination some spatial modulation is transmitted through such a system up to a frequency of $2\nu_c$; but for coherent illumination the modulation transmission function is unity up to ν_c and zero for $\nu > \nu_c$, as shown in Fig. 6. For the purpose of calculation it is convenient to define for the mask a transmission function that has sinusoidal intensity, as illustrated at the top of Fig. 11. In each case,

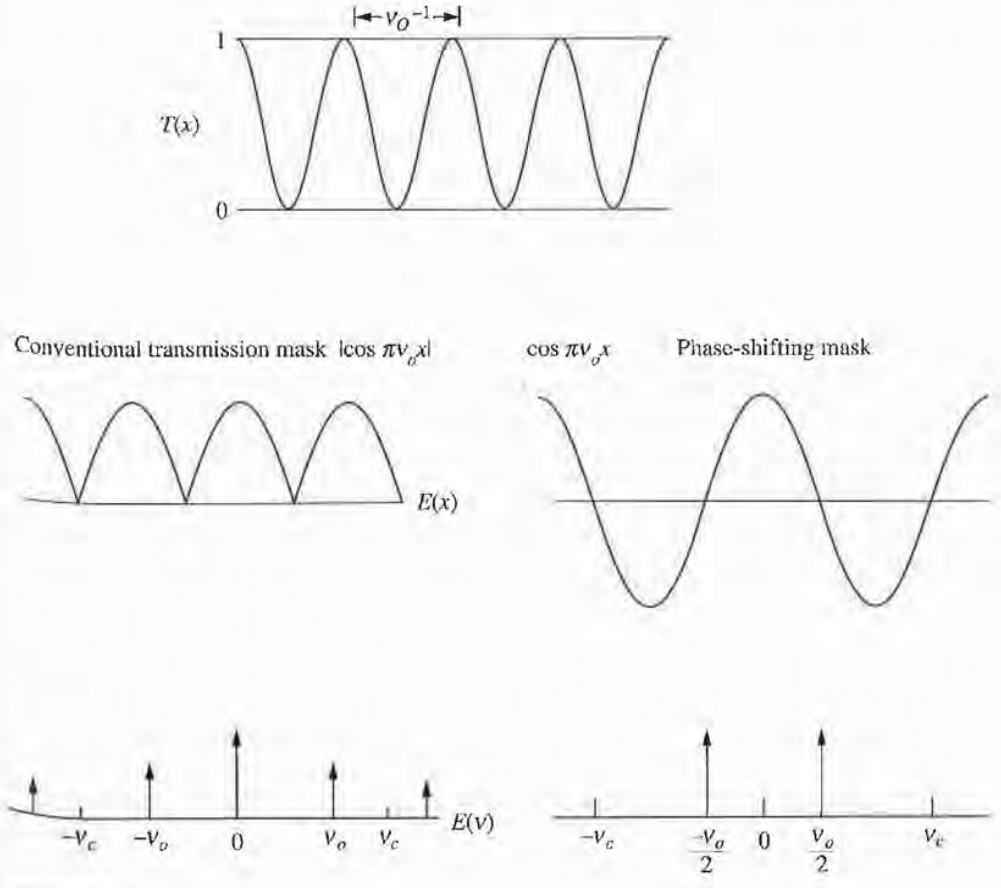


FIGURE 11
Fourier analysis of the optics of transmission and phase-shifting masks. (*After Levenson et al., Ref. 14.* © 1992 IEEE.)

the intensity transmitted by the mask is assumed to be

$$T(x) = \tfrac{1}{2}(1 + \cos 2\pi\nu_0 x) \tag{6.10}$$

The electric field at the mask plane can have two possible forms when the illumination is normally incident and coherent. A transmission mask would produce an electric field profile $\mathscr{E}$ that is proportional to $|\cos \pi\nu_0 x|$, as in the left side of Fig. 11, whereas a phase-shifting mask would produce the electric field on the right, or $\mathscr{E} \propto \cos \pi\nu_0 x$. Both profiles can be Fourier-analyzed as shown at the bottom of Fig. 11. The transmission mask yields a Fourier spectrum of the form

$$\mathscr{E}_T(\nu) = \frac{4}{\pi}\left\{\frac{1}{2}\delta(\nu) + \frac{1}{6}[\delta(\nu - \nu_0) + \delta(\nu + \nu_0)] - \frac{1}{30}[\delta(\nu - 2\nu_0) + \delta(\nu + 2\nu_0)] + \cdots\right\} \tag{6.11}$$

where $\delta(\nu)$ is the Dirac delta function and higher harmonics have been ignored. The phase-shifting mask projects only two Fourier components:

$$\mathscr{E}_\phi(\nu) = \tfrac{1}{2}[\delta(\nu - \nu_0/2) + \delta(\nu + \nu_0/2)] \tag{6.12}$$

The same information, in the sense of an optical intensity, is transmitted in each case, but the phase-shifting mask requires less spatial bandwidth than the transmission mask.

The optical system through which these Fourier components must propagate can be modeled as having a response function equal to unity for $\nu \leq \nu_c$ and zero for $\nu > \nu_c$, where ν_c is the cutoff frequency. Thus, the Fourier components of the electric field at the image plane are (except for a scale factor that can be set to unity) identical to those at the mask plane for $\nu < \nu_c$ and zero otherwise. From this treatment, the critical frequency for the transmission of modulation produced by a phase-shifting mask is $\nu_0 = 2\nu_c$, twice that of a transmission mask.

It is expected that an I-line stepper can resolve features down to 0.2 μm using a phase-shifting mask without reducing the DOF. The problem with this technology is the restriction on the pattern layout. It is very difficult to apply this technology to random patterns, although it is very useful in a periodic layout. Several ideas have been proposed[15] to solve the problem, as shown in Fig. 12. The halftone method, shown in Fig. 12*a*, has advantages such as not requiring extra data and ease of processing, but the resolution is not so good as the alternative method of Fig. 12*c*.

Phase-shifting masks are made by using an electron-beam (e-beam) direct-writing method. After the first layer is etched, the second layer is exposed on a newly coated resist, referenced to the alignment marks made in the first layer. In Fig. 12*b* or 12*c*, two kinds of patterns are necessary for each wafer layer. The problems of phase-shift technology are the design of the shifter pattern and the complexity of the mask process. Other key technologies required to make phase-shifting masks are improved inspection and repair technologies for the shifter patterns and the opaque patterns.

Off-axis illumination[16] is another useful method to enhance the resolution and DOF. The basic principle is to tilt the illumination to an angle that passes through both the undiffracted light and the first-order diffracted light symmetrically as shown

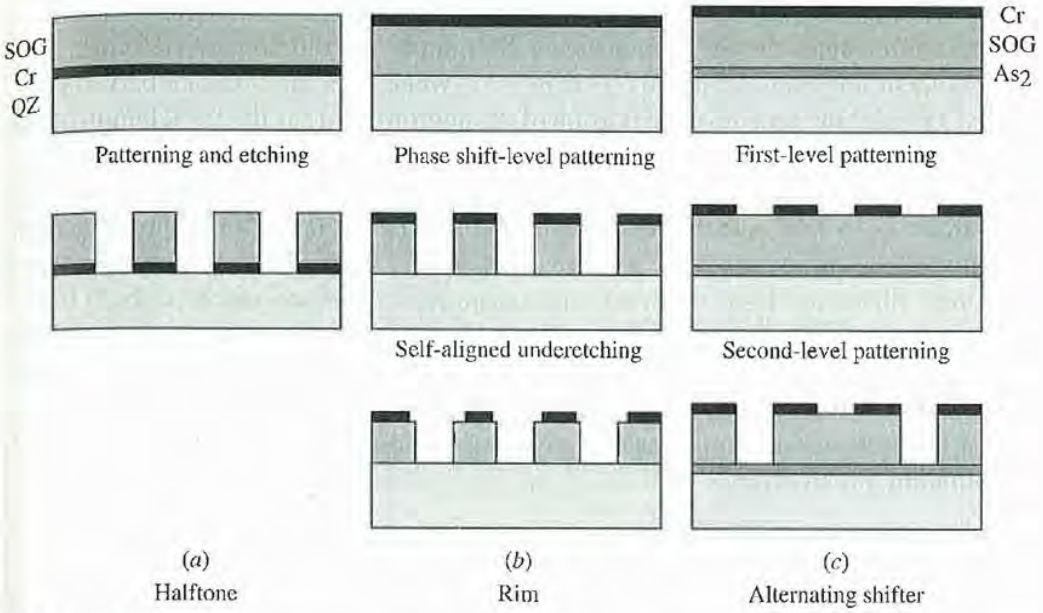


FIGURE 12
Examples of phase-shifting masks. (*After Ronse et al., Ref. 15.*) (*a*) Halftone. (*b*) Rim. (*c*) Alternating shifter.

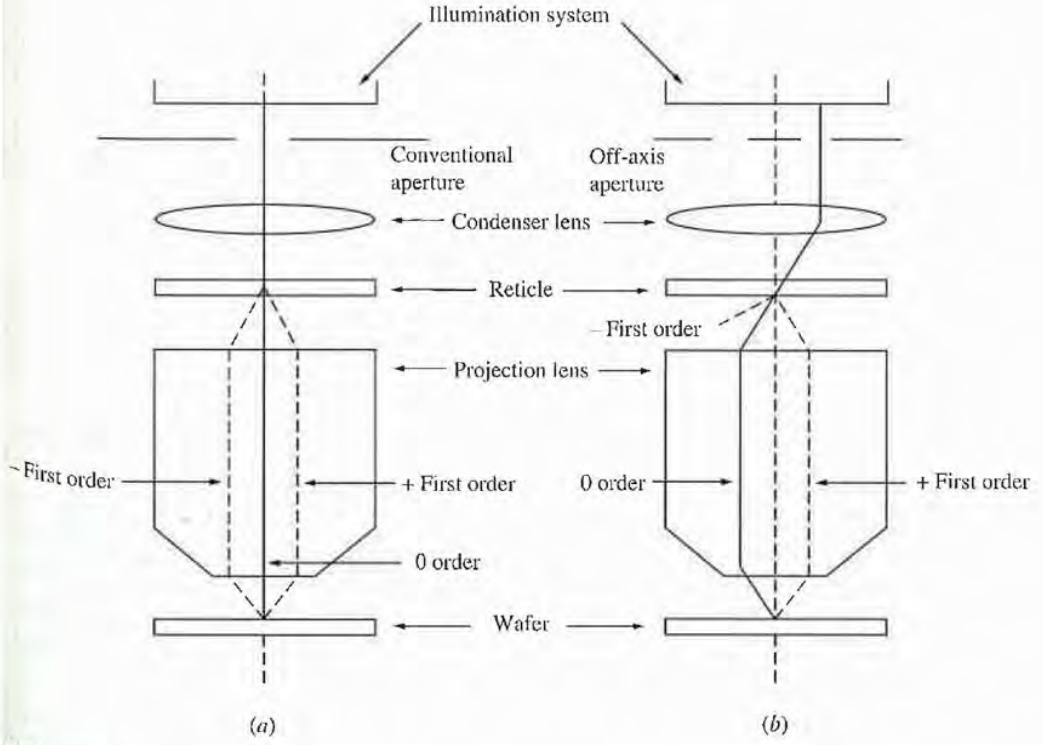


FIGURE 13
The off-axis illumination principle. (*After Shiraishi et al., Ref. 16.*) (*a*) Conventional illumination. (*b*) Off-axis illumination.

in Fig. 13*b*, compared with the conventional method in Fig. 13*a*. The light beams transmitted through the off-axis aperture illuminate the reticle pattern with a specific incident angle ϕ, defined as $f \sin \phi = x$, where x is the distance between the optical axis and the transmissive portion of the aperture and f is the focal length of the condenser. The light beams are diffracted by the reticle pattern. The −first-order or higher-order diffracted light will not enter the projection lens because the pitch of the pattern is so fine and the diffraction angle ($\sin \theta$) is larger than the numerical aperture of the projection lens, as explained in Fig. 5. As a result, only zero- and first-order diffracted light will interfere on the wafer surface and contribute to the image formation. It is reported[16] that this technique has a line-and-space resolution of 0.27 μm and a 2.8-μm DOF for a 0.35-μm line-and-space pattern with a 1.18-μm resist thickness. It is necessary to adjust the incident angle depending on pattern spatial frequencies and to increase the exposure time to compensate for the loss of illumination due to tilting.

6.2.4 Overlay Accuracy

To this point we have emphasized resolution, but overlay accuracy is another fundamental technology component.[12] Overlay error is generated mainly in two process steps—the alignment and the exposure. Many types of errors are included in the exposure process, such as the relative placement error between the mask and the wafer, distortion or magnification error of the lens, and mechanical instability of the exposure system caused by vibration, temperature drift, or atmospheric pressure drift.

Alignment methods are classified into two types: off-axis and TTL (through the lens). Off-axis offers the possibility of using nonactinic rays and of broad-band and high-NA viewing, with the flexibility of brightfield, darkfield, or even phase-contrast viewing. It is also easy to upgrade the off-axis alignment system because of its total separation from the imaging optics. However, off-axis requires high mechanical stability and precision as well as accurate means to refer the positions of the mask, the wafer, and the alignment microscopes to each other.

TTL offers the inherent advantage of a direct mask-to-wafer reference and thus relaxes the requirement for high mechanical stability and precision. However, some accuracy of the direct reference is lost if nonactinic alignment is used or if the wafer has to be moved into the exposure position after alignment. Both of these process steps require compensation.

Brightfield alignment suffers from multiple interferences of the wafer film stacks in case of TTL narrow-bandwidth illumination. The alignment signal is sensitive to variation in film thickness, on the order of 10 nm. This problem is reduced by using broad-band illumination. Alternatively, darkfield alignment removes the dependence on film thickness but does not completely nullify the effect of uneven resist coverage. Also, darkfield alignment has the potential problem of signal degradation on a grainy substrate. The solution to this problem is using the slit beam and the periodic alignment marks. One way to solve the aforementioned problems is to remove the resist over the alignment marks in the positive resist process, but this is impossible in the case of multilayer resist and negative resist processes.

An alignment system[17] is illustrated in Fig. 14. Off-axis global alignment measures the relative position of the wafer with a He-Ne laser. Die-by-die alignment is

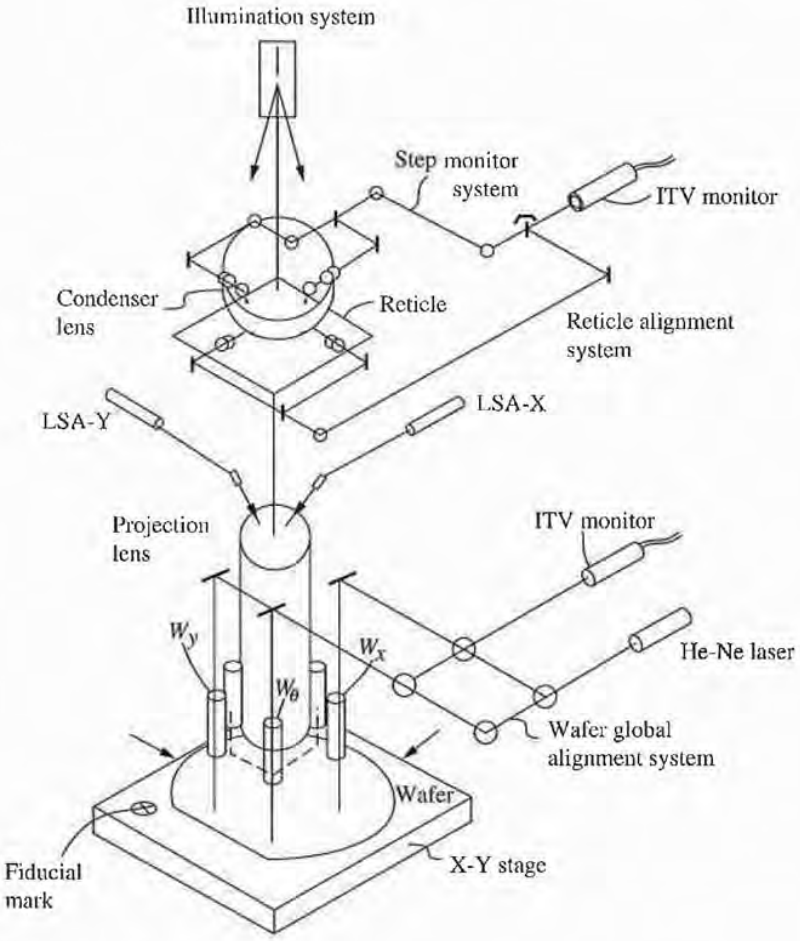


FIGURE 14
Wafer alignment system of an optical stepper. (*After Shiotake and Yoshida, Ref. 17.*)

more accurate because it uses an extra laser source (LSA-X, LSA-Y) through the projection optical path, i.e., the TTL method. Step-and-repeat exposure offers level-to-level registration, which is independent of wafer size, by separate alignment of each exposure field. In the reduction projection system, the mask pattern dimensions are larger than in systems imaged with unity magnification, and therefore it is more convenient to adjust the overlay position.

The basic sources of overlay error are listed in Table 2. The problem of overlay is common to both optical and x-ray lithography. X-ray lithography and e-beam direct write are considered distortionless. The registration error, defined as the ability to locate the alignment marks, is assumed to be identical for all optical alignment systems. Resist-induced alignment error is considered to be systematically removable. Wafer inplane distortion is caused by various hot processes. The major part of the magnification error of reduction systems is eliminated by fine tuning, but not easily, owing to the inherent symmetry of 1× optical systems. Coupled with the linear part of wafer inplane distortion, the magnification error of 1× optical systems can be as large as 0.15 μm. Similarly, x-ray systems have an error of 0.1 μm as a result

TABLE 2
Sources of overlay errors[12]

Items	5× optics (μm)	1× optics (μm)	1× x-ray (μm)	E-beam direct write (μm)
1. Lens distortion	0.07	0.06	0	0
2. Registration error	0.05	0.05	0.05	0.05
3. Wafer inplane distortion	0.05	0.05	0.05	0.05
4. Magnification error	0.01	0.1	0.1	0.01
5. Mask-to-mask placement error	0.05	0.05	0.05	0.05
6. Total contributions (RSS)	0.1	0.15	0.13	0.09
7. Total without lens distortion	0.07	0.13	0.13	0.09
8. Total without magnification error	0.1	0.11	0.09	0.09
9. Total without distortion or magnification error	0.07	0.09	0.09	0.09

of wafer inplane distortion and mask heating during exposure. The mask-to-mask placement error reflects the accuracy of e-beam mask making. For x-ray systems, the placement error of the membrane mask can be worse than for the optical mask, but it is not included here. In the case of e-beam direct write, the mask error is simply the composite placement error of the e-beam between two masking levels. In the case of 5× lithography, the mask-to-mask placement error is divided by 5 before being included in the RSS (square root of the sum of the squares) total of the contributions.

6.2.5 Optical Resists

Photoresist is the general term for polymers that can create patterns by the use of solvents after irradiation. The development of the resist is based on a chemical reaction and depends on the solubility difference between irradiated and unirradiated areas. Photoresists are of two types: negative, which on exposure to light become less soluble in a developer solution, and positive, which become more soluble.

Negative resists generally consist of a chemically inert polyisoprene rubber, which is the film-forming component, and a photoactive agent. The photoactive agent releases nitrogen gas on exposure to light, and the radicals generated react with the double bonds to form cross-links between rubber molecules, making the rubber less soluble in an organic developer solvent. The reactive species formed during the exposure can react with oxygen and be rendered ineffective for cross-linking, so an inert atmosphere is used. The developer solvent dissolves the unexposed resist. The exposed resist has low molecular weight, so it swells as the uncrosslinked molecules are dissolved. The swelling distorts the pattern features and limits resolution to 2 to 3 times the initial film thickness. For VLSI/ULSI applications, the use of negative resists has been supplanted by positive resists because negative resists have a resolution limit of about 2 μm, although they have many advantages such as resistance to etching and good adhesion to the substrate.

Positive resists have two components: a resin and a photoactive compound dissolved in a solvent. The photoactive compound is a dissolution inhibitor. When the photoactive compound is destroyed by exposure to light, the resin becomes more

soluble in an aqueous developer solution. The unexposed regions have high molecular weight and so do not swell much in the developer solution; therefore higher resolution, suitable for ULSI, is possible with positive resists. The development process of projection-printed images in positive resists has been modeled theoretically;[18] it is an isotropic etching or removal process.

One of the fundamental properties of a resist, which determines resolution, is the contrast.[19] Contrast, in the case of negative resist, is related to the rate of formation of a cross-linked network and may be measured by exposing pads of known area to varying radiation doses. D_g^i is the minimum dosage required to form the first insoluble film. Thereafter, the film thickness, and consequently the thickness remaining after development, increases with increasing dose until, ultimately, the thickness of the developed pad is not detectably different from the original film thickness (in practice it is slightly less owing to volume contraction during cross-linking). This process is shown schematically in Fig. 15*a*. The resist contrast γ_n is then defined as

$$\gamma_n = 1/(\log D_g^0 - \log D_g^i) \tag{6.13}$$

For a positive resist, the film thickness of the irradiated region that remains after development decreases with increasing dose until eventually a dose D_c is reached, which results in complete removal of the film on development, as shown in Fig. 15*b*. The contrast of positive resist γ_p is defined as

$$\gamma_p = 1/(\log D_c - \log D_0) \tag{6.14}$$

When photoresist films are exposed using monochromatic radiation, standing waves are formed in the resist.[20] These are caused by coherent interference effects due to a reflecting substrate. Coherent interference results in periodic intensity distributions in the direction perpendicular to the plane of the resist with a period $\lambda_r/2$, where λ_r is the wavelength in the resist. The refractive index of most resists is about 1.6, leading to an optical mismatch at the air/resist interface. Figure 16 shows

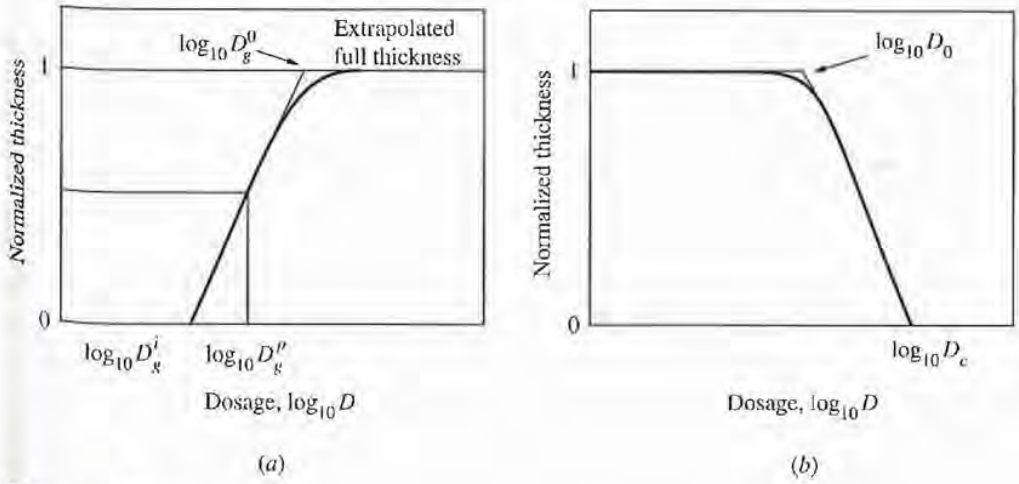


FIGURE 15
Sensitivity curves of negative and positive resists. (*After Thompson, Ref. 19.*) (*a*) Negative resist. (*b*) Positive resist.

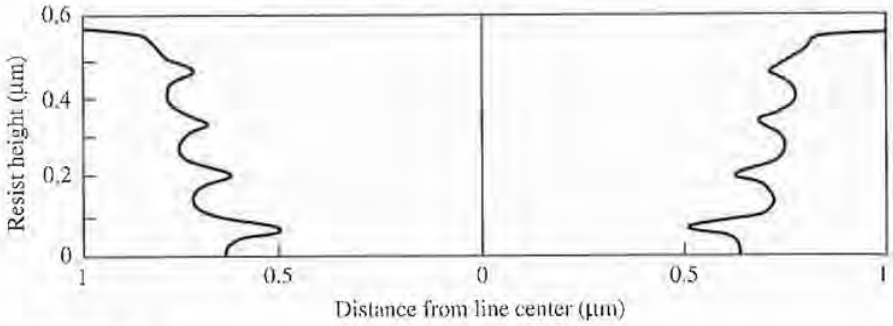


FIGURE 16
Simulated standing wave edge profile for a nominal 1-μm line in AZ1350 photoresist developed for 85 sec in 1 : 1 AZ developer : water. (*After Dill et al., Ref. 20.* © 1975 IEEE.) Substrate: Si + SiO_2, 60 nm; resist thickness: 583.6 nm; exposure wavelength: 435.8 nm (G-line); exposure energy: 57 mJ/cm^2

the standing wave. This variation in peak intensity with resist thickness becomes less with decreasing substrate reflectivity and increasing absorption by the resist. Since resist thickness varies at a step in the substrate topography, the resulting difference in effective exposure leads to size variations in the resist image.

Commercially available resists are listed in Table 3. The negative RD2000N-type resists, which have a diazide compound as a reactor with polyvinylphenol resin, have a resolution similar to a positive resist and do not swell on development. Photoresists are being developed for exposure at shorter wavelengths where higher resolution is possible. A few such deep-UV resists are PMMA, sensitive for $\lambda < 250$ nm; polybutene sulfone, sensitive for $\lambda < 200$ nm; and Microposit MP-2400, sensitive for $\lambda = 250$ nm. At these shorter wavelengths, the radiation quantum is large enough to produce scission of the molecular chain.

The other very important resists for deep UV are the chemically amplified resists, which exhibit high photo-speed, excellent resolution, and process tolerance.[21] The chemically amplified resist consists of a polymer host and a generator of acid in the presence of light. In the postexposure bake the photogenerated acid catalyzes thermal reactions that alter the solubility of the exposed region. Generally, resolution and sensitivity are affected by postexposure bake conditions. Some chemically amplified resists are applicable to electron lithography.

Other important properties of resists are adhesion to the substrate and resistance to wet and dry etch processing. In general, the commercially available optical resists are compatible with such processes.

Some inorganic materials are also effective as resists. Fundamental functions are based on the Ag-doping effect, which is induced by photoirradiation or electron-beam irradiation on a stacked-layer system of Ag and Se-Ge chalcogenide-glass film. When doped with Ag, the chalcogenide-glass films become almost insoluble in an alkaline solution. The light source, such as a Hg lamp, should have a wavelength that is shorter than the absorption edge of Se-Ge. This method is expected to have high resolution, with applications in dry processing.[22,23]

TABLE 3
Commercial optical resists

Name	Maker	Type	Chemical structure	Remarks
OMR83	Tokyo-oka	Negative	Isoprene rubber, azide compound	1× aligner
CIR	Nippon Synthetic Rubber	"	"	"
RD-2000N	Hitachi Chemical	"	Phenol resin, azide compound	Deep UV
MES-U	Nippon Synthetic Rubber	"	Chloropolystyrene	"
THMR-iN200	Tokyo-oka	"	Novolac resin, azide	I-line
TDUR-N7	"	"	Phenol resin	Deep UV
Microposit 1350J	Shipley	Positive	Novolac resin, O-kyniadiazide	G-line
AZ1350J	Hoechst	"	"	"
OFPR800	Tokyo-oka	"	"	"
Microposit2400	Shipley	"	"	Deep UV
ODUR-120	Tokyo-oka	"	PMMA	"
ODUR1000	"	"	PMIPK	I-line
THMR-IP3000	"	"	Novolac resin, azide	"
TSCR-80I	"	"	"	"
PFI28	Sumitomo Chemical	"	"	"
PFRIX060	Nippon Synthetic Rubber	"	"	"
PFRIX061	"	"	"	"
FHI-3950	Fuji Hunt	"	"	"
AZ7500	Hoechst	"	"	"
NPR-A18SH7	Nagase	"	"	"

6.2.6 Process Technologies

Increasing the NA and decreasing the wavelength are logical approaches to improving resolution. However, if these techniques are used, DOF decreases, based on Eq. (6.3). In the actual production process, many possible causes of interference to focusing arise. Some occur during the wafer process, such as surface topography and warping. Others are due to the exposure process, such as vibration, illumination nonuniformity, and aberrations. One solution to these problems is the combination of planarization and surface imaging. The thick bottom planarization layer reduces optical reflection and pattern-width degradation due to the surface topography beneath. The thin top layer enhances resolution, pattern-width control, and process latitude.

Surface-imaging resist processes are summarized in Fig. 17. The first item is treatment through a development cycle to enhance the contrast of the resist.[24] In this cycle the wafer is dipped in the developer, rinsed, and dried. This sequence is repeated until the resist film of the exposed area is completely dissolved. This interrupted development process promotes the formation of a passivation layer on the

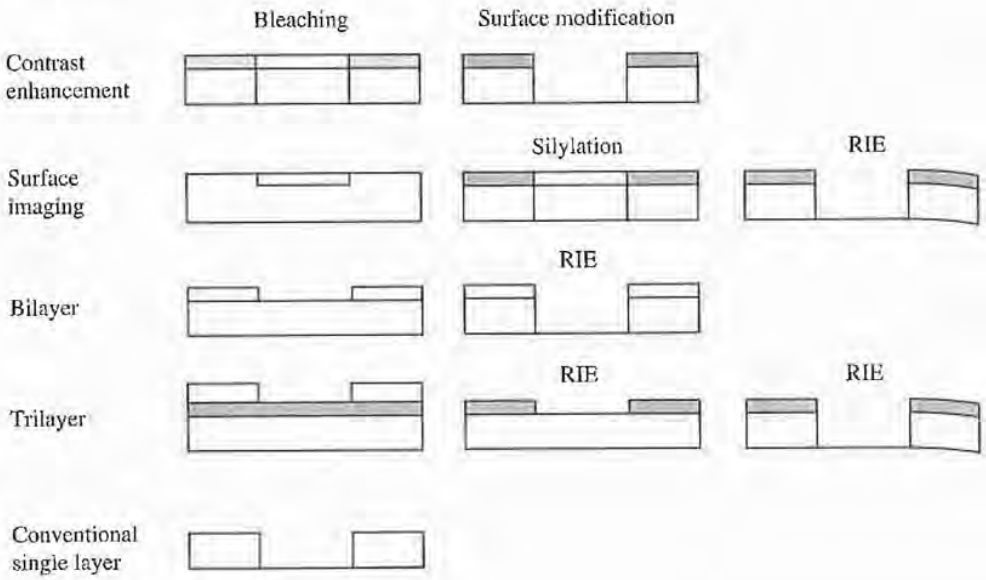


FIGURE 17
Advanced resist processes.

unexposed surface. As a result, the side walls become vertical and sensitivity is improved 30 to 50% in AZ resist. This technology is also used in electron lithography.

The second process in Fig. 17 is silylation.[25] A single layer of PLASMASK® (UCB from Belgium) resist is spincoated onto the substrate, with a thickness from 1.5 to 2.5 μm, and prebaked to a self-planarizing layer. After exposure, the wafers are treated with a gas-phase silylating agent (hexamethyldisilazane) at elevated temperature to reduce the silylation time. The resist material of the exposed areas selectively bonds chemically with silicon to a depth of 100–200 nm, and remains stable for a long time. The wafers are developed in oxygen plasma. During this treatment, the silicon is converted into silicon dioxide, which forms a thin protective mask that stops the etching of these exposed areas. It is reported that 0.35-μm features can be achieved using a 248-nm source.[26]

In the bilayer method, the image of the top layer formed by exposure and development is transferred to the bottom planarization layer by reactive ion etching (RIE). Resolution better than 0.5 μm is achieved by using a resist containing tungsten, which has a high resistance to etching, as the top layer.[26] There are some disadvantages in the bilayer method, such as intermixing between the resist layers and the formation of cracks while different materials are coated and baked.

The trilayer resist structure typically consists of an inorganic interlayer that is 0.1 to 0.2 μm thick, sandwiched between a top organic imaging layer that is 0.3 to 0.4 μm thick and a bottom organic planarization layer that is 1 to 2 μm thick. The trilayer method is more complex than the bilayer but is more flexible because of the intermediate layer.[27,28] Because of the intermediate barrier, the processes for the top layer and for the bottom layer are independently selectable, so many combinations of processes can be achieved by the proper choice of the intermediate layer materials and the top layer resists. The thickness of the top layer can be minimized to get high

resolution with enough etching duration for the intermediate layer (usually oxide). This method is also used in electron lithography.

6.3 ELECTRON LITHOGRAPHY

Electron lithography has the possibility of higher resolution than optical lithography because of the small wavelength (less than 1 angstrom) of the 10–50 keV electrons. Resolution in electron lithography systems is not limited by diffraction but by electron scattering in the target materials including the resist and by the various aberrations of the electron optics. Scanning electron beam pattern generators have been under development for more than 20 years and were derived from the scanning electron microscope.[29,30] Because of the serial nature of the pattern writing, throughput is much lower than for optical systems. However, a wide variety of applications is available in the pattern-generating function for electron lithography, such as mask fabrication for optical or x-ray lithography, direct writing on the wafers, and direct reaction with some materials on the substrate. Electron lithography is classified into two types—scanning and projection—and the scanning type can be either raster or vector scanning.

The exposure scheme is illustrated in Fig. 18. The ULSI pattern is composed by a computer-aided design (CAD) system. The output format from the CAD system is converted into the internal format of the individual exposure systems. The electron

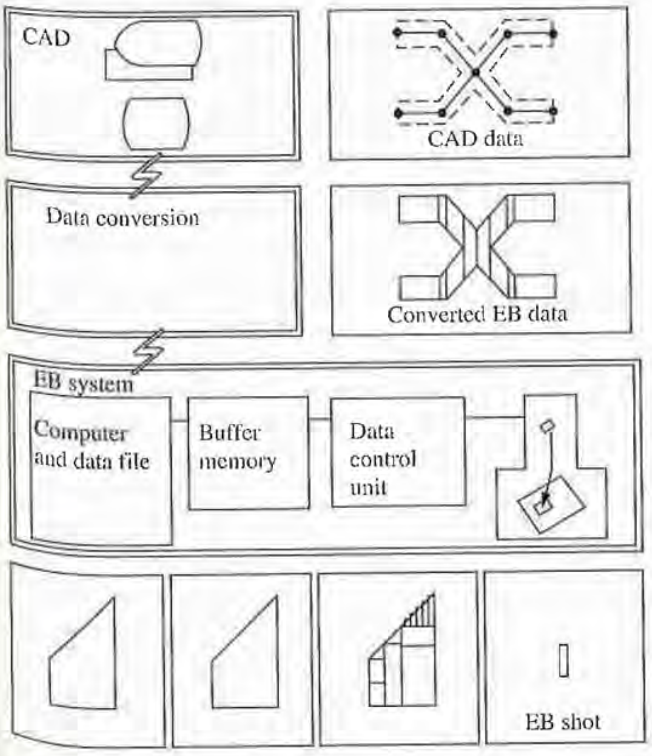


FIGURE 18
Flow of data in electron beam exposure systems.

issues are the pattern-width uniformity and the overlay accuracy. If the tolerances of these items are defined as 10% of the feature size, they should be less than 0.02 μm. It is very difficult to guarantee these accuracies because the accuracy of present measurement tools cannot meet this requirement.

Pattern-geometry accuracy is degraded in each step of the device manufacturing process. In general, the patterning sequences are (1) electron beam irradiation, (2) resist patterning on the mask, (3) chromium patterning on the mask, (4) stepper exposure, (5) resist patterning on the wafer, and (6) pattern transfer to a specified layer on the wafer. To minimize error, two methods are available. One is to skip some manufacturing steps, for example, using a direct reaction on the wafer by electron beam with gas molecules. The second is a statistical approach, which is used because of the difficulty in analyzing the cause of errors in the intermediate steps. This method has recently been applied to mask writing to improve the pattern-placement accuracy by multiple exposures. The more active approach is the evaluation and repair of function units of the final product, such as memory or logic cells, instead of inspecting and repairing in the intermediate fabrication steps such as masks.

It is very difficult to predict the trend in lithography beyond 0.2 μm or the architecture or structure of future ULSI. Optical technology is considered almost impossible to use for devices that have less than 0.2-μm geometry because of its resolution limit. We have only two choices: electron beam direct writing or an x-ray technology. However, it is very difficult to make perfect x-ray masks, either 1× transmission or 5× reflection, because of the complicated process steps including inspection and repair. In conclusion, the only available choice is electron beam direct writing. The throughput of this method is just a few 8-inch wafers per hour, as shown in Fig. 33, if the cell projection method is applied to production. For the mass production application, we must minimize the cost and footprint (required floor area) of the machine and develop high-sensitivity resists that are easy to process. It is considered possible that all of these requirements can be fulfilled.

REFERENCES

1. E. G. Cromer, Jr., "Mask Aligners and Steppers for Precision Microlithography," *Sol. State Technol.;* **36,** 23 (Apr. 1993).
2. B. J. Lin, "Deep UV Lithography," *J. Vac. Sci. Technol.,* **12,** 1317 (1975).
3. B. J. Lin, "Optical Methods for Fine Line Lithography," *Fine Line Lithography,* Roger Newman, Ed., North-Holland, Amsterdam, Ch. 2, 1980.
4. J. G. Skinner, "Some Relative Merits of Contact, Near-Contact, and Projection-Printing," *Proceedings of Microelectronics Seminar, Kodak Interface,* **73,** 53 (1974).
5. R. K. Watts, "Advanced Lithography," *Very Large Scale Integration, Fundamentals and Applications,* D. F. Barbe, Ed., Springer, New York, Ch. 2, 1982.
6. M. Born and E. Wolf, *Principles of Optics,* 6th ed., Pergamon Press, New York, 1980.
7. M. C. King and M. R. Goldrick, "Optical MTF Evaluation Techniques for Microelectronic Printers," *Sol. State Technol.,* **20,** 37 (Mar. 1977).
8. J. D. Cuthbert, "Optical Projection Printing," *Sol. State Technol.,* **20,** 59 (Aug. 1977).
9. M. Lacombat, J. Massin, G. M. Dubroeucq, and M. Brevignon, "Laser Projection Printing," *Sol. State Technol.,* **23,** 115 (Aug. 1980).

10. B. J. Lin, "The Optimum Numerical Aperture for Optical Projection Microlithography," *Proc. SPIE,* **1463,** 42 (1991).
11. W. J. Smith, "Modern Optical Engineering," *The Design of Optical Systems,* 2nd ed., McGraw-Hill, New York, Ch. 13, 1991.
12. B. J. Lin, "The Path to Subhalf-Micrometer Optical Lithography," *Proc. SPIE,* **922,** 256 (1988).
13. R. Unger, C. Sparkes, P. A. DiSessa, and D. J. Elliott, "Design and Performance of a Production-Worthy Excimer-Laser-Based Stepper," *Proc. SPIE,* **1674,** 708 (1992).
14. M. D. Levenson, N. S. Viswanathan, and R. A. Simpson, "Improving Resolution in Photolithography with a Phase-Shifting Mask," *IEEE Trans. Electron. Dev.,* **ED-29,** 1828 (1982).
15. K. Ronse, R. Jonckheere, C. Juffermans, and L. Van den Hove, "Comparison of Various Phase Shift Strategies and Application to 0.35 μm ASIC Designs," *Proc. SPIE,* **1927,** 2 (1993).
16. N. Shiraishi, S. Hirukawa, Y. Takeuchi, and N. Magome, "New Imaging Technique for 64M-DRAM," *Proc. SPIE,* **1674,** 741 (1992).
17. N. Shiotake and S. Yoshida, "Recent Advances of Optical Step-and-Repeat System," *Proc. SPIE,* **537,** 168 (1985).
18. F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of Positive Photoresist," *IEEE Trans. Electron. Dev.,* **ED-22,** 445 (1975).
19. L. F. Thompson, "Design of Polymer Resist for Electron Lithography," *Solid State Technol.,* **17,** 27 (1974).
20. F. H. Dill, A. R. Neureuther, J. A. Tuttle, and E. J. Walker, "Modeling Projection Printing of Positive Photoresists," *IEEE Trans. Electron. Dev.,* **ED-22,** 456 (1975).
21. H. Ito and C. G. Willson, "Applications of Photoinitiators to the Design of Resists for Semiconductor Manufacturing," *Polymer in Electronics,* ACS Symposium Series **242,** T. Davidson, ed., American Chemical Society, 11, 1984.
22. A. Yoshikawa, O. Ochi, and Y. Mizushima, "Dry Development of Se-Ge Inorganic Photoresist," *Appl. Phys. Lett.,* **36,** 107 (1980).
23. K. L. Tai, R. G. Vadimisky, C. T. Kemmerer, J. S. Wagner, V. E. Lamberti, and A. G. Timko, "Submicron Optical Lithography Using an Inorganic Resist/Polymer Bilevel Scheme," *J. Vac. Sci. Technol.,* **17,** 1169 (1980).
24. K. G. Chiong, K. Petrillo, F. J. Hohn, and A. D. Wilson, "Contrast and Sensitivity Enhancement of Resists for High-Resolution Lithography," *J. Vac. Sci. Technol.,* **B6,** 2238 (1988).
25. B. Roland, R. Lombaerts, C. Jakus, and F. Coopmans, "The Mechanism of the DESIRE Process," *Proc. SPIE,* **771,** 69 (1987).
26. G. R. Misium, M. Tipton, and C. M. Garza, "Surface Imaging Lithography at 248 nm," *J. Vac. Sci. Technol.,* **B8,** 1749 (1990).
27. J. M. Moran and D. Maydan, "High Resolution, Steep Profile Resist Patterns," *J. Vac. Sci. Technol.,* **16,** 1620 (1979).
28. L. E. Stillwagon, A. Kornblit, and G. N. Taylor, "Thin Titanium Dioxide Films as Interlayers in Trilayer-Resist Structures," *J. Vac. Sci. Technol.,* **B6,** 2229 (1988).
29. R. F. M. Thorney and T. Sun, "Electron Beam Exposure of Photoresists," *J. Electrochem. Soc.,* **112,** 1151 (1965).
30. H. C. Pfeiffer, "Recent Advances in Electron-Beam Lithography for the High-Volume Production of VLSI Devices," *IEEE Trans. Electron. Dev.,* **ED-26,** 663 (1979).
31. D. R. Herriott, R. J. Collier, D. S. Alles, and J. W. Stafford, "EBES: A Practical Electron Lithographic System," *IEEE Trans. Electron. Dev.,* **ED-22,** 385 (1975).
32. A. J. Speth, A. D. Wilson, A. Kern, and T. H. P. Chang, "Electron-Beam Lithography Using Vector-Scan Techniques," *J. Vac. Sci. Technol.,* **12,** 1235 (1975).

33. D. S. Alles, C. J. Biddick, J. H. Bruning, J. T. Clemens, R. J. Collier, E. A. Gere, L. R. Harriott, F. Leone, R. Liu, T. J. Mulrooney, R. J. Nielsen, N. Paras, R. M. Richman, C. M. Rose, D. P. Rosenfeld, D. E. A. Smith, and M. G. R. Thomson, "EBES4: A New Electron-Beam Exposure System," *J. Vac. Sci. Technol.*, **B5,** 47 (1987).
34. G. A. Burns and J. A. Schoeffel, "Performance Evaluation of the ATEQ CORE-2000 Scanning Laser Reticle Writer," *Proc. SPIE,* **772,** 269 (1987).
35. E. Goto, T. Soma, and M. Idesawa, "Design of a Variable-Aperture Projection and Scanning System for Electron Beam," *J. Vac. Sci. Technol.*, **15,** 883 (1978).
36. H. C. Pfeiffer, "Variable Spot Shaping for Electron-Beam Lithography," *J. Vac. Sci. Technol.*, **15,** 887 (1978).
37. A. M. Carroll and J. L. Freyer, "Measuring Performance of the AEBLE™150 Direct-Write e-Beam Lithography Equipment," *Proc. SPIE,* **537,** 25 (1985).
38. M. Fujinami, N. Shimazu, T. Hosokawa, and A. Shibayama, "EB60: An Advanced Direct Wafer Exposure Electron-Beam Lithography System for High-Throughput, High-Precision, Submicron Pattern Writing," *J. Vac. Sci. Technol.*, **B5,** 61 (1987).
39. T. Takigawa, H. Wada, Y. Ogawa, R. Yoshikawa, I. Mori, and T. Abe, "Advanced e-Beam Lithography," *J. Vac. Sci. Technol.*, **B9,** 2981 (1991).
40. K. Meissner, W. Haug, S. Silverman, and S. Sonchik, "Electron-Beam Proximity Printing of Half-Micron Devices," *J. Vac. Sci. Technol.*, **B7,** 1443 (1989).
41. R. Ward, A. R. Franklin, I. H. Lewin, P. A. Gould, and M. J. Plummer, "A 1 : 1 Electronstepper," *J. Vac. Sci. Technol.*, **B4,** 89 (1986).
42. Y. Sakitani, H. Yoda, H. Todokoro, Y. Shibata, T. Yamazaki, K. Ohbitsu, N. Saitou, S. Moriyama, S. Okazaki, G. Matsuoka, F. Murai, and M. Okumura, "Electron-Beam Cell-Projection Lithography System," *J. Vac. Sci. Technol.*, **B10,** 2759 (1992).
43. H. Y. Liu, M. P. deGrandpre, and W. E. Feely, "Characterization of a High-Resolution Novolak Based Negative Electron-Beam Resist with 4 $\mu C/cm^2$ Sensitivity," *J. Vac. Sci. Technol.*, **B6,** 379 (1988).
44. D. F. Kyser and N. S. Viswanathan, "Monte Carlo Simulation of Spatially Distributed Beams in Electron-Beam Lithography," *J. Vac. Sci. Technol.*, **12,** 1305 (1975).
45. M. Parikh, "Self-Consistent Proximity Effect Correction Technique for Resist Exposure (SPECTRE)," *J. Vac. Sci. Technol.*, **15,** 931 (1978).
46. F. Murai, S. Okazaki, N. Saito, and M. Dan, "The Effect of Acceleration Voltage on Linewidth Control with a Variable-Shaped Electron Beam System," *J. Vac. Sci. Technol.*, **B5,** 105 (1987).
47. E. Kratschmer and T. R. Groves, "Resist Heating Effects in 25 and 50 kV e-Beam Lithography on Glass Masks," *J. Vac. Sci. Technol.*, **B8,** 1898 (1990).
48. M. Fujita, K. Shiozawa, T. Kase, H. Hayakawa, F. Mizuno, R. Haruta, F. Murai, and S. Okazaki, "Application and Evaluation of Direct-Write Electron Beam for ASIC's," *IEEE J. Solid State Circuits,* **23,** 514 (1988).
49. F. Murai, Y. Nakayama, I. Sakama, T. Kaga, Y. Nakagome, Y. Kawamoto, and S. Okazaki, "Electron Beam Direct Writing Technology for 64-Mb DRAM LSIs," *J. Appl. Phys.*, **29,** 2590 (1990).
50. N. Samoto, Y. Makino, K. Onda, E. Mizuki, and T. Itoh, "A Novel Electron-Beam Exposure Technique for 0.1 μm T-Shaped Gate Fabrication," *J. Vac. Sci. Technol.*, **B8,** 1335 (1990).
51. K. L. Lee and M. Hatzakis, "Direct Electron-Beam Patterning for Nanolithography," *J. Vac. Sci. Technol.*, **B7,** 1941 (1989).
52. E. M. Clausen, J. P. Harbison, C. C. Chang, H. G. Craighead, and L. T. Florez, "Electron Beam Induced Modification of GaAs Surfaces for Maskless Thermal Cl_2 Etching," *J. Vac. Sci. Technol.*, **B8,** 1830 (1990).

53. D. L. Spears and H. I. Smith, "High-Resolution Pattern Replication Using Soft X Rays," *Electron. Lett.*, **8,** 102 (1972).
54. D. Fleming, J. R. Maldonado, and M. Neisser, "Prospect for X-Ray Lithography," *J. Vac. Sci. Technol.*, **B10,** 2511 (1992).
55. R. Viswanathan, R. E. Acosta, D. Seeger, H. Voelker, A. Wilson, I. Babich, J. Maldonado, J. Warlaumont, O. Vladimirsky, F. Hohn, D. Crockatt, and R. Fair, "Fully Scaled 0.5 μm Metal-Oxide Semiconductor Circuits by Synchrotron X-Ray Lithography: Mask Fabrication and Characterization," *J. Vac. Sci. Technol.*, **B6,** 2196 (1988).
56. I. Okada, Y. Saitoh, S. Itabashi, and H. Yoshihara, "A Plasma X-Ray Source for X-Ray Lithography," *J. Vac. Sci. Technol.*, **B4,** 243 (1986).
57. A. Heuberger, "X-Ray Lithography," *Sol. State Technol.*, **29,** 93 (Feb. 1986).
58. B. S. Fay and W. T. Novac, "Advanced X-Ray Alignment System," *Proc. SPIE,* **632,** 146 (1986).
59. R. B. McIntosh, G. P. Hughes, J. L. Kreuzer, and G. R. Conti, "X-Ray Step-and-Repeat Lithography System for Submicron VLSI," *Proc. SPIE,* **632,** 156 (1986).
60. C. N. Archie, J. I. Granlund, R. W. Hill, K. W. Kukkonen, J. A. Leavey, L. G. Lesoine, J. M. Oberschmidt, A. E. Palumbo, C. Wasik, M. Q. Barton, J. P. Silverman, J. M. Warlaumont, A. D. Wilson, R. J. Anderson, N. C. Crosland, A. R. Jorden, V. C. Kempson, J. Schouten, A. I. C. Smith, M. C. Townsend, J. Uythoven, M. C. Wilson, M. N. Wilson, D. E. Andrews, R. Palmer, R. Webber, and A. J. Weger, "Installation and Early Operating Experience with the Helios Compact Synchrotron X-Ray Source," *J. Vac. Sci. Technol.*, **B10,** 3224 (1992).
61. A. C. Chen, C. J. Progler, F. F. Couch, T. A. Gunther, H. Fair, and K. A. Cooper, "First X-Ray Stepper in IBM Advanced Lithography Facility," *J. Vac. Sci. Technol.*, **B10,** 2628 (1992).
62. J. Conway, C. Alcorn, D. Patel, J. Ricker, R. Yandow, L. Hsia, and A. Flamholz, "Fabrication of High-Density SRAM Chips Using Mix-and-Match X-Ray Lithography," *Proc. SPIE,* **1924,** 309 (1993).
63. A. M. Hawryluk and L. G. Seppala, "Soft X-Ray Projection Lithography Using an X-Ray Reduction Camera," *J. Vac. Sci. Technol.*, **B6,** 2162 (1988).
64. M. Komuro, N. Atoda, and H. Kawakatsu, "Ion Beam Exposure of Resist Materials," *J. Electrochem. Soc.*, **126,** 483 (1979).
65. R. L. Seliger, J. W. Ward, V. Wang, and R. L. Kubena, "A High-Intensity Scanning Ion Probe with Submicrometer Spot Size," *Appl. Phys. Lett.*, **34,** 310 (1979).
66. G. Stengel, G. Bosch, A. Chalupka, J. Fegel, R. Fischer, G. Lammer, H. Loscher, L. Malek, R. Nowak, C. Traher, P. Wolf, P. Mauger, A. Shimkunas, S. Sen, and J. C. Wolfe, "Ion Projector Wafer Exposure Results at 5× Ion-Optical Reduction Obtained with Nickel and Silicon Stencil Masks," *J. Vac. Sci. Technol.*, **B10,** 2824 (1992).
67. J. L. Bartelt, "Masked Ion Beam Lithography: An Emerging Technology," *Sol. State Technol.*, **29,** 215 (May 1986).
68. S. Okazaki, "Resolution Limits of Optical Lithography," *J. Vac. Sci. Technol.*, **B9,** 2829 (1991).

PROBLEMS

1. A proximity printer operates with a 10-μm mask-to-wafer gap and a wavelength of 436 nm. Another printer uses a 30-μm gap with wavelength 365 nm. Which offers higher resolution?

CHAPTER 7

Etching

Y. J. T. Lii

7.1 INTRODUCTION

Devices are built from a number of different layer materials deposited sequentially. The lithography techniques described in Chapter 6 are used to replicate circuit and device features, and the desired pattern is transferred by means of etching. As we enter the era of ULSI manufacturing, the etching process grows more important for fabricating features with sub-half-micrometer dimensions. Dry etching is synonymous with plasma-assisted etching, which encompasses several techniques that use plasmas in the form of low-pressure gas discharges. Integrated circuit fabrication processes that use reactive plasmas are commonplace in today's semiconductor production lines because of their potential for very-high-accuracy transfer of resist patterns, i.e., anisotropic etching, as shown in Fig. 1. By contrast, wet chemical etching results in isotropic etching, where both vertical and lateral etch rates are comparable.

In the early 1970s, photo lithography with Novolak-based resists was used to pattern devices with dimensions less than three micrometers. Oxygen plasmas were explored for stripping of resist. The need to maintain dimensional tolerance during etching became a critical requirement that led to numerous advances in plasma processing and improved reactors during the 1980s. Today, there is a broad base of empirical knowledge and some qualitative understanding of etch mechanisms.[1] These insights and systematic experiments led to the process developments needed to meet the requirements of today's submicrometer device manufacturing.

Etch techniques consist of dry and wet etch methods. Dry etch methods include plasma etching, reactive ion etching, sputtering, ion beam etching, and reactive ion beam etching. For dry etch methods, this chapter concentrates only on plasma and reactive ion etching because they are the most popular methods used in semiconductor industry. This chapter starts with a discussion of the fundamentals of

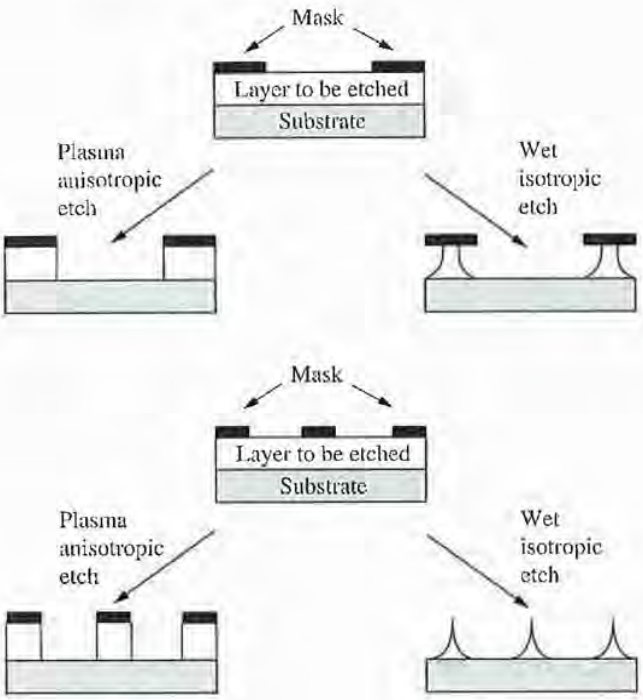


FIGURE 1
Illustration of the difference between plasma (anisotropic) and wet (isotropic) etching.

low-pressure discharge. Next, fluorocarbon plasma chemistry is introduced to review the means for achieving etching directionality and selectivity in reactive etching using glow discharges. Specific plasma processes used to etch different materials important to semiconductor processing are reviewed. Plasma diagnostics, process control, and plasma-induced damage are introduced briefly. In addition, relevant trends in new, magnetically enhanced reactive ion etching systems, in high-density plasma systems, and in integrated etching clustering are discussed. Finally, alternative wet etching methods used in the semiconductor industry are briefly presented.

7.2 LOW-PRESSURE GAS DISCHARGE

Plasma is an ionized gas with approximately equal amounts of positively charged particles (positive ions) and negatively charged particles (usually electrons and a small amount of negative ions). The plasma useful to ULSI processing is a weakly ionized plasma, called a "glow discharge," containing a significant density of neutral particles—more than 90% in most etchers. Although plasma is neutral in a macroscopic sense, it behaves quite differently from a molecular gas, because it consists of charged particles that can be influenced by applied electric and magnetic fields.

Plasma etching systems are extremely complex, and many details of the physical and chemical interactions, both within the plasma and on the surface exposed to the plasma, are not yet understood. The interaction of plasma with surfaces will be discussed in order to describe some of the mechanisms thought to be important in etching. A comprehensive review of this subject can be found in Ref. 2.

7.2.1 AC Plasma Discharge

In plasma etching, a plasma is created by applying a voltage across two electrodes, between which a gas is confined at low pressure. Simple dc power can be used to generate plasma; however, insulating materials cannot be used to cover electrode surfaces, because the insulating materials stop the current flow between the two electrodes and cause surface charging. To deal with the surface charging problem, an ac discharge is used so that the positive charge accumulated during one half-cycle can be neutralized by electron bombardment during the next half-cycle. In plasma etching, an rf field is normally used to generate the gas discharge. One reason for doing so is that the electrodes do not have to be made of a conducting material. The other reason is that electrons can pick up sufficient energy during field oscillation to cause more ionization by electron–neutral atom collisions. As a result, the plasma can be generated at pressures as low as 10^{-3} torr.

The free electrons in the gas are accelerated by the electric field and gain energy. Energy is released when free electrons collide with gas atoms; the gas atoms are then ionized, and more electrons are freed. Electron collisions with neutrals in plasma cause dissociation, ionization, and light emission. These collisions generate atoms, molecules, molecular fragments, and ions. The ions and electrons generated from these collisions raise the density of the plasma. Additional electrons are generated by secondary emission from energetic, positive ions colliding with the electrode surfaces. However, some collisions on the electrode surfaces lower plasma density, such as an ion-ion recombination and an electron-ion recombination. The process continues until equilibrium is reached, that is, until the number of ion-electron pairs created is the same as the number of pairs lost at the electrode surface and the chamber wall. At this point, the plasma is self-sustaining. The density for the plasmas of interest ranges from 10^9 to 10^{12} cm^{-3}. The density of gas molecules at 1 torr is about 10^{16} cm^{-3}, so these discharges are weakly ionized. Therefore, at low pressure, plasma is a partially ionized gas containing electrons, ions, and neutral atoms and/or molecules.

7.2.2 Properties of RF Discharge

When a plasma comes in contact with an electrode or surface, a dark space or *sheath* is formed next to it. The potentials of the rf discharge near the electrodes are important in determining the energies of ions incident on surfaces in the plasma.[3] The potentials near the electrodes, as shown in Fig. 2, result from currents of different particles, specifically electrons and ions. A light particle such as an electron will accelerate more than a heavier particle such as a positive ion in a given electric field. A large number of electrons rush toward electrodes after an electric field is applied. The depletion of electrons from the plasma and a negative charge buildup on the electrodes are inhibited by the development of a negative charge barrier that repels electrons and attracts positive ions. In Fig. 2, V_c is the potential at the rf-powered electrode measured with respect to ground, and V_p is the plasma potential with respect to ground. The potentials across the ion sheaths are $V_p - V_c$ at the rf-powered electrode

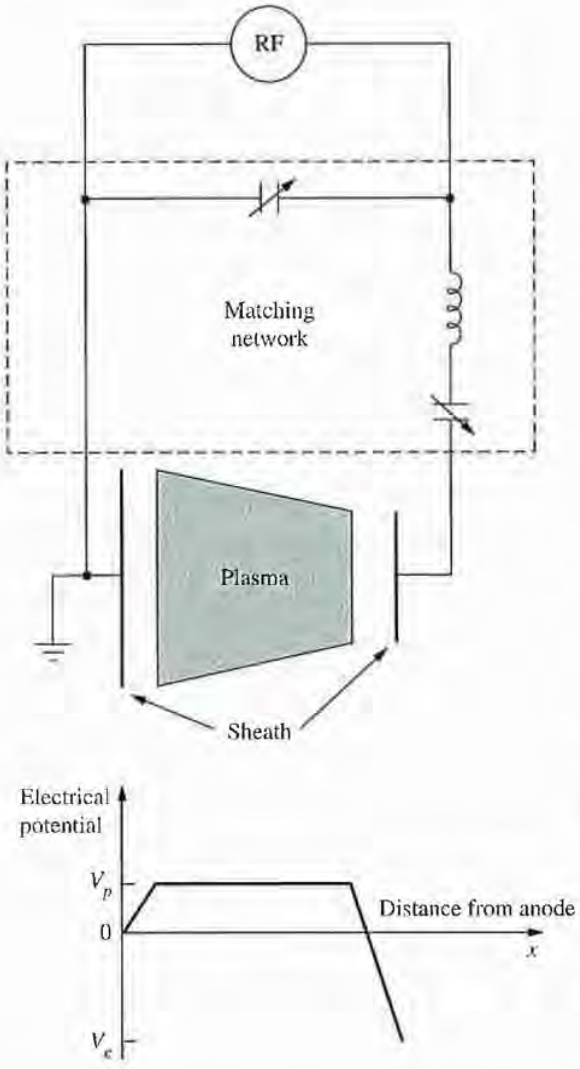


FIGURE 2
Representation of a generic, reactive-ion plasma etcher. RF power is supplied to the electrode supporting the wafer by a generator and matching network. The potential is shown as a function of position in the discharge for the case in which the area of the powered electrode is much less than the area of all grounded surfaces.

and V_p at the grounded surface. The potential of the electrode or surface with respect to that of the plasma determines the maximum energy of the ions bombarding that electrode or surface.

To a first approximation, the sheaths are treated as capacitors. The current density of the positive ions across sheaths is uniform and is equal at both electrodes. The ratio, R, of the area of the rf-powered electrode to the area of all grounded surfaces in contact with the plasma is a key parameter in determining how the applied voltage is distributed among the plasma sheaths. The potential $V_p - V_c$ increases as R decreases. A classic theory of plasma characteristics can be expressed as[2]

$$\frac{V_p - V_c}{V_p} = \left(\frac{A_a}{A_c}\right)^4 \tag{7.1}$$

where A_a and A_c are the grounded surface and rf-powered electrode areas, respectively. The sheath voltage near the small rf-powered electrode is larger than the

sheath voltage near the grounded surface. In a typical diode reactive-ion etcher, $V_p - V_c$ can be in the several-hundred-volt range when V_p is less than 100 volts.

7.2.3 Matching Network

A matching network is usually located between the rf power supply and the plasma, as shown in Fig. 2. The purpose of this network is to increase the power dissipation in the discharge and to protect the rf power supply. An rf power supply normally has a purely resistive output with a usual value of 50 ohms. However, an rf plasma normally has a partly capacitive impedance. Therefore, a load equal to the impedance of the rf power-supply output can be simulated by combining the plasma impedance with a variable matching network load. Most plasma etching systems lose power in the matching network; but with careful system design, the power loss can be minimized.

7.2.4 Glow Discharge Chemical Phenomena

Electron energies achieved at low pressure can be so high that they dissociate molecules into free radicals, ions, and new electrons, thus cascading the electrical breakdown of the gas and igniting or sustaining the plasma. At rf frequencies higher than 1 MHz, the ions, because of their low mobility, do not respond to the rf field instantaneously. Electrons undergo primarily inelastic electron-molecule collisions, which supply fresh species to the plasma. The steady-state constitution of a glow discharge is governed by the equilibrium of the rates of production and loss of the various species.

Two main reactions take place in a plasma system:

1. Homogeneous reactions in the bulk plasma, caused mainly by electron impact with neutral species
2. Heterogeneous reactions at the surface contact with the plasma, initiated by ions, electrons, and neutrals

Some examples of the most important types of electron–neutral particle collisions are summarized here:

Simple ionization

$$Ar + e^- \rightarrow Ar^+ + 2e^- \tag{7.2}$$

$$O_2 + e^- \rightarrow O_2^+ + 2e^- \tag{7.3}$$

In molecular gases, ionization may be concurrent with fragmentation; in this case, dissociative ionization is said to occur.

Dissociation

$$CF_4 + e^- \rightarrow CF_3^+ + F + 2e^- \tag{7.4}$$

$$CF_4 + e^- \rightarrow CF_3^+ + F^- + e^- \tag{7.5}$$

Light emission

$$O_2 + e^- \rightarrow O_2^* + e^- \tag{7.6}$$

$$O_2^* \rightarrow O_2 + h\nu \tag{7.7}$$

In the plasma, each formation step balances various loss processes. The equilibrium between formation and loss determines the steady concentrations of species in a discharge. For the charged species in a plasma, these formation and loss processes may be grouped into a few categories as follows:

Electron-ion recombination $O_2^+ + e^- \rightarrow 2O$ (7.8)

Ion-ion recombination $CF_3^+ + F^- \rightarrow CF_4$ (7.9)

Dissociative attachment $e^- + CF_4 \rightarrow CF_3 + F^-$ (7.10)

Atoms and radicals can be lost either by homogeneous reactions or by heterogeneous reactions. The type of loss reaction that will dominate for any species depends on many factors such as pressure, the type of surface, and the ratio of surface area to the volume of discharge.

7.3 ETCH MECHANISMS, SELECTIVITY, AND PROFILE CONTROL

Much of the early work in plasma etching focused on the discovery and optimization of fluorocarbon chemistry.[4] Although this initial work provided many useful industrial etching processes, an understanding of the associated etch mechanisms was typically lacking. Plasma etching is a process in which a solid film is removed by a chemical reaction with ground-state or excited-state neutral species, and plasma etching is often enhanced or induced by energetic ions generated in a gaseous discharge. Some important processing issues discussed in this section include basic etch mechanisms, etch selectivity, and profile control.

7.3.1 Etch Mechanisms

Unlike ion sputtering, which can be compared to sandblasting on an atomic scale, plasma etching methods are based on chemical reactions between the solid to be etched and an active species that arrives on the surface of the solid from the plasma. The microscopic processes of plasma etching of a silicon wafer are shown schematically in Fig. 3.[5] Electrons are accelerated by an applied rf field, and the energetic electrons collide with gas atoms and molecules. Some atoms and molecules become electronically excited, and light emission (glow) is produced from their deexcitation. Other atoms and molecules are dissociated or ionized to form radicals, atoms, and ions by colliding with high-energy electrons. The active species are then transported to the wafer surface where they are adsorbed. The adsorbed species can react with the wafer to form etch products or desorb from the wafer surface without reaction. Etch products then desorb into a gas phase if they are volatile. A brief summary of processes that occur during plasma etching is that a molecular gas, usually relatively inert, is used to generate reactive species, such as F or Cl radicals. By choosing proper chemistry, the radicals react with the wafer surface to form "volatile" products. The chemical compound formed must be volatile with a high vapor pressure

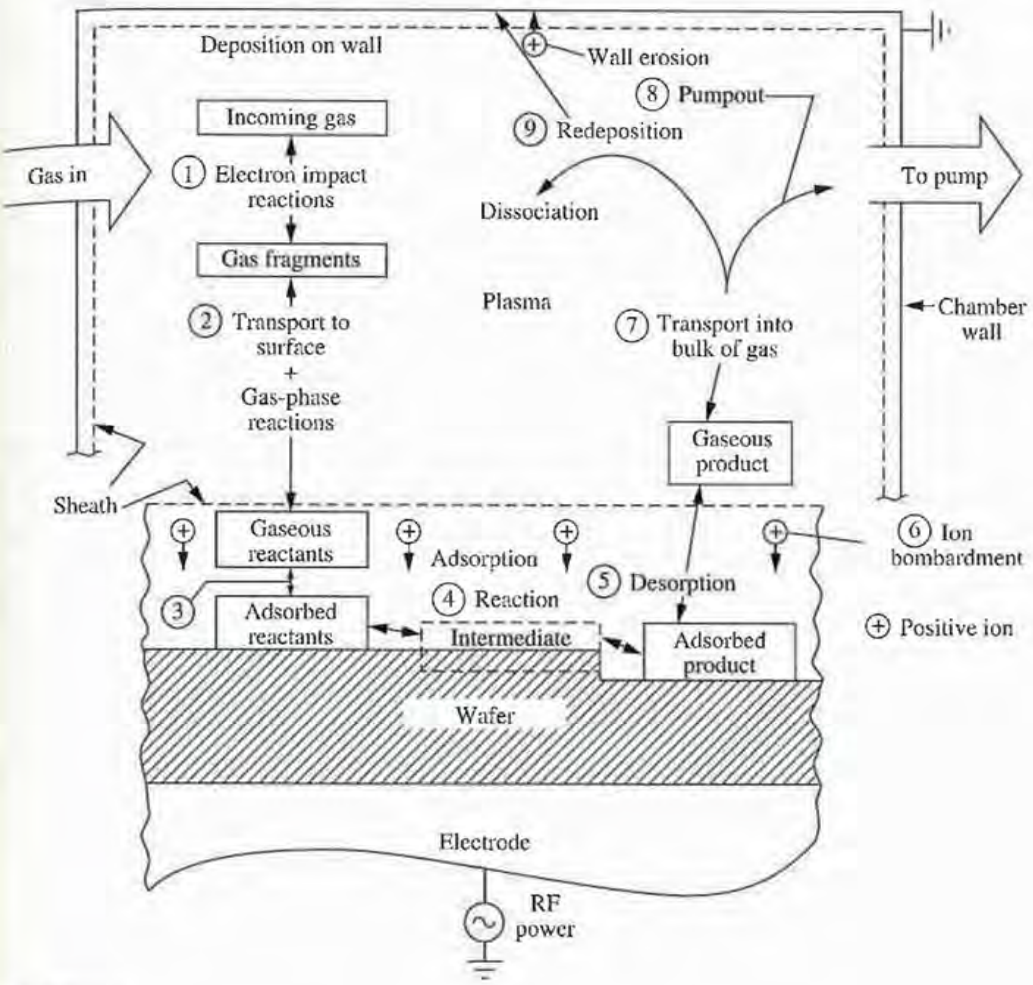


FIGURE 3
Schematic view of microscopic processes that occur during etching of a silicon wafer. (*After Oehrlein and Rembetski, Ref. 5.*)

and stable with a high bonding energy so that it is not dissociated in the plasma and, consequently, can be pumped out of the reactor.

Plasma etching comprises chemical, ion-sputter, and ion-enhanced plasma etching. All these methods are based on the generation of plasma by an rf discharge in a gas at low pressure. But two basic methods can be distinguished—physical methods and chemical methods. The former include sputter etching and the latter include pure chemical etching. There is also a hybrid technique, which uses a physical method to enhance chemical etching. In chemical etching, neutral reactive species generated by the plasma interact with the material's surface to form volatile products. In physical etching, positive ions bombard the surface at high speed; small amounts of negative ions formed in the plasma cannot reach the wafer surface and therefore play no direct role in plasma etching. Chemical and physical etch mechanisms have different basic characteristics. Chemical etching exhibits a high etch rate and good selectivity, produces low ion-bombardment-induced damage, and yields isotropic profiles. Physical etching can yield anisotropic profiles, but it is associated with low

etch selectivity and high bombardment-induced damage. Combinations of chemical and physical etching give anisotropic etch profiles, reasonably good selectivity, and moderate bombardment-induced damage.

7.3.2 Fluorocarbon Chemistry

In the early 1980s, molecular gases containing one or more fluorine atoms were used for plasma etching of silicon and silicon compounds. These gases were selected because they are often nontoxic, and fragments produced in a plasma react with silicon or silicon compounds to form volatile compounds at room temperature. For example, CF_4 is extremely stable but dissociates into F atoms and fluorinated fragments (CF_x) in a plasma. Atomic F is the active etchant for Si and SiO_2 by the formation of the volatile SiF_4 and O_2.

$$Si + 4F \rightarrow SiF_4 \tag{7.11}$$

$$SiO_2 + 4F \rightarrow SiF_4 + O_2 \tag{7.12}$$

Energetic electrons colliding with CF_4 are responsible for the generation of radicals, atoms, and ions. However, these radicals and atoms can react with other species, such as gas molecules, to alter the overall chemistry. For example, the etch rates of Si increase dramatically, as shown in Fig. 4,[6] if O_2 is added to CF_4. A maximum

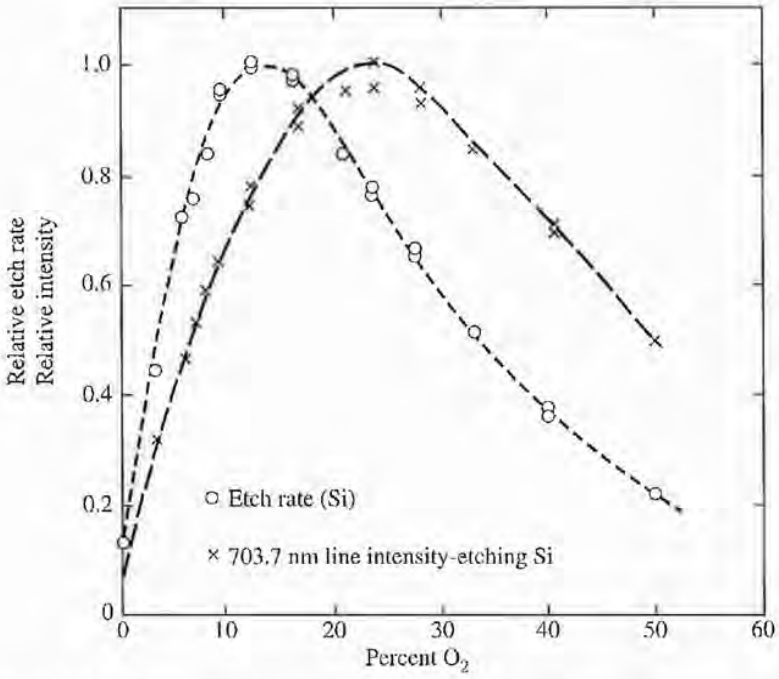


FIGURE 4
Dependence of both the normalized silicon etch rate and the normalized intensity of the emission from electronically excited F atoms on oxygen concentration in the CF_4/O_2 mixture. (*After Mogab, Adams, and Flamm, Ref. 6.*)

etch rate occurs with a CF_4/O_2 mixture that has 12% O_2. A similar effect is observed for SiO_2, where the maximum etch rate is reached at 20% O_2. These effects can be explained by the decreasing recombination rate of F and CF_x ($x \leq 3$). Adding O_2 to the plasma leads to the formation of COF_2, CO, and CO_2, which decreases the concentration of CF_x, thereby decreasing the recombination rate of F and CF_x. This results in a higher F concentration and, therefore, higher Si and SiO_2 etch rates. F concentration increases until the effect of O_2 dilution by CF_4 becomes dominant. This point is reached more quickly with Si etching because oxygen chemsorbs on the Si surface more readily than on the SiO_2 surface thereby blocking the reaction of F with Si.

When H_2 is added to the CF_4 plasma, quite different etch rate changes are observed. Figure 5 shows the etch rates of Si and SiO_2 in a CF_4/H_2 plasma as a function of H_2 concentration.[7] The SiO_2 etch rate is relatively insensitive to the H_2 concentration, but the Si etch rate decreases significantly as the H_2 concentration increases. When the H_2 concentration reaches 40%, the silicon etch stops. A generally accepted explanation of these effects is that two main mechanisms result from H_2 addition. First, H_2 reacts with F to form HF. This reaction consumes the F concentration and decreases the Si etch rate.

$$H + F \rightarrow HF \tag{7.13}$$

Second, the reaction of CF_x with SiO_2 forms etch products SiF_4, CO, CO_2, COF_2, etc. However, the reaction of CF_4 with Si forms a carbon or hydrocarbon polymer on the silicon surface, which blocks silicon etching by F.

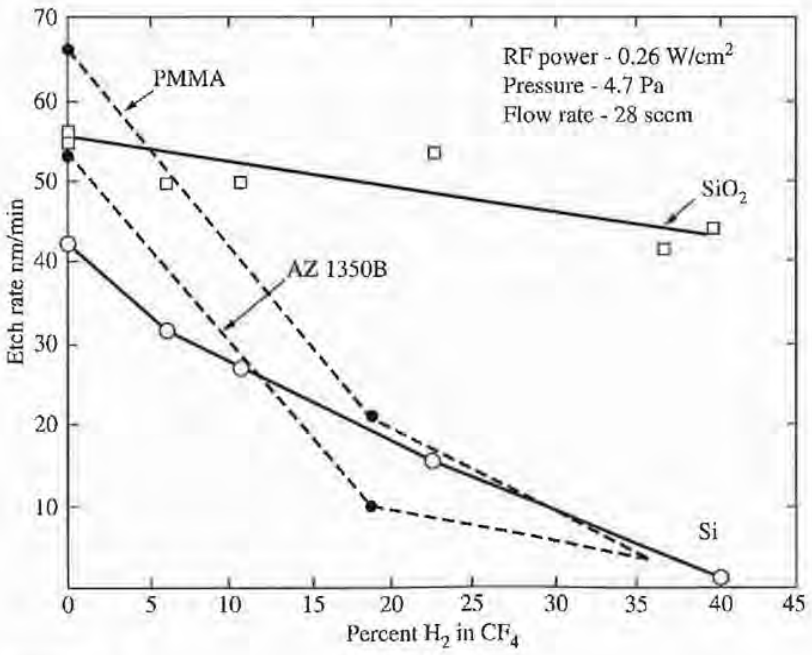


FIGURE 5
Etch rates of silicon (solid line, round points), silicon dioxide (solid line, square points), and PMMA and AZ1350B resists (dashed lines) vs. percentage of H_2 in CF_4. (*After Ephrath, Ref. 7.*)

$$CF_4 + SiO_2 \rightarrow SiF_4 + (CO, CO_2, COF_2) \quad (7.14)$$

$$CF_4 + Si \rightarrow SiF_x + (C, CF_y, CH_xF_y) \quad (7.15)$$

Similarly, CHF_3 can be used to replace the CF_4/H_2 mixture for selective silicon dioxide–to-silicon etch. A mixture of CHF_3/CF_4 is often used in industry for selective silicon dioxide–to-silicon etch, because etch selectivity can be controlled by changing the ratio of CHF_3 to CF_4 concentrations.

7.3.3 Anisotropy

Etch profiles are extremely important for submicrometer semiconductor fabrication, and careful control of the surface etch processes is required to ensure directional etching. Etch mechanisms can be understood by analyzing surface etching processes. Plasma consists of neutrals, ions, and electrons, and the effects of each species on etch results are not easily identified. Much of the fundamental understanding of the surface processes that occur during plasma etching comes from well-controlled plasma experiments in which beams of ions, electrons, and neutrals are directed, either together or alternately, at well-characterized surfaces, using apparatus such as that shown in Fig. 6.[8] Etch products are usually analyzed by mass spectrometry. The surface compositions are diagnosed by in-situ or ex-situ x-ray photoelectron spectroscopy, Auger spectroscopy, and secondary ion mass spectrometry. Without radiation or any external energy, the probability of a reaction of Si with F is approximately 10^{-2}, which could be enhanced by ion-beam bombardment. Figure 7 shows the etch rate of silicon with inert Ar ions and with reactive flux to the silicon surface. This etch rate is larger than the sum of the individual etch rates for ion sputtering and neutral etching because chemical etching is enhanced by ion bombardment.[9] Without experiments like this, it is difficult to identify the dominant etch mechanism.

The precise anisotropic etch mechanisms for any given material system are not fully understood, but it is clear that the ions striking the surface must themselves be anisotropic, that is, traveling mainly in a direction perpendicular to the wafer surface. Ions are oriented and accelerated in a sheath. Two proposed mechanisms for anisotropic plasma etching are as follows:

1. Perpendicular ion bombardment creates a damaged surface that is then more reactive toward neutral etchants.
2. Ions help to desorb etch-inhibiting species, such as etch products, from the surface.

In both mechanisms, ion transport must be perpendicular to the surface so that only the etch rate of the bottom surface is enhanced. An illustration of these anisotropic mechanisms is shown in Fig. 8.[4]

Ion–neutral particle collisions from the plasma sheath result in a fraction of the ions hitting the sidewall, and some lateral etching may occur. The number of ion–neutral collisions in the sheath is directly proportional to the sheath thickness and inversely proportional to the ion mean free path. Because the ion mean free path is

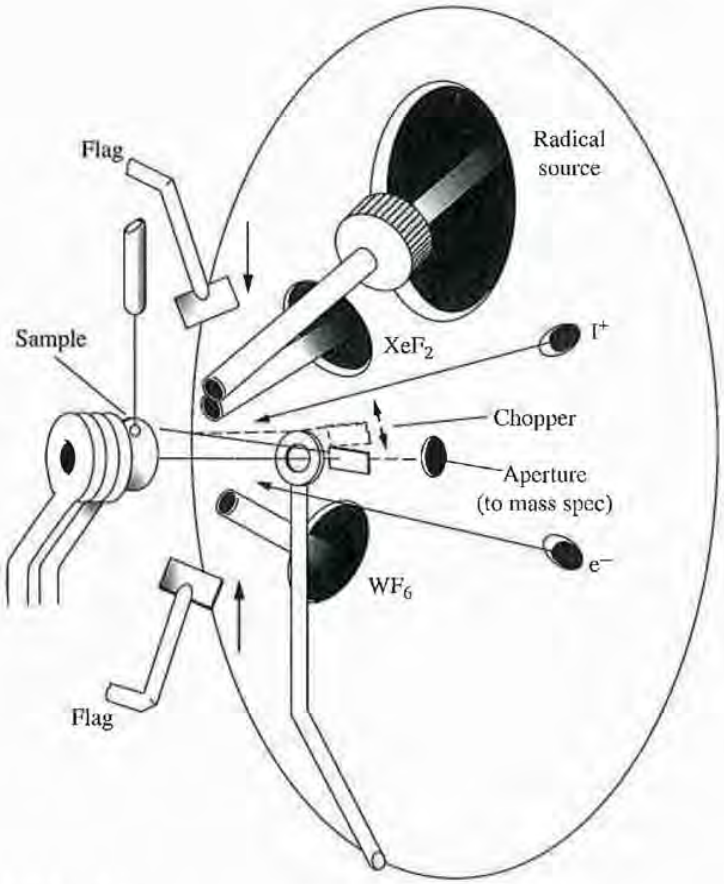


FIGURE 6
Beam-surface experimental apparatus used for the simulation of plasma etching. Such experiments are important for understanding mechanisms of surface-plasma interactions. (*After Winters and Plumb, Ref. 8.*)

usually proportional to pressure, lowering the pressure reduces ion–neutral collisions and enhances anisotropic etching.

In mechanism (2), the formation of sidewall films during etching plays a critical role in the development of the anisotropic etch profile; the films protect the sidewalls from etching by nonperpendicular incoming ions. In other cases, such as during silicon trench etching, thick deposits, often in the upper area of the trench, lead to progressive narrowing of the trench opening and a round bottom.[10] Various mechanisms contribute to sidewall film growth. Three cases are important in the formation of sidewall films:[11,12]

1. Redeposition of etch products or sputtered material from the bottom of the etched structure by adsorption and recombination
2. Deposition from glow discharge
3. Redeposition of sputtered mask material

A schematic description of sidewall deposition phenomena during silicon trench etching by HBr is shown in Fig. 9. Sidewall passivation is governed by the

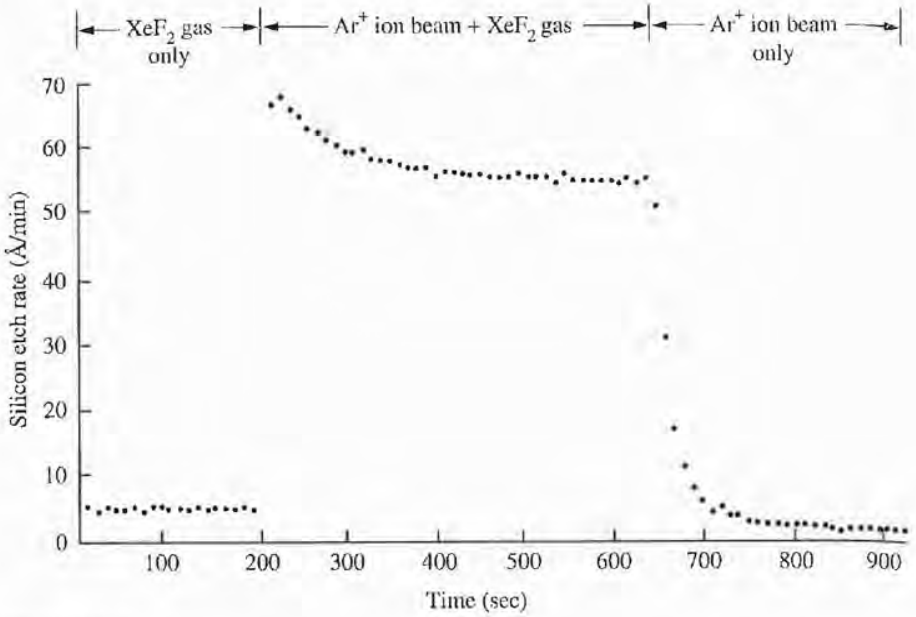


FIGURE 7
Etch rate of silicon as a function of time. This graph illustrates ion-enhanced chemical etching of silicon by an Ar^+ ion beam and XeF_2 flux. (*After Coburn and Winters, Ref. 9.*)

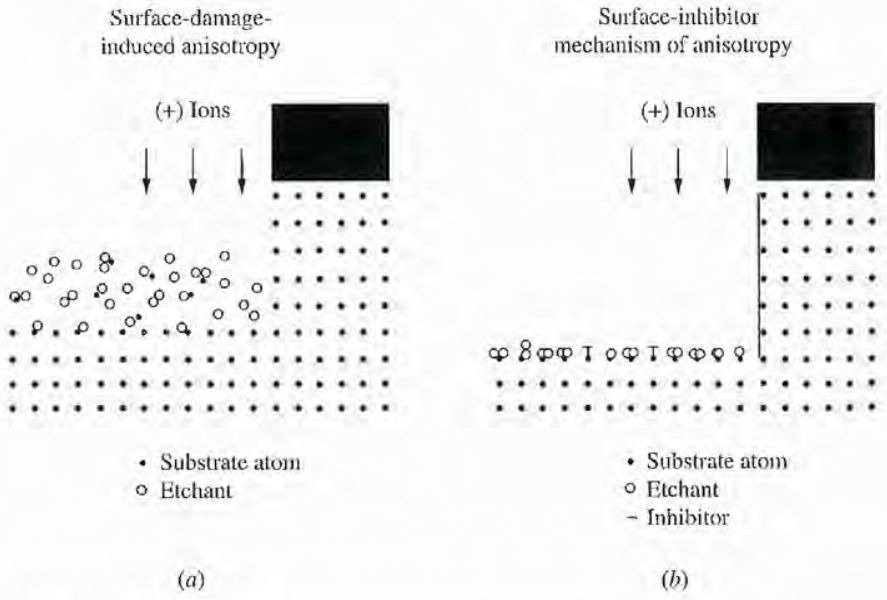


FIGURE 8
Schematic of two proposed mechanisms for ion-enhanced surface etching. Ion-enhanced surface etching results in anisotropic etch profiles. (*After Flamm and Donnelly, Ref. 4.*)

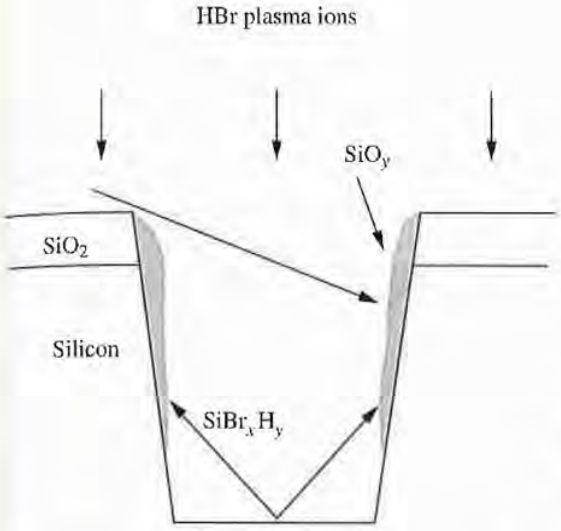


FIGURE 9
Schematic of sidewall formation mechanisms during silicon trench etching by HBr plasma.

equilibrium between deposition and ionic sputtering on vertical surfaces. Sidewall film thickness control becomes critical if tapered sidewalls and dimensional loss is not desired. For example, if the passivation layer becomes too thick during polysilicon gate etching, the gate dimensions increase and problems result.

7.3.4 Selectivity

Selectivity is defined as one film etching faster than another film under the same etching conditions. A higher etch rate ratio (ERR) between different layers is the crucial advantage of reactive ion etching over physical sputtering. The etch rate difference between two different materials is due to different surface etch mechanisms, such as adsorption, reaction, and desorption. During etching, the selectivities with respect to the masking material and the underlying layer require careful control. The required selectivity is defined according to a specified percentage of overetch and film thickness. The gate oxide thickness is in the range of 10 nm for today's 0.8-micrometer, complementary metal-oxide-silicon (CMOS) technologies, and an etch selectivity of polysilicon to thin gate oxide of more than 10 : 1 is required. In the case of a contact hole etch, a large amount of overetch time is required because of the different thicknesses of the oxide layers over the polysilicon gate and over the source and drain. Therefore, a high etch selectivity between silicon and oxide is required for oxide contact hole etch. Thinner masks, such as photoresist, are preferred for optical lithography. A thinner mask film can be used in the etching if a high ERR between the etched film and the mask is obtainable.

Selectivity usually results from two main mechanisms. First, the gas chemistry chosen should be able to generate thermodynamically favorable reactions with the primary film. For example, chlorine-based plasma is usually employed for a polysilicon–oxide etch, and the selectivity arises from the fact that Cl atoms etch SiO_2 very slowly. Etching SiO_2 with chlorine is not a thermodynamically favored reaction. Second, gas chemistry is selected to produce a volatile etch product with

the primary film. For example, to etch SiO_2 over aluminum, a fluorine-based chemistry is used because of the high vapor pressure of SiF_4 from the SiO_2 etching. A similar principle has also been applied to SiO_2-over-silicon etching, and the selectivity is usually achieved by forming a passivation layer on silicon. Fluorocarbon polymer forms are used to etch oxide selectively with respect to silicon. Carbon reacts with oxygen during oxide etch, and a fluorocarbon polymer film forms on the silicon surface and provides a good selectivity.

7.3.5 Microscopic Uniformity

Some problems in achieving microscopic uniformity occur because etching rates and profiles depend on feature size and pattern density. Microscopic uniformity problems can be grouped into two categories—aspect ratio–dependent etching (ARDE) and pattern-dependent etching, known as microloading.[13] The cause of the problems is ion and neutral transport.

A typical trench is shown in Fig. 10 with an aspect ratio of 8. Significant ARDE is seen, and trenches with a large aspect ratio etch more slowly than trenches with a small aspect ratio,[14] as shown in Fig. 11. This phenomenon became serious when the

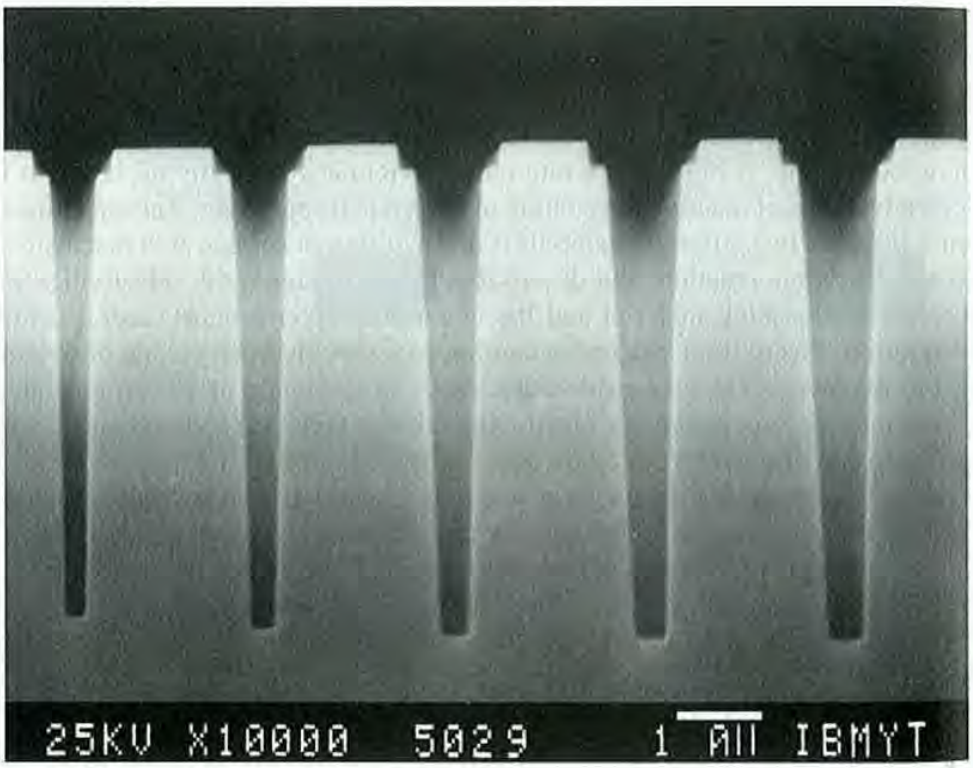


FIGURE 10
Effect of aspect ratio–dependent etching on a silicon trench etch. The etching rate decreases as the trench dimensions are reduced.

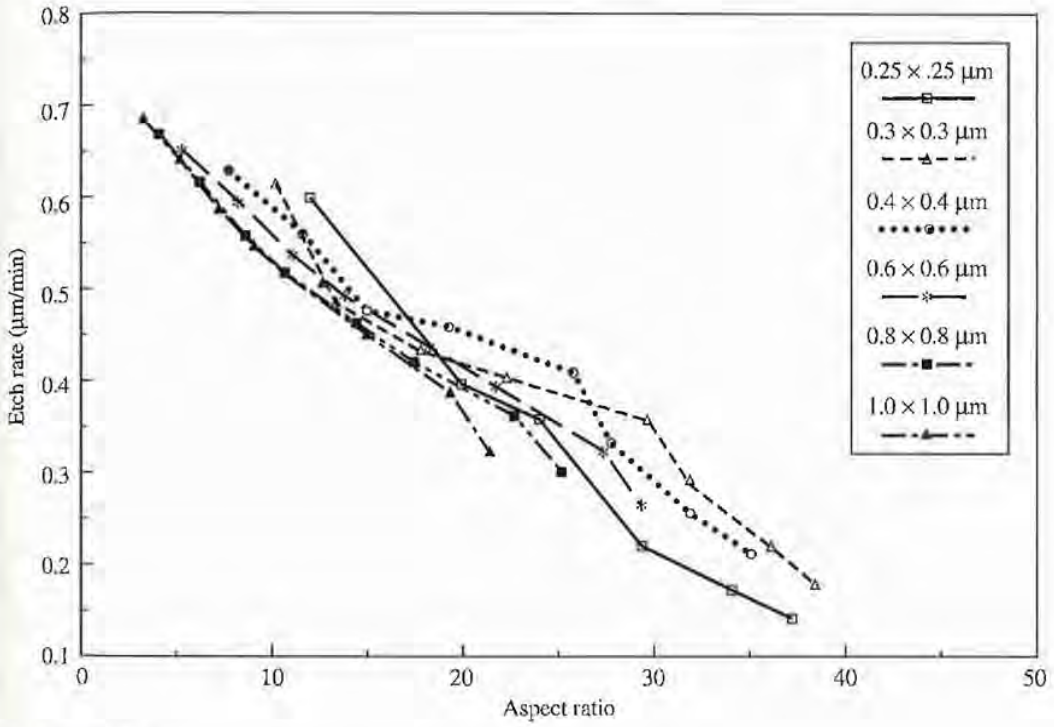


FIGURE 11
Dependence of average silicon trench etch rate on aspect ratio.

era of submicrometer etching began in recent years. Ion bombardment, electron bombardment, reactive neutral species, product desorption, and redeposition all appear to be important in determining the relative etch rates of trenches. Four different mechanisms are the main contributors to ARDE: limited Knudsen transport of neutrals, ion shadowing from ion–neutral collisions in the plasma sheath, neutral shadowing caused by geometrical effects, and ion scattering from insulator charging.[15,16,17]

The term *microloading* usually refers to the dependence of the etching rate on pattern density for identical features.[18] Microloading results from depletion of reactants because the wafer has a local, higher-density unmasked area.[14]

7.4 REACTIVE PLASMA ETCHING TECHNIQUES AND EQUIPMENT

Plasma reactor technology in the electronics industry has changed dramatically since the first application of plasma processing to photoresist stripping. A reactor for plasma etching contains a vacuum chamber, pump system, power supply generators, pressure sensors, gas flow control units, and an end-point detector. Plasma etch tools may be categorized according to the etch mechanism each of them employs. Table 1 shows the similarities and differences in the types of etch equipment that are commercially available. A comparison of pressure operating ranges and ion energies

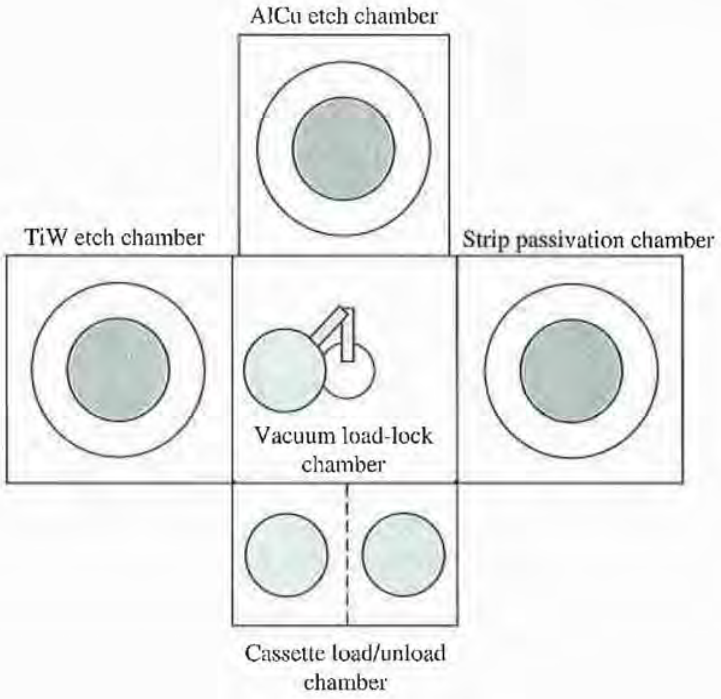


FIGURE 21
Cluster reactive ion etch tool for multilayer metal (TiW/AlCu/TiW) interconnect etching.

7.5 PLASMA PROCESSING PROCESSES

The chemical and physical complexity of plasma-surface interactions makes computer-based modeling and plasma simulation inadequate for developing plasma reactors. As a result, the detailed descriptions required to guide the transfer of processes from one reactor to another or to scale processes from a small to a large reactor are not available. Therefore, the plasma processes for fabricating microelectronic devices have been developed largely by time-consuming, costly, empirical methods.

Excellent critical dimension control and minimized profile microloading are the two most important issues for ongoing research. Extensive research in plasma etching explores ways to optimize the performance of etching processes, characterize and avoid contamination and damage effects, identify the important reactive species and measure their densities, and, in general, improve our limited understanding of plasma etching. See Refs. 28 and 29 on various aspects of plasma etching. Other studies concentrate on the limitations of conventional rf plasma. This research led to the development of enhanced plasma source configurations and processes that are still being examined.

Etch chemistries also play a critical role in the performance of etch processes. Table 3 lists some conventional and new etch chemistries for different etch processes.[30] A truly useful plasma process must have a robust process window.

TABLE 3
Etch chemistries of different etch processes[30]

Material being etched	Conventional chemistry	New chemistry	Benefits
PolySi	Cl_2 or BCl_3/CCl_4 / CF_4 / $CHCl_3$ / CHF_3 } sidewall passivating gases	$SiCl_4/Cl_2$	No carbon contamination
		BCl_3/Cl_2	
		$HBr/Cl_2/O_2$	Increased selectivity to SiO_2 and resist
		HBr/O_2	No carbon contamination
		Br_2/SF_6	
		SF_6	Higher etch rate
		CF_4	
Al	Cl_2	$SiCl_4/Cl_2$	Better profile control
	BCl_3 + sidewall-passivating gases	BCl_3/Cl_2	No carbon contamination
	$SiCl_4$	HBr/Cl_2	
Al-Si (1%); Cu (0.5%)	Same as Al	$BCl_3/Cl_2 + N_2$	N_2 accelerates Cu etch rate
Al-Cu (2%)	$BCl_3/Cl_2/CHF_3$	$BCl_3/Cl_2 + N_2 + Al$	Additional aluminum helps etch copper
W	$SF_6/Cl_2/CCl_4$	SF_6 only	No carbon contamination
		NF_3/Cl_2	Etch stop over TiW and TiN
			No carbon contamination
TiW	$SF_6/Cl_2/O_2$	SF_6 only	
WSi_2, $TiSi_2$, $CoSi_2$	CCl_2F_2	CCl_2F_2/NF_3	Controlled etch profile
		CF_4/Cl_2	No carbon contamination
Single-crystal Si	Cl_2 or BCl_3 + sidewall-passivating gases	CF_3Br	Higher-selectivity trench etch
		HBr/NF_3	
SiO_2 (BPSG)	CCl_2F_2	CCl_2F_2	
	CF_4	CHF_3/CF_4	CFC alternatives
	C_2F_6	CHF_3/O_2	
	C_3F_8	CH_3CHF_2	
Si_3N_4	CCl_2F_2	CF_4/O_2	CFC alternatives
	CHF_3	CF_4/H_2	
		CHF_3	
		CH_3CHF_2	

Developing an etch process usually means optimizing many characteristics, such as etch rate, selectivity, profile control, endpoint control, damage, etc., by adjusting a large number of process parameters. Because we do not fully understand etch processes, systematic experiments are required to study the large number of process parameters. The most common method is to use the design experiment and analyze the results with a surface response methodology. In this section, several of the most important etch processes for ULSI fabrication are introduced.

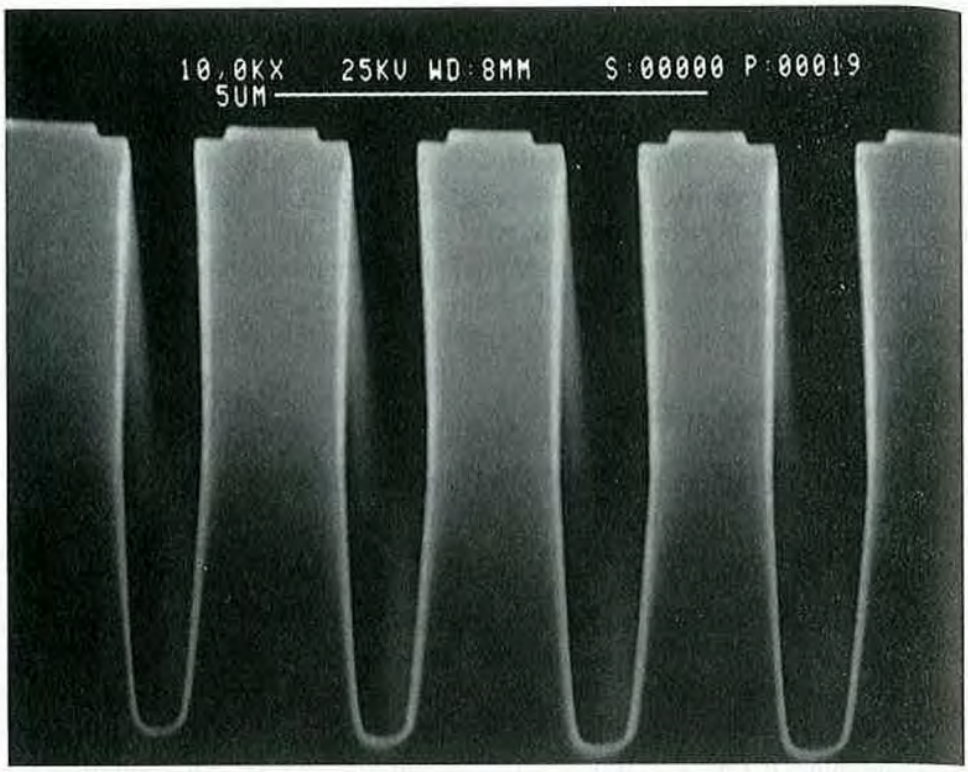
7.5.1 Silicon Trench Etching

As feature sizes in ULSI decrease, a corresponding decrease is needed in the area occupied by both the isolation between circuit elements and the storage capacitor of the DRAM memory cell. This area can be reduced by etching trenches into the silicon substrate and filling them with suitable dielectric or conductive materials. Silicon trenches can also be used as alignment marks for optical or electron beam lithography. Deep silicon trenches, usually with a trench depth larger than 5 μm, are used mainly for forming trench capacitors. Shallow silicon trenches, usually with a trench depth less than 1 μm, are often used for isolation between device elements.

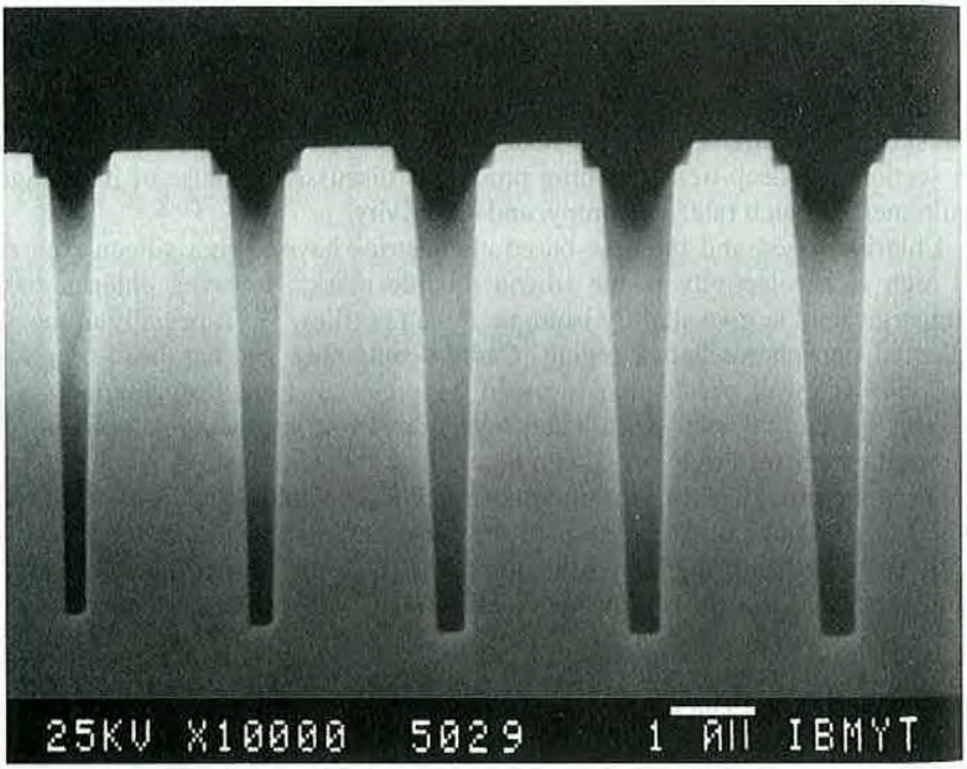
To guarantee a void-free trench refill, anisotropic etching with a high etch selectivity to mask material is desired for submicrometer-deep silicon trench etching. This precludes the use of fluorocarbon chemistry, which tends to undercut the mask material.[18] Fluorine chemistry has a high silicon etch rate, but it has low etch selectivity to the oxide mask and to the isotropic trench profile. However, it is often used for etching shallow silicon trenches with trench depths less than 1 μm, because it satisfies the requirement of etch selectivity of silicon to the photoresist mask. In this section, the deep-trench etching process is discussed because of its stringent requirements of etch rate, anisotropy, and selectivity.

Chlorine-based and bromine-based chemistries have a high silicon etch rate and high etch selectivity to the silicon dioxide mask. However, chlorine-based chemistries tend to give slightly isotropic trench profiles,[31,32] especially in the ion-implanted phosphorus-doped region. Carbon-containing gas has been employed to provide sidewall passivation to avoid undercutting from fast chlorine-silicon reactions.[33] The sidewall passivation layer, however, causes contamination and reduces the etch selectivity of silicon to a silicon dioxide mask. A simple carbon-free etch chemistry leading to minimum thickness of the sidewall passivation is desired to guarantee high accuracy in the pattern transfer. As a general rule, the reactivity of the halogens follows the order F > Cl > Br. Recently, bromine-based chemistries have become popular for etching silicon with anisotropy and extremely high selectivity to the oxide mask.[34]

Sidewall passivation protects the sidewall from lateral etch attack and local ion-enhanced trenching. The desired trench shapes with smooth vertical sidewalls are obtained by controlling the balance between the etch and the inhibition by sidewall passivation. The sidewall shape depends significantly on the wafer temperature. Increasing temperature results in less deposition of sidewall passivation and more lateral etching. Fig. 22[10] refers to the effects of backside cooling under different He pressure. Greater pressure of He results in a cooler wafer temperature.



(a)



(b)

FIGURE 22
Effect of wafer backside cooling on a silicon trench profile: (*a*) backside He pressure 0.5 torr (*b*) backside He pressure 6.0 torr. (*After Lii, Ng, and Danner, Ref. 10.*)

Aspect ratio–dependent etching (ARDE) is often observed in submicrometer-deep silicon trench etching caused by limited ion and neutral transport within the trench. Trenches with large aspect ratios etch more slowly than trenches with small aspect ratios. Ion scattering results from ion–neutral collisions in the plasma sheath, and the electrical charging on the masks causes ARDE. Some neutrals are transported to the bottom of the trench by Knudsen diffusion, also contributing to ARDE. Low gas pressure reduces the ARDE effect, and chlorine-based chemistry shows less ARDE than fluorine-based chemistry during deep trench etching, because ion-assisted etching is dominant in chlorine-based chemistry.[35]

7.5.2 Polysilicon and Polycide Gate Etching

Low-pressure chemical vapor deposited (LPCVD) polysilicon is usually used as a gate material for MOS devices because of its superior interface property with thin gate oxide at high temperature. Metal silicide over polysilicon has also been used for the MOS gate because of its high electrical conductivity. Anisotropic etching and high etch selectivity between polysilicon and the gate oxide are the most important requirements for gate etching. Achieving high selectivity and etch anisotropy at the same time is difficult for most ion-enhanced etching processes. Therefore, multistep processing is used in which different etch steps in the process are optimized for etch anisotropy and selectivity.

Most chlorine-based and bromine-based chemistries can be used for gate etching to achieve the required etch anisotropy and selectivity. The main etchant today is bromine-based chemistry, because bromine gives higher etch selectivity of polysilicon to gate oxide than chlorine does. Figure 23 shows a scanning electron microscope (SEM) cross section of a polysilicon gate etched by HBr plasma[36] that has a 0.1 μm gate width and a 45 Å gate oxide thickness. The thin native oxide on top of the polysilicon should be removed by fluorine-based plasma before polysilicon etching, i.e., initiation, to avoid any micromasking formation during polysilicon etching. To further improve the selectivity of polysilicon to gate oxide requires removing carbon-containing gases and carbon-containing materials from the etch process and the etch reactor, because carbon tends to reduce the selectivity of polysilicon to gate oxide,[37] as shown in Fig. 24. It is suspected that carbon reacts with the oxygen atoms of silicon dioxide. An enhanced etch attack at the edge of the gate, a phenomenon called "trenching," is shown in Fig. 25. Trenching results in broken gate oxide and is usually observed at low gas pressure. The trenching problem can be solved by increasing etch selectivity, such as adding bromine- and oxygen-containing gases to the etch process.

7.5.3 Oxide and Nitride Etching

An oxide etching process is often used to open a contact window down to a silicon device or metal conductor; therefore, the selectivity of SiO_2 to Si must be high. It is well known that selectivity is achieved by increasing the ratio of CF_x to F atoms

FIGURE 23
SEM cross section of HBr-plasma-etched polysilicon gate with a 0.1 μm gate width and a 45 Å gate oxide thickness. (*After Lii et al., Ref. 36.*)

in the plasma.[7] CHF_3/CF_4 chemistry is a common oxide etchant because it provides high selectivity to Si from carbon polymer deposition on the silicon surface. Therefore, a further silicon surface cleaning must be added to remove the polymer passivation layer after oxide etching to avoid high contact resistance between metal and silicon.

LPCVD or plasma-enhanced CVD silicon nitride are often used as a sidewall around the polysilicon gate in MOS devices for source and drain extension or used as a final passivation layer. Sidewall formation for a polysilicon gate requires anisotropic etching and high selectivity of silicon nitride to silicon dioxide. Si_3N_4 can be etched by both fluorine atoms and CF_x-containing plasma. CF_4 and CF_4/O_2 are widely used to etch silicon nitride but are not selective to silicon or silicon dioxide.[38] Si_3N_4 can be etched anisotropically in an RIE mode with high selectivity to Si and SiO_2 by CHF_3/O_2, CH_2F_2, or CH_3F.[39,40] Details of why the etch mechanism results in high selectivity are not clear. Fluorine-deficient fluorocarbon polymer is thought to contribute to this high selectivity.

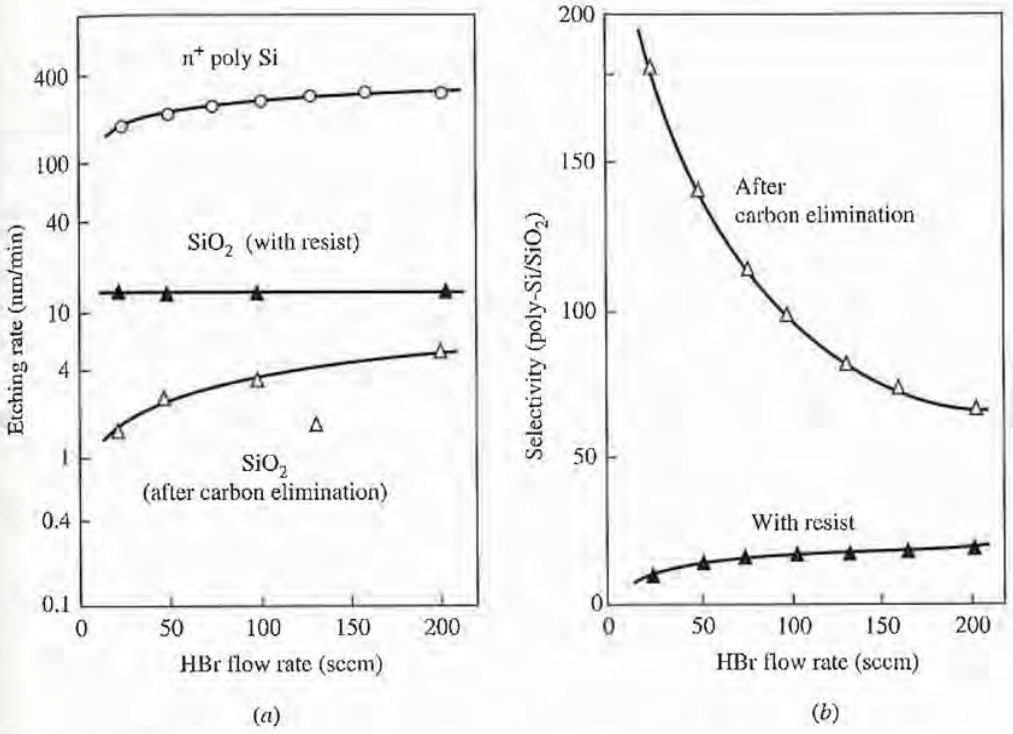


FIGURE 24
Dependence of (*a*) the polysilicon etch rate and (*b*) Si/SiO_2 selectivity on HBr flow rate, carbon reducing Si/SiO_2 selectivity. (*After Nakamura, Iizuka, and Yano, Ref. 37.*)

7.5.4 Aluminum and Tungsten Metal Etching

Etching of a metallization layer is a very important step in semiconductor device fabrication. Aluminum and tungsten are the most popular materials used for interconnection, and anisotropic etching of these materials is usually required. An aluminum etch with fluorine-based chemistry does not work, because of the very low vapor pressure of the etch product AlF_3. Chlorine-based chemistry has been used for Al etching most frequently, and bromine-based chemistry has been investigated recently.[41]

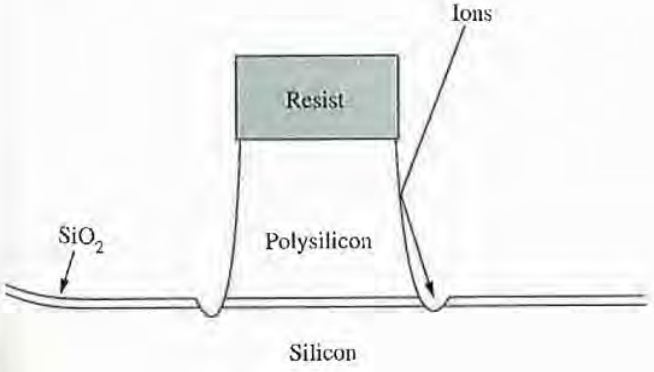


FIGURE 25
Schematic of the trenching mechanism during polysilicon gate etching.

Cl_2 spontaneously etches clean aluminum. Chlorine has a very high chemical etch rate with aluminum and tends to produce an undercut during aluminum etching. Carbon-containing gas is added to form sidewall passivation during aluminum etching to obtain etch directionality. The Al_2O_3 layer, which is always present on air-exposed aluminum, is etched very slowly in chlorine-based gases. Energetic ion bombardment, for example, sputtering, is required to remove Al_2O_3 before aluminum etching. In the semiconductor industry, BCl_3 plasma is often used to remove Al_2O_3 on the aluminum surface.[42] BCl_x also appears to recombine with chlorine at the wafer surface to ensure an anisotropic etch profile during Al etching with a Cl_2/BCl_3 mixture. A small percentage of silicon and copper are usually added to enhance the electromigration resistance of aluminum conductors, but Cu causes micromasking during aluminum etching with chlorine-based plasma because it generates low-volatility etch products. Micromasking can be suppressed by using stronger ion bombardment during etching and addition of aluminum chloride gas.[43] More recently, a barrier metal layer, such as TiW and TiN, has been used to prevent Si migration into Al. TiW and TiN etch readily in chlorine-based plasma.

Corrosion of the aluminum pattern exposure to the ambient is another problem in aluminum etching. Residual chlorine on both the Al sidewall and the photoresist tends to react with water to form HCl, which etches aluminum. In-situ exposure of the wafer to a CF_4 discharge to exchange Cl with F and then to an oxygen discharge to remove the resist, followed by immediate immersion in deionized (DI) water, is successful in eliminating Al corrosion.[44] High-temperature heating of the wafer to evaporate chlorine or hydrolyzing chlorine with hydrogen-containing gases can be used to remove chlorine trapped in the resist and the sidewall.

LPCVD tungsten has been widely used for contact hole filling and first-level metalization because of its excellent deposition conformability. Both fluorine- and chlorine-based chemistry etch W and form volatile etch products.[45,46] Fluorine-based chemistry also etches the silicon dioxide layer; thus it is very difficult to obtain selectivity between W and SiO_2. Chlorine-based chemistries etch tungsten with selectivity to SiO_2. Two tungsten etch processes, blanket W etchback to form a W plug and W pattern etch, are discussed briefly in this section.

To form a W plug, blanket LPCVD W is deposited on top of a TiN barrier layer as shown in Fig. 26. Reducing loading effects during the etchback of W is very important to avoid a W recess in the plug. A two-step process is used for W etchback. In the first step, 90% of the W is etched at a high etch rate and excellent uniformity to achieve high throughput. In the second step, the etch rate is reduced to remove the

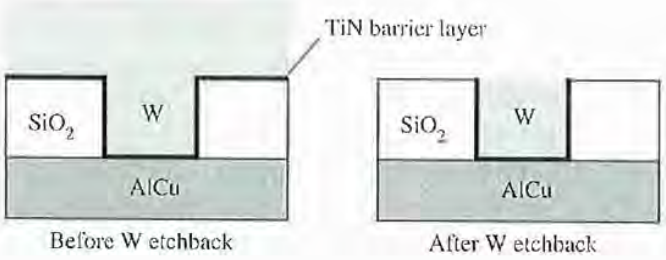


FIGURE 26
Formation of tungsten plug in a contact hole by depositing blanket LPCVD W and then using RIE etchback.

remaining W with an etchant that has a high W-to-TiN-barrier-layer etch selectivity and a minimum loading effect.[47] The loading effect is reduced by decreasing the gas pressure and wafer temperature because of the low chemical etch rate of F with W at these process conditions.

For W interconnect etching, the selectivity between W and the photoresist mask is very low for both fluorine- and chlorine-based chemistries. N_2 is added to the SF_6/Cl_2 mixture to get a slight selectivity increase.[48] Inorganic material, such as SiO_2, can be used as a hard mask during W etching with chlorine-based chemistry because of high etch selectivity of W to the hard mask.

7.5.5 Organic Materials Etching

Oxygen plasmas have been used with a barrel reactor for resist etching since the 1970s. Oxygen atoms are the primary species responsible for etching polymers at a pressure of 1–3 torr. Oxygen atoms initiate polymer etching by extracting hydrogen atoms from a polymer chain to form radicals. These radicals then react with oxygen molecules to form volatile products under high wafer-temperature conditions or under ion bombardment.

Recently, multilayer lithography has become increasingly important for ULSI processing because it provides high resolution that is independent of the underlying materials and substrate topography. The trilayer scheme is a very popular multilayer lithography method. It uses a conventional resist to pattern an intermediate layer, such as silicon dioxide or nitride, that acts as a mask during the subsequent planarization of the organic layer by oxygen RIE. A high degree of anisotropic and selective etching is required in these applications. This implies that both chemical and physical processes are needed, because purely physical etching is not selective, whereas purely chemical etching is not anisotropic. Reactive ion etching contributes both chemical and physical processes. To achieve anisotropic etching, sidewall passivation is required to suppress chemical and physical etching on the sidewalls of the feature.[49] A low wafer temperature and extra carbon- or chlorine-containing gases increase the deposition rate of sidewall films and improve etch anisotropy.[50]

7.5.6 Planarization Etch

With the advent of three or more levels of interconnection, more than 50% of the cost of chip manufacturing will be in the fabrication of electrical interconnects. Between these interconnection levels are inorganic insulators or polymer films, such as doped silicon dioxide glass or polyimide. Because the lithographic exposure achieves maximum resolution over only a limited depth of focus, these films must be flat, or planarized, over the optical exposure area to construct higher levels of interconnection. If the temperature budget does not allow planarization of the oxide layer by means of heat flow, planarization is sometimes accomplished by a repetitive sequence of plasma etching and deposition. An organic material with good flow properties is

deposited on top of the oxide layer, and the wafer is etched in a plasma that removes the organic material and the oxide at an approximately equal rate, leaving a uniform oxide layer.

Mixtures of fluorine-based chemistry, such as NF_3, CF_4, and CHF_3, with oxygen are suitable for a nonselective silicon dioxide/organic material etch. Good etch uniformity and a high etch rate are the most important requirements for a planarization etch. The planarization etching step also reduces surface topography because different etch rates are used at different surface angles.

7.6 DIAGNOSTICS, END POINT CONTROL, AND DAMAGE

Dry etching differs from traditional wet chemical etching, and it usually does not have enough etch selectivity to the layer underneath. The plasma reactor must be equipped with a monitor that indicates when the etching process is to be stopped; an end point detection system is needed to avoid etching the underlying layer. Several different types of end-point systems are available. Most of them detect the change in the amount of active reactants or etch products in the discharge. Two of the most popular methods are introduced here. Some etching systems use a combination of these methods for end point detection.

7.6.1 Optical Emission Spectroscopy

Optical emission spectroscopy is a method of analyzing the light emitted by excited atoms and molecules in the gas discharge. The concentration of excited residual products or excited reactants during the etching process is studied by equipping a detector with a filter that lets light of a specific wavelength pass through to be detected, as shown in Fig. 27. Photons passing through a tunable monochromator are detected by a photomultiplier. The presence of neutrals and ions can be determined by correlating these spectra with previously determined spectral series. However, absolute concentrations cannot be obtained unless a complex actinometry technique is used. The emission signal derived from the primary etchant or by-product begins to rise or fall at the end of the etch cycle.

7.6.2 Laser Interferometry

Laser interferometry uses laser light directed toward the wafer surface. If the films on the wafer are transparent, the laser beams reflected from the top and bottom of the etched layer will interfere with each other. As the thickness of the etched layer is reduced, the degree of interference will change, either constructively or destructively. The time between light maxima and minima can be used to determine the etching rate. At the end of the etching process, the interference stops and the signal flattens out. The period of the oscillation is related to the change in film thickness, Δd where