Limits of Lithography

LLOYD R. HARRIOTT

Invited Paper

Lithography technology has been one of the key enablers and drivers for the semiconductor industry for the past several decades. Improvements in lithography are responsible for roughly half of the improvement in cost per function in integrated circuit (IC) technology. The underlying reason for the driving force in semiconductor technology has been the ability to keep the cost for printing a silicon wafer roughly constant while dramatically increasing the number of transistors that can be printed per chip. ICs have always been printed optically with improvements in lens and imaging material technology along with decreases in wavelength used fueling the steady improvement of lithography technology. The end of optical lithography technology has been predicted by many and for many years. Many technologies have been proposed and developed to improve on the performance of optical lithography, but so far none has succeeded. This has been true largely because it has always been more economical to push incremental improvements in the existing optical technology rather than displace it with a new one. At some point in time, the costs for pushing optical lithography technology beyond previously conceived limits may exceed the cost of introducing new technologies. In this paper, I examine the limits of lithography and possible future technologies from both a technical and economic point of view.

Keywords*—*Electron beams, IC manufacturing, lithography, nanotechnology, photolithography, printing.

I. Introduction

Lithography has been one of the key drivers for the semiconductor industry. Moore's Law states that the number of devices on a chip doubles every 18 months. There are three main constituents of the technology improvements that have kept the industry on this pace for more than 30 years. They are lithography, increased wafer size, and design. Roughly half of the density improvements have been derived from improvements in lithography. With the cost to fabricate a wafer remaining roughly constant, independent of size or content, this has resulted in a 30% reduction in cost per function per year over this period [1].

This very powerful economic engine has driven and been driven by advances in lithography. From a purely technical point of view, it is possible to pattern features to atomic dimensions with advanced research techniques. However, these nanolithography methods tend to be very slow and would not meet the requirements for the economic model described above. In this paper, the limits of lithography in the context of printing integrated circuits (ICs) will be discussed. Current lithographic techniques as well as their evolution will be described along with possible future alternative technologies. In this context, the ultimate limits of lithography will be determined by the interplay of science, technology, and economics.

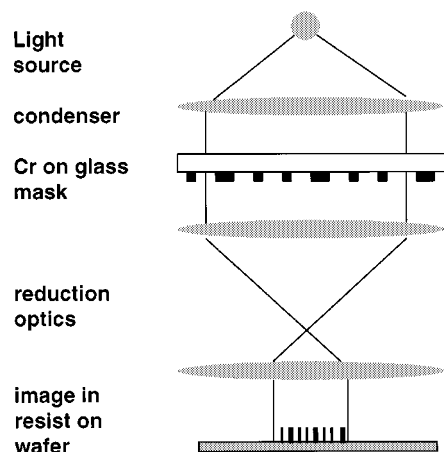

Fig. 1. Schematic diagram of optical step-and-repeat lithography tool. Pattern on the mask usually represents one level of one (or sometimes a few) chip. Mask pattern is printed onto the wafer, then the wafer is moved to a new location, and the process is repeated until the entire wafer is exposed. Actual projection optical systems have 25 or more lens elements for aberration compensation and field flattening to achieve diffraction-limited performance.

Leading-edge production lithography employs optical projection printing operating at the conventional Rayleigh diffraction limit. The image of the master pattern or mask (usually reduced by four or five times) is projected onto the wafer substrate that has been coated with a photosensitive material (resist). A schematic diagram of an optical lithography system is shown in Fig. 1. The solubility of the resist is changed by exposure to light so that a pattern emerges upon development (much like a photograph). The remaining

Manuscript received February 5, 2000; revised October 4, 2000.

The author is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA.

Publisher Item Identifier S 0018-9219(01)02071-0.

0018–9219/01$10.00 © 2001 IEEE

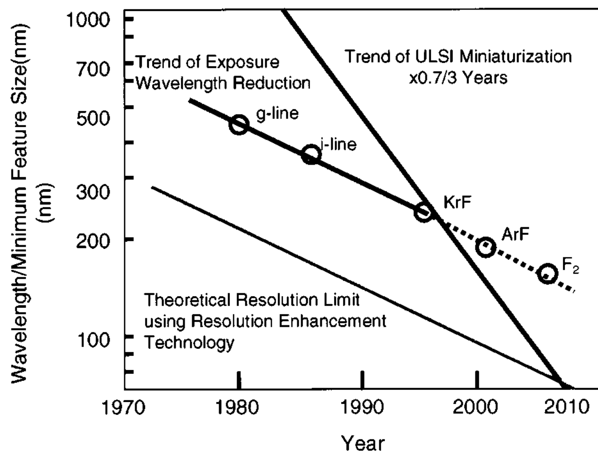

Fig. 2. Comparison of lithography wavelength trends with IC feature size trend. Courtesy of Dr. S. Okazaki, Hitachi Ltd.

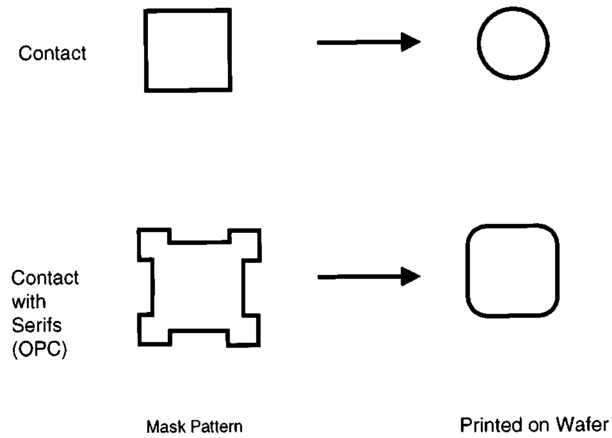

Fig. 3. Example of using OPC serif features in contact hole printing. (top) Square feature on the mask prints as a circle at the wafer due to diffraction effects. (bottom) Serifs are added to make the corners of the printed image more square.

resist pattern is then used for subsequent process steps such as etching or implantation doping. The optical projection systems used today have very complex multielement lenses that correct for virtually all of the common aberrations and operate at the diffraction limit. The resolution of a lithography system is usually expressed in terms of its wavelength and numerical aperture (NA) as

$$\text{Resolution} = k_1 \frac{\lambda}{\text{NA}} \tag{1}$$

where the constant k_1 is dependent on the process being used. In IC manufacturing, typical values of k_1 range from 0.5 to 0.8, with a higher number reflecting a less stringent process. The NA of optical lithography tools ranges from about 0.5 to 0.6 today. Thus, the typical rule of thumb is that the smallest features that can be printed are about equal to the wavelength of the light used. Historically, the improvements in IC lithography resolution have been driven by decreases in the printing wavelength, as shown in Fig. 2. The illumination sources were initially based on mercury arc lamps filtered for different spectral lines. The figure shows the progression from G-line at 435 nm to I-line at 365 nm. This was followed by a switch to excimer laser sources with KrF at 248 nm and, more recently, ArF at 193 nm. The most advanced IC manufacturing currently uses KrF technology with introduction of ArF tools beginning sometime in 2001. It can also be seen from the figure that the progress in IC minimum feature size is on a much steeper slope than that of lithography wavelength. Prior to the introduction of KrF lithography, the minimum feature sizes printed in practice have been larger than the wavelength with the crossover at the 250-nm generation and KrF. With the introduction of 180-nm technology in 1999, the most advanced IC manufacturing was done with feature sizes significantly below the wavelength (248 nm).

The ability to print features significantly less than the wavelength of the exposure radiation can largely be attributed to improvements in the imaging resist materials. Modern resists exhibit very high imaging contrast and act as a thresholding function on the aerial image produced by the optical system. In other words, even though the light intensity image has less than full modulation for the small features, the combination of high-contrast imaging material and good process (exposure dose) control can reliably produce subwavelength features. In this way, the improvements in imaging resists have lowered the value for k_1. It is interesting to note that while this is true for KrF lithography at 248 nm, it is not yet true for ArF exposures at 193 nm. That is, the resist materials are not yet developed to the point of producing superior images even though the wavelength is smaller. Currently, the best lithographic performance is seen at 248 nm. This also implies that unless resist materials for 193 nm or shorter wavelengths such as 157 nm (F_2 excimer) can be developed to a performance point equal to or better than that for 248-nm materials, continued feature size shrinkage through wavelength reduction is not feasible.

Some compensation for the image degradation from diffraction are possible by predistorting the mask features. A simple example is a correction for corner rounding by using serifs. The addition of subresolution features does enhance the quality of the image on the wafer somewhat, but requires the addition of these correction features on the mask, increasing its complexity and cost. An example of using serifs to improve the printing fidelity of contact hole patterns is shown in Fig. 3. This approach is referred to as optical proximity effect correction (OPC) [2].

Image size reduction is an important factor in lithography. As stated above, the image of the mask is generally reduced by a factor of four or five when it is printed on the wafer. The main reason for this is due to the mask-making process. Masks are patterned by a scanned electron or laser beam primary pattern generator. The resolution and placement accuracy of the pattern generator are the basis for that of the optical printing system. Reduction imaging relaxes the requirements on the pattern generators. Thus, the specifications for wafer lithography are generally four or five times better than those of the pattern generators. In the regime where feature sizes are printed that are less than the exposure wavelength, the process is highly nonlinear. In terms of the mask, this introduces a complication referred to as mask

error enhancement factor (MEEF) [3]. In the regime where the printed feature size is much less than the wavelength of the light used to print it (i.e., $k_1 < 0.5$), less of the light diffracted by the mask is accepted by the entrance pupil of the imaging system and contrast is lost. One of the consequences of lithography in this nonlinear imaging regime is the enhancement of the printing of the inevitable linewidth control errors on the mask. Although the overall demagnification of the image from mask to wafer may still be four or five, small errors on the mask are not suppressed by the magnification ratio. The MEEF is defined as the derivative of the critical dimension (CD) at the wafer to that at the mask (corrected for the magnification ratio)

$$\text{MEEF} = \frac{\partial \text{CD}_{\text{wafer}}}{\partial \text{CD}_{\text{mask}}}. \quad (2)$$

For example, at a k_1 factor of 0.35, for equal lines and spaces, the mask errors are enhanced by about 2.5 times. This means that the requirements for feature size control on the mask are 2.5 times more stringent than one would otherwise expect with reduction lithography performed at larger $k_1(CD \geq \lambda)$. Thus, the demands on mask making and thus costs are increasingly difficult as optical lithography is pushed below the wavelength.

II. Advanced Optical Lithography

Increased NA is, of course, another route to improved resolution in optical lithography. Improved optical designs aided by sophisticated computer modeling are enabling larger NA lenses to be designed. KrF systems with NA $= 0.7$ will be available shortly with NA $= 0.85$ systems being designed. The penalties for these very high-NA systems are primarily in cost and depth of focus (DOF). The cost of the lens and, thus, the lithography system and wafers printed by it scales roughly with the cube of the NA (volume of lens material). At these large NAs, the weight and size of the lenses also presents many practical issues. The DOF of a system can also be characterized by the wavelength and NA as

$$\text{DOF} = k_2 \frac{\lambda}{\text{NA}^2} \quad (3)$$

where k_2 is also a process-dependent parameter generally taken to have the same value as k_1. This clearly shows the penalty in DOF for high-NA systems. Reduced DOF requires extremely tight control and planarity in the wafer process. For even a modest (0.6) NA system, the DOF is only a few hundred nanometers. Usable NA will be limited to something less than the theoretical limit of NA $= 1$ by DOF considerations as well as effects caused by refraction of the high angle light and polarization effects in the resist film. A practical maximum is currently thought to be NA $= 0.85$.

Improvements in illumination methods have also been used to improve lithographic performance. The resolution limit expressed in (1) reflects the fact that the first diffracted order must be captured in the lens for the image information to be transferred by the optical system. Off-axis illumination (OAI) [4] uses a tilted illumination to capture one of the first diffracted orders while allowing the zero order also to pass. In other words, resolution performance can be doubled in principle since the aperture of the lens now must cover the zero and one of the first diffracted orders, whereas in a conventional system the aperture covers the span of the -1 to $+1$ orders. In practice, however, this method is limited by the fact that the illumination must be tailored to the mask pattern since the diffraction pattern will be different for different mask patterns. Specific illumination patterns (annular or quadrupole symmetry) are chosen for specific types of circuit patterns to emphasize the performance of specific features such as line-space gratings.

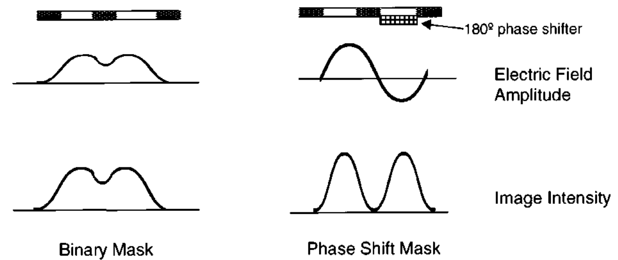

Fig. 4. Schematic comparison of diffraction limited imaging with a binary mask and with a phase-shift mask.

Other strategies to improve resolution include phase modulation in addition to the image amplitude modulation introduced by the mask pattern in the imaging system [5]. Conventional photomasks are often referred to as binary masks because they create image contrast using opaque and transparent materials. Most photomasks are made using a transparent quartz (synthetic fused silica) substrate with a patterned metal layer (usually Cr) as an absorber. The mask and, thus, image are then binary, full, or zero intensity. The optical system then creates an aerial image at the wafer plane that is a convolution of the binary intensity pattern and the diffraction point-spread function of the lens. Phase modulation or phase shifting can improve the effective resolution of the system through constructive and/or destructive interference in addition to the amplitude modulation. There are a great many implementations of phase-shift masks (PSMs), but the earliest form was the alternate-aperture PSM. Fig. 4 illustrates the principle and compares it to that of the binary mask. In the case of a pattern of lines and spaces (grating), the binary mask consists of alternate transparent and opaque areas. In the alternate-aperture PSM, every other transparent area is designed such that the light passing through it is shifted in phase by 180°. In this way, the electric field in the aerial image is forced to change sign at the feature edge and, thus, cross zero intensity. The resolution of the image in the resist is thus improved by "pulling down" the minima in the intensity pattern. The alternate-aperture PSM is limited to patterns of repeating features and is difficult to apply directly to circuit designs. It has been used for dynamic random access memory (DRAM) circuits in limited applications. The practical implementation aspects include difficulties with how to terminate the shifted and nonshifted areas at the ends of the features (sometimes called the three-color map problem). As a result, most implementations are done manually in a

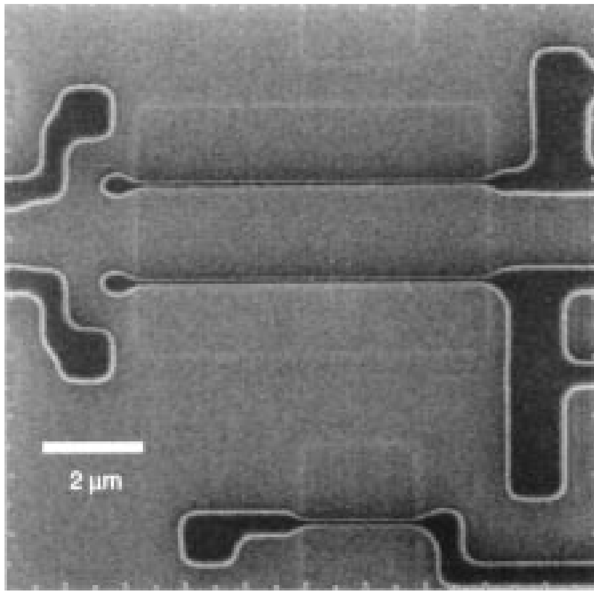

Fig. 5. Scanning electron micrograph of gate-level pattern of a DSP chip design. Gates were printed at 120 nm using 248-nm DUV lithography and the two-mask phase-shift approach.

cell-based or repetitive pattern and are not amenable to automated computer-aided design-based approaches with the exception of the two-mask approach for gate levels described below.

There are any other variations of PSM, including so-called rim shifters where 180° phase-shift areas are placed at feature edges to improve the resolution, and chromeless shifters where the pattern uses only phase modulation to produce line patterns. In general, these methods are also difficult to apply to real circuits, but illustrate the potential of this approach.

A much more practical approach to PSM uses two masks [6]. In this approach, a binary mask is used to print the gate level of a circuit with conventional design rules. Then, a second mask is used to reexpose the same resist pattern with a phase-shift pattern. This "trim" mask acts to reduce the size of the gates and can, in principle, achieve linewidths as small as half that achieved using conventional masks. The "trim" mask is created using software that compares the gate-level pattern with that of the underlying thinox layer to isolate the gate features in the pattern data. The PSM pattern is then created automatically. This method has been recently applied to the gate level of a three million transistor digital signal processor (DSP) chip yielding gate lengths of 120 nm using 248 KrF lithography. A micrograph of this is shown in Fig. 5. Currently, this method is limited to gate lithography because the automated software must have a logical basis upon which to decide which areas to phase shift.

An important aspect of the two-mask PSM approach and many other of the resolution enhancement technologies (RETs) is that while they improve the minimum feature size that can be printed, they do not improve the density of features over conventional binary mask lithography [7]. The phase-shift approaches improve the printing of isolated features such as gates, but do not improve the minimum distance (half pitch) for the patterns. In terms of IC performance, smaller gates will improve speed and power consumption of circuits; they do not increase the packing density or number of circuit elements per chip. This has led to a bifurcation of the roadmap for the IC industry over the past few years. The progress in the minimum half pitch or density of circuit elements has been driven mainly by memory applications and has continued to follow the Moore's Law trend with some amount of acceleration due to improvements in resist and NA. Meanwhile, minimum feature sizes are driven by the speed requirements of microprocessors and have been following a more aggressive trend. In general, gate sizes are about one generation ahead of circuit density or minimum half pitch as a result of RET.

III. Economics—Cost of Ownership

The driving force behind the improvement in lithography and Moore's Law in general are economic. The premise is that the cost of producing a wafer should stay roughly constant while the number of circuit elements on each chip doubles every 18 months. This implies that the costs related to lithography must remain roughly constant as performance is improved. Lithography is perhaps the most critical of the processing steps since about half of the capital equipment cost for a wafer fab is in lithography.

There are many sophisticated models for calculating lithography costs. Here, I will use a very simplified version to illustrate the critical issues in lithography. The cost for printing a circuit level on a wafer has three main contributors. They are the costs pertaining to the exposure equipment, the costs for the mask, and the cost of the process

$$\text{Cost/Level} = \frac{\text{Tool}}{T'\text{put}} + \frac{\text{Mask}}{\text{Usage}} + \text{Process}. \tag{4}$$

Equation (4) shows an estimate for these factors. The tool related costs are obtained by taking the cost to operate an exposure tool for an hour (depreciation, labor, space, etc.) and dividing it by the throughput or number of wafers printed in an hour. Traditionally, this has been the dominant factor in lithography costs and has driven equipment manufacturers to produce tools with high throughput. The contribution from the mask is simply the cost of the mask divided by the number of wafers to be printed with that mask. This number is usually limited by design considerations rather than some wear-out or failure mechanism of the mask itself. In DRAM or microprocessor manufacturing, the designs are continually improved and shrunk, resulting in new mask sets. For application-specified integrated circuit (ASIC) and system-on-a-chip (SOC), the total number of parts required for each design is usually small. Industry average numbers for DRAM are 3000–5000 wafers per mask set and 1500 for a microprocessor and 500 or less for ASIC and SOC. The trend is that mask costs are increasing rapidly at the same time that usage numbers are diminishing, mostly due to rapid design revisions. The mask contribution is becoming the dominant factor in lithography costs, particularly as minimum feature sizes fall below the exposure wavelength and MEEF becomes large and enhancements such as PSM

or OPC are required. The third term in the estimate of lithography cost per wafer level is the process. This includes the cost of the resist application and development. Today, the process cost is relatively small compared to the tool and mask terms and is based on a single-layer resist process. Antireflection coatings (ARCs) are often used to suppress interference effects in the resist. These ARCs tend to improve linewidth control and chip performance but add to the process cost. The decreased DOF of short-wavelength high-NA optical lithography may require new types of surface imaging resist materials. However, the additional costs of the extra pattern transfer steps must be taken into account.

Any future lithography technology must address the issues of tool cost, throughput, mask costs, and process costs in order to be viable. The economic driving forces behind Moore's Law will require this and if an economic solution cannot be found for lithography, the historical trend will not continue.

IV. Next-Generation Lithography Technologies

In Fig. 2, the collision of Moore's Law with optical lithography has already taken lithography to the subwavelength regime. A second line parallel to the wavelength trend at $\lambda/2$ is drawn to indicate the theoretical limit to optical lithography. The figure shows that Moore's Law will cross this line somewhere between the 100- and 70-nm technology nodes. It should be noted that the nomenclature for the nodes refers to the minimum half pitch and not the minimum feature size. For example, the 100-nm node requires 100-nm lines and spaces and 70-nm isolated lines (gates). This crossover of the requirements for IC production and performance of optical lithography drives the need for alternate lithography technologies. In addition, the high costs associated with subwavelength lithography could drive the transition to a new lithography technology before this limit is reached. The transition for lithography will likely begin at or during the 100-nm technology node.

There are three main next-generation lithography (NGL) technology contenders for 100-nm lithography and below. They are 157-nm (F_2) optical lithography [8], extreme ultraviolet lithography (EUV) [9]–[11], and electron projection lithography (EPL) [12], [13]. Here, 157 optical is considered to be a departure from current optical lithography methods and is considered as an NGL technology. I will summarize the elements of each of the contending technologies in the following sections.

A. 157-nm Optical Lithography

In the natural progression to shorter exposure wavelengths, the most obvious next generation lithography would be 157 nm that uses an F_2 excimer laser source. Any lithography technology must include the exposure tool, mask, and process. In the case of 157 nm, the exposure tool technology is probably the best understood of these elements. Because of the increased absorption of fused silica at 157-nm wavelength, the optics for a 157-nm system must be made from CaF_2. In order for this technology to be competitive with projected high-NA 193-nm systems, it, too, must be introduced at high NA. This will require large amounts of CaF_2 for building the lenses and result in much higher cost for the system (thought to be about twice that for a 193-nm system). Optical absorption in the mask substrates is also a significant issue for this technology. Normal fused silica is not transparent enough for use as a mask substrate material. On the other hand, CaF_2 is considered to be too expensive and would significantly affect the cost of ownership. Recently, new forms of fused silica have been developed (with reduced OH content and fluorine doping) that show much improved optical transmission. This development has been considered a key enabler for 157-nm technology. The most serious shortcoming of 157-nm lithography, however, is the resist material. All photoresist materials are hydrocarbon based with polymer backbones and specialized additives to improve performance. Hydrocarbon materials absorb 157-nm radiation very strongly, thus preventing a single-layer resist approach. Alternatives include thin-layer or surface-imaging materials or fluorocarbon-based materials. Currently, there are no viable resist materials for 157-nm lithography. If entirely new materials need to be developed, it is not clear whether this can happen in time for the 100-nm IC generation.

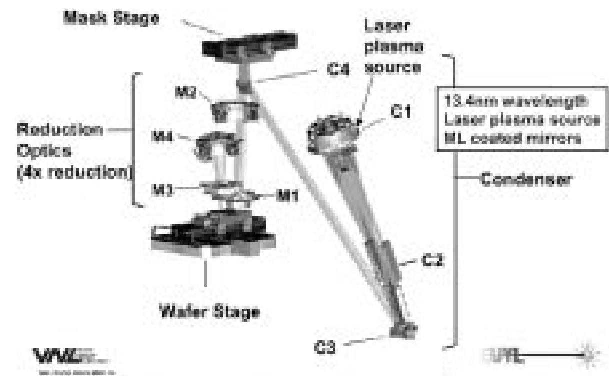

Fig. 6. Schematic diagram of an EUV lithography system showing reflective optics design.

B. Extreme Ultraviolet Lithography

In principle, this is also a logical extension of optical lithography to very short wavelengths (10–14 nm). Extension of conventional refractive optical lithography to wavelengths below the F_2 excimer (157 nm) is problematic due to absorption in the refractive elements. It is also very difficult to design and construct reflective optical systems with sufficiently large NAs to allow printing at or below the wavelength of the illumination source. The idea of EUV lithography is to use small NA reflective optical systems at wavelengths much shorter than the circuit dimensions (see Fig. 6). This combination of small NA and small (compared to the feature sizes being printed) wavelength allows simultaneous achievement of high resolution and large DOF. For example, an EUV system with wavelength of 14 nm and NA of 0.1 can yield 100-nm resolution and 1-μm DOF (assuming a k factor of about 0.7, which is conservative).

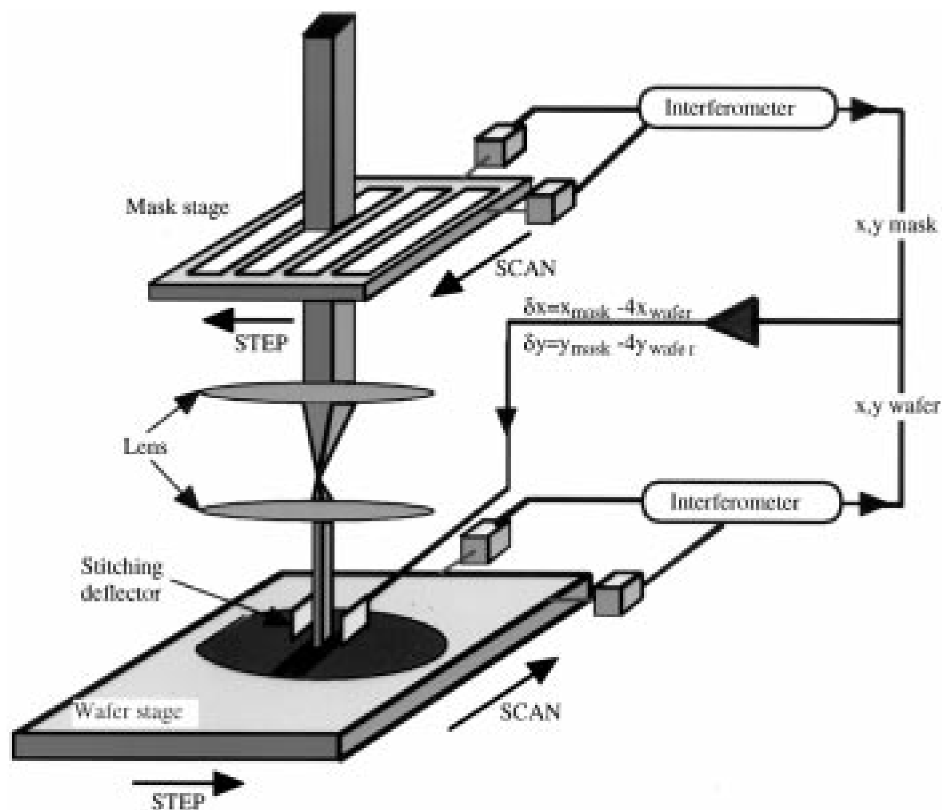

Fig. 7. Schematic diagram of SCALPEL system illustrating the exposure method.

The reflective elements for EUV use multilayer mirrors to produce reflectivities up to nearly 70% at 14 nm. The mask in an EUV system is reflective and also uses the same type of multilayers. A plasma- or synchrotron-based source is used to illuminate the mask, which is imaged by a system of mirrors onto the resist-coated wafer with a reduction factor of four. The optical systems require mirrors with unprecedented tolerances with respect to figure and finish. That is, the shape of the mirror must be correct in addition to the surface being smooth. The specifications are in the angstrom and, in some cases, subangstrom range posing serious challenges for mirror fabrication, coating, and mounting. The masks are made by depositing a multilayer coating on a bare silicon wafer or other flat substrate. An absorber is then deposited and patterned to complete the mask. The most serious issues for EUV masks are in creating multilayer coatings across a wafer with no defects. Even very small (30 Å) defects in the multilayers can print unwanted features on the wafers. EUV radiation is absorbed by most materials over a very short range, typically 20–50 nm. Resist layers required for IC processing must be nearly 1 μm in thickness, therefore requiring some sort of surface imaging method for EUV. This differs significantly from current practice of single thick resist layers and poses many challenges.

C. Electron Projection Lithography

EPL is not an entirely new idea and has been practiced in the past using stencil mask technology. In this case, the mask is a solid membrane with holes in it representing the pattern (stencil). The electron beam is absorbed in the solid parts and passes through the holes, thus imparting the pattern onto the beam. One of the principal difficulties with this approach is that the electrons absorbed in the stencil will deposit a significant amount of energy, causing it to heat up and distort. Accurate image placement is critical to any lithography technology since as many as 25 or more lithographic layers must be overlaid to make a chip. Typically, each layer has an overlay tolerance of one third of the minimum feature size or about 23 nm for NGL at the 70-nm node. Mask heating and the subsequent overlay errors have prevented this type of technology from being widely used. Scattering with angular limitation projection electron beam lithography (SCALPEL)[1] uses masks with scattering contrast to overcome both the mask heating problem and the requirement for multiple exposures in stencil mask technology. Fig. 7 shows a schematic diagram of SCAPLEL systems illustrating the exposure method. The mask is a thin membrane of low atomic number material (1000 Å of SiN, for example), which is transparent to the electron beam (typically 100 keV). The pattern is formed in a thin layer of high atomic-number material (250 Å of tungsten), which is also transparent to the electron beam, but will scatter the electrons. The thickness is usually three mean free path lengths or more. After passing through the mask, contrast is produced in the image with an aperture at the back focal plane of the imaging system that stops the scattered beam and allows the

[1]The SCALPEL program was supported in part by DARPA and SEMATECH.

unscattered beam to pass. The beam is formed from a conventional thermionic emission filament source and focused with magnetic lenses. The image of the mask is projected onto the resist-coated wafer with a 4× demagnification. The resists used for SCALPEL are based on the same chemical platform as those used in today's deep-ultraviolet (DUV) lithography. In some cases, the same material performs well for both electron and DUV exposures. The critical issue for SCALPEL technology is throughput. As the beam current is increased to increase the exposure rate, electron–electron scattering tends to degrade the resolution of the image (stochastic space charge effect). Thus, charged particle lithography such as SCALPEL has a fundamental link between spatial resolution and printing speed. The open question is whether SCALPEL can have sufficient wafer printing speed to make it economically viable. An important aspect of the cost of a lithography technology is the mask technology. In this respect, SCALPEL has a significant advantage over the other technologies, particularly optical with RET. The masks for SCALPEL are made from a fairly simple blank structure with a very thin patterned layer (W). This allows wet chemical etching to be used, even for 100-nm feature sizes, and allows existing infrastructure for optical masks to be used. Further, SCALPEL is a true 4× imaging technology that does not require any form of RET, which has a significant effect on mask costs.

It is not entirely clear which of these technologies will be used in IC manufacturing or exactly when. The key for introduction of 157-nm lithography will be the timely availability of a suitable resist material. In order for 157-nm lithography to represent an improvement over its predecessor (193 nm), high-performance (contrast) imaging materials must be developed. If the technology is developed in time, 157-nm lithography may be used at the 100-nm or perhaps the 70-nm node. It is very unlikely that is extensible beyond 70 nm (less than $\lambda/2$). EPL represents an opportunity for the segment of the IC industry with a low wafer usage per mask (ASIC, SOC) to continue on Moore's curve in an economic way due to relatively low mask costs. EUV looks to be a higher throughput but higher mask cost solution and may be more suitable for the high-volume segment of the industry. The timing of the current active programs in these technologies is such that EPL and 157 nm will be introduced at roughly the same time with EUV two to three years later. One scenario might be that EPL and 157 nm will be used at the 100 nm and possibly the 70 nodes with EPL used for low volume and 157 nm for high volume. At the 50-nm node and beyond, it could be EPL for low volume and EUV for high volume. It seems likely that starting at the 100-nm node, there will be different advanced lithography solutions for different industry segments. This is a departure from the historical trend, but may be necessary to maintain the economic driving forces for each segment.

V. Beyond Next-Generation Lithography

The NGL technologies described above will be useful for the 50 nm and perhaps the 35-nm generations of ICs, but what lies beyond is not entirely clear. Research on transistors and circuits with dimensions of 30 nm or less is being conducted now using direct write electron beam lithography. A very fine (<10-nm diameter) electron beam is used to pattern polymer resist material directly from the pattern data file without the use of a mask. As a result of this serial writing, the process is quite slow. Although useful in research, the serial nature of the process precludes it from use in general IC manufacturing. There are several promising approaches that are being pursued for 50-nm generation circuits and below.

A. Multicolumn Direct-Write E-Beam

The essence of this approach is to use an array of electron beams to overcome the issues of space charge, which limit the performance of EPL as described above [14]. In EPL, all of the electrons that expose the resist on the wafer pass through a common crossover point in the electron optics. The space charge effect is strongest in this region of high current density in the beam. In the multibeam approach, an array of electron beams may produce the same or larger total beam current, but there is no common crossover, thus minimizing the space charge effect. This can allow much smaller features to be printed at a given beam current or throughput. The performance would be limited by aberrations and diffraction and not tied to throughput. At 100-keV electron beam energy, the wavelength of the electrons is 3.7 pm. With a typical NA in the range of 10^{-3}, the resolution of these systems can, in principle, be about 4 nm.

B. Scanning Probe Arrays

Scanning force microscopy is capable of atomic scale imaging resolution. These scanning probes can also be used to affect surface chemistry and, thus, patterning. One approach is to use an electrical bias on the scanning probe tip to generate field emission [15]. The resultant beam is proximity focused onto the sample (wafer). The current impinging on the wafer has been used to desorb hydrogen from oxide layers and change their solubility. This can produce relief images that could be used for circuit patterning. This is also inherently a serial process and will require a large array of scanning tips in order to operate at economically viable throughput. The technology developed for micromechanical devices is now being used to develop such a tip array. The number of tips required in the array depends on the process used. It is estimated that 10^6–10^9 tips will be required, a daunting but perhaps not impossible task.

C. Nanoimprinting

Nanoimprinting uses a molding or stamping process to replicate patterns from a master. It is somewhat analogous to the process currently used to replicate compact disks. There are many variations of this approach, but all essentially use a mold to transfer a pattern by mechanical means to a layer on a substrate. In one variant, a relief mold is fabricated in a hard layer of SiO_2. This mold is then pressed into a relatively soft polymer layer on the substrate. This method has been shown to be capable of features as small as 10 nm and has promise

as a simple replication method. There are two primary difficulties in applying this to IC manufacture. The first is the generation of the mold. These molds must be patterned by some other method that has the same (or better) lithographic quality as the final image made by imprinting. As discussed above, reduction approaches have significant advantages in this regard. The second issue is alignment. The level-to-level alignment accuracy required in IC manufacturing is generally one third of the minimum feature size. It is not clear how this can be achieved in nanoimprinting due to errors in the mold fabrication, distortion of the mold across the chip or wafer.

VI. Alternatives and Conclusion

The requirements for lithography include resolution, CD control, overlay accuracy, and throughput. The first three are quality metrics while the fourth is an economic consideration. All four of these criteria are important for a successful technology. Most of the focus of this article and indeed lithography research is on resolution. This is due to the fact that lithography has always been based on optics and its extensions. Improvements in resolution have been the primary driver and improvements in control and accuracy have followed. As described above, throughput has remained constant and even increased with the evolution of optical lithography. In considering alternate solutions, this assumption is not necessarily valid. It is also true that device scaling, as applied to lithography, has been an essentially linear process. That is, as resolution and critical feature sizes scaled downward, the requirements for CD control and overlay scaled directly. The requirements of CD control have been set at $\pm 10\%$ of the critical dimension, whereas overlay accuracy has been set at about one third of CD.

Alignment and overlay performance are also critical parameters in lithography. In many cases, the tolerance to which layers in a circuit can be aligned will determine the packing density of the features. Much of the progress in lithography has been aimed at improvements in resolution. However, without commensurate improvements in overlay performance, some of the benefits of improved resolution cannot be realized. In a full-field lithography tool, where all of the chip is printed in a single optical field (as depicted in Fig. 1), overlay performance is determined primarily by alignment capability and residual distortions in the optical system. Alignment schemes using diffraction patterns from corresponding target marks on the mask and wafer have been improved in sophistication. Alignment marks have been tailored to specific levels and processes to optimize performance. Further, distortions in the lenses have been minimized by improved multielement optical designs. The residual (mostly fifth-order) distortions result in features being printed on a grid that varies from the ideal. When subsequent layers are printed on different tools, the differences in these distortions can result in overlay errors. For this reason, sometimes dedicated tools are used for critical layers for improved overlay performance. In the future, subfield printing techniques such as EPL offer the possibility to adjust feature placement within the fields and modify the distortion signature of the exposure tool to match underlying layers and achieve superior overlay performance.

Recent work in nanotransistors has demonstrated that the relationship between gate size and transistor performance is no longer scaling linearly [16]. In fact, the data shows that for transistors with gate sizes of 50 nm and below, CD control requirements will be much tighter, perhaps as small as $\pm 5\%$. This will be an important factor in the limits of lithography for transistor fabrication. It would also tend to eliminate nonlinear lithography approaches that use masks and also those that do not use demagnification.

As transistor gate sizes scale to below 30 nm, the requirements on CD control approach atomic dimensions. There are many factors that will make scaling of any conventional polymer resist-based lithography problematic. The molecular size of the resist constituents will then be larger than the required feature size control.

Alternative transistor designs and architectures may offer answers to further scaling. Vertical replacement gate (VRG) transistors change the fine lateral dimensional control requirements to layer thickness control requirements [17]. In the VRG, the gate length is controlled by layer thickness with the sides of the patterned layer acting as the gate control for the transistor. Thin film deposition methods are capable of atomic-level thickness control across large areas and may offer the solution to CD control. Advanced lithography is still required for the device patterning, particularly for isolation but the transistor performance will depend less on lithographic performance than metal-oxide-semiconductor (MOS) transistors. The VRG technology addresses mainly the gate size and control issue, but does not improve packing density over complementary metal-oxide-semiconductors (CMOSs) using the same baseline process toolset. It is in some ways analogous to the PSM technology that yields improved transistor performance, but not density.

Lithography scaling, the key enabler for Moore's Law, will continue for the next several IC generations well beyond the 100-nm technology node and likely to the 30-nm node. There is no shortage of solutions for high-resolution printing methods that can go well beyond 30 nm, but the ultimate limit to lithography scaling will more likely be set by CD control requirements and economics rather than purely resolution performance.

Acknowledgment

The author would like to thank Dr. S. Okazaki of Hitachi Ltd. (currently on assignment to ASET) for the use of Fig. 1 for this paper.

References

[1] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors: 1999*. Austin, TX: SEMATECH, 1999.

[2] A. Nakae, K. Kamon, T. Hanawa, K. Moriizumi, and H. Tanabe, "Improvement in optical proximity correction by optimizing second illumination source," *Jpn. J. Appl. Phys., Pt. 1*, vol. 35, p. 6396, 1996.

[3] B. G. Kim, S. W. Choi, W. S. Han, and J. M. Sohn, "Beneath the MEEF," *Solid State Technol.*, vol. 43, no. 8, p. 107, 2000.
[4] K. Kamon, T. Miyamoto, Y. Myoi, H. Nagata, M. Tanaka, and K. Horie, "Pholithography system using annular illumination," *Jpn. J. Appl. Phys.*, pt. 1, vol. 30, p. 3012, 1991.
[5] M. D. Levinson, N. S. Viswanathan, and R. A. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Trans. Electron Devices*, vol. ED-29, p. 1828, 1982.
[6] H.-Y. Liu, L. Karklin, Y.-T. Wang, and Y.C. Pati, "The application of phase-shifting masks to 140 nm gate patterning (II): Mask tolerances," *Proc. SPIE*, vol. 3334, p. 2, 1998.
[7] M. D. Levenson, "Wavefront engineering from 500 to 100 nm CD," *Proc. SPIE*, vol. 3051, p. 2, 1997.
[8] T. M. Bloomstein, M. Rothschild, R. R. Kunz, D. E. Hardy, T. B. Goodman, and S. T. Palmacci, "Critical issues in 157 nm lithography," *J. Vac. Sci. Technol. B*, vol. 16, p. 3154, 1998.
[9] W. T. Silvast and O. R. Wood II, "Tenth micron lithography with a 10 Hz 37.2 nm sodium laser," *Microelectron. Eng.*, vol. 8, p. 3, 1988.
[10] A. M. Hawryluk and L. G. Seppala, "Soft X-ray projection lithography using an X-ray reduction camera," *J. Vac. Sci. Technol. B*, vol. 6, p. 2162, 1988.
[11] C. W. Gwyn, R. Stulen, D. Sweeney, and D. Attwood, "Extreme ultraviolet lithography," *J. Vac. Sci. Technol. B*, vol. 16, p. 3142, 1998.
[12] J. M. Gibson and S. D. Berger, "A new approach to projection-electron lithography with demonstrated 0.1 linewidth," *Appl. Phys. Lett.*, vol. 57, p. 153, 1990.
[13] H. C. Pfeiffer and W. Stickel, "Prevail: An e-beam stepper with variable axis immersion lenses," *Microelectron. Eng.*, vol. 27, p. 143, 1995.
[14] T. H. P. Chang, D. P. Kern, and L. P. Muray, "Arrayed miniature electron beam columns for high throughput sub-100 nm lithography," *J. Vac. Sci. Technol. B*, vol. 10, p. 2743, 1992.
[15] K. Wilder, H. T. Soh, A. Atalar, and C. F. Quate, "Hybrid atomic force/scanning tunneling lithography," *J. Vac Sci. Technol. B*, vol. 15, p. 181, 1997.
[16] D. M. Tennant, G. L. Timp, L. Ocola, M. Green, T. Sorsch, A. Kornblit, F. Klemens, R. Kleinman, D. Muller, Y. Kim, and W. Timp, "Gate technology issues for silicon MOS nanotransistors," in *Proc. Materials Research Soc. Symp.*, 1999.
[17] J. M. Hergenrother, D. P. Monroe, F. P. Klemens, A. Kornblit, G. R. Weber, W. M. Mansfield, M. R. Baker, F. H. Baumann, K. J. Bolan, J. E. Bower, N. A. Ciampa, R. A. Cirelli, J. I. Colonell, D. J. Eaglesham, J. Frackoviak, H. J. Gossman, M. L. Green, S. J. Hillenius, C. A. King, R. N. Kleinman, W. Y.-C. Lai, J. T.-C. Lee, R. C. Liu, H. L. Maynard, M. D. Morris, C.-S. Pai, C. S. Rafferty, J. M. Rosamilia, T. W. Sorsch, and H.-H. Vuong, "The vertical replacement-gate (VRG) MOSFET: A 50 nm vertical MOSFET with lithography-independent gate length," in *Proc. Int. Electron Devices Meeting*, 1999.

Lloyd R. Harriott received the M.A. and Ph.D. degrees in physics from the State University of New York, Binghamton, in 1976 and 1980, respectively.

He joined Bell Laboratories in 1980 and was part of the team that developed the EBES4 field emission electron beam lithography system for mask and wafer lithography. From 1983 to 1990, he developed fundamental technology and applications for focused ion beams, including defect repair and circuit diagnosis. He became a Technical Manager in the Advanced Lithography Research Department in 1993 and Director of the department in 1996 with responsibility for advanced optical and next generation lithographies. He is currently the Program Manager of the SCALPEL projection electron beam lithography program at Bell Laboratories and Lucent Technologies. His research includes liquid metal ion sources, ion-solid interactions, beam induced chemistry for deposition, and developing an all-vacuum lithography process combining molecular beam epitaxy, low damage dry etching, and focused ion beam patterning for development of laser and quantum structures in InP materials. Recent research includes advanced optical lithography including DUV at 248 nm and 193 nm as well as resolution enhancement technologies such as phase-shift masks and optical proximity correction as well as early work in EUV lithography.

Dr. Harriott received the Bell Laboratories Distinguished Member of Technical Staff award in 1987.