

Modeling X-ray proximity lithography

by Jerry Z. Y. Guo
F. Cerrina

Advanced semiconductor circuits, such as DRAMs, are based on very complex fabrication processes. Because of the cost and complexity involved, it is rapidly becoming impossible to adopt a "trial-and-error" approach in the development stage of a new process. Fortunately, the advances in computer power spurred by the new semiconductor devices have made it possible to compute the response of complex systems in a reasonable time on workstations. Thus, the study of a virtual representation of the process (that is, a model) can represent a solution to the high cost of process development—of course, after verification of the model accuracy through controlled experiments. A correct physical interpretation of the process under study is necessary in order to implement a model that is both accurate and extendible. This is particularly true for new approaches, such as those involved in X-ray lithography. We have studied the process of image formation in X-ray lithography and have implemented several models to predict the intensity distribution at the wafer plane. The models can be applied to the definition of an optimal exposure system

that will provide the maximum exposure latitude, and to the study of new types of X-ray masks.

Introduction

Although X-ray lithography was invented over 20 years ago [1], it has only recently become a real manufacturing tool in microelectronic processing. Much progress has recently been made in light sources, illumination optics, mask fabrication and inspection, and stepper technology. A commonly held view is that X-ray proximity lithography will compete economically with optical projection lithography when dimensions shrink to a quarter micron and below.

Proximity printing is often referred to as "shadow-casting" imaging. A stencil-like object (mask) is held close to a recording surface (wafer) and illuminated with light (X-rays) of suitable wavelength. The light casts an image which is formed by illuminated areas under the transmitting parts and by shadows under the absorbing parts. Of course, the light must be energetic enough to introduce a change in the recording medium (the photoresist). Although not immediately noticeable (hence "latent"), the change can be evidenced by a subsequent treatment (development). We notice that there are no optical elements between the mask (object plane) and the wafer (image plane), so it is not really an imaging system

©Copyright 1993 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of this paper must be obtained from the Editor.

using conventional optics terminology. Broadly speaking, the sharpness of the image is controlled by the diffraction of the light as it propagates from the mask across the gap to the wafer. In an edge diffraction, the distance from the edge to the peak of the diffraction pattern is given by a simple relation:

$$\text{blur} \sim \sqrt{\lambda g}, \quad (1)$$

where λ is the wavelength and g is the proximity gap. Hence, the faithful transfer of features from the mask to the wafer over a finite distance must rely on the use of short-wavelength radiation. For deep submicron resolution, soft X-rays are required. The short wavelength of soft X-rays has associated with it many issues involving source, materials properties, and diffraction and absorption processes that must be addressed in detail to produce a realistic model of the global imaging processes. First, the light source (synchrotron radiation or plasma source) for X-ray lithography is quite different from any light source ever used in lithography, so its role in the image formation had not been fully studied until recently. The effect of the optical system used to relay the radiation to the mask (if any) can be included in the description of a new, effective source. Second, because of the wavelength range, relatively thick (in wavelength units) mask absorbers must be used to provide the modulation in the transmitted light. This is the case even for high-atomic-number materials such as Au or W, where for instance the complex electric permittivity $\epsilon = 1 - \alpha + i\beta$, where $\beta = 7-10 \times 10^{-4}$ [2]. The effect is quite different from that of the thin chrome mask used in optical lithography, where $\beta = 3.4$. Finally, X-ray photons are about 1000 times more energetic than UV photons. A given amount of energy is deposited by a correspondingly smaller number of photons, and the use of continuous field approximation implicit in intensity calculations is good only as a prediction of statistical expectation.

In this paper, we give a comprehensive description of these aspects of the image-forming system and their effects on image formation.

• X-ray lithography systems

X-ray lithography is based on the use of an intense X-ray source, a radiation relay system, a mask, and a resist system. These are complex subsystems, large and expensive to build. Clearly, it is difficult if not impossible to test several different types of beamlines or masks in order to select the best process. A much better approach is based on the definition of a suitable model for the various subsystems and for the system itself. This model, residing in the memory of a computer, can be exercised in order to evaluate many alternative strategies and to finally select the best approach. It is critical, we believe, that models be physically based, rather than "black boxes." Although it is

possible to specify the response of a system in limited conditions by using a system transfer function obtained experimentally, this is very risky, since some hidden variables may have been neglected. Furthermore, extrapolation beyond the measurement range is dangerous, and the margin of error increases rapidly with the distance in parameter space. It is much better to establish and verify a correct physical model, since the extrapolation will be much more reliable. Also, increased understanding of the physical (and chemical) processes underlying the model often provides novel ideas and extensions.

In the following sections, we concentrate on the development of physical models for the various systems, and we investigate the conditions under which they can be linked together. Of foremost importance is the definition of fast and efficient algorithms that can be executed on a workstation (such as an IBM RISC System/6000® processor) on a time scale of minutes, and rapidly visualized. It is also important that data be presented with quantitative information that can be related directly to the output of measuring tools for verification of the model itself. We do not present the algorithms themselves, but rather discuss their basis and the limits of their application. The various codes are part of a comprehensive model developed at, and available from, CXrL, the Center for X-Ray Lithography at the University of Wisconsin-Madison.

System description

A typical X-ray proximity lithography system, as shown in Figure 1, consists of an X-ray source, illumination optics, a mask, and a photoresist-coated wafer separated by a gap in the range of 5–40 μm . Like any optical system, it can be divided into an input field, propagation through optics, and an output field. For X-ray proximity lithography, the input field is the field immediately *beyond the mask*, the optical system is the *free-space propagation*, and the output field is the field at the *resist surface*. For the simple case of a coherent source, the transfer function of a cascaded system can simply be multiplication of the transfer functions of the subsystems. For the partially coherent cases, this is not possible unless the optical system satisfies some restrictive conditions, and more complex treatments need to be carried out.

Input field

The input field is defined by the mask and by the illuminating radiation. A radiation source is fully characterized by its spectral and geometrical properties of the radiated power flux Φ —in particular:

1. Spectral distribution (in frequency ν , $d\Phi/d\nu$).
2. Spatial extension (in lateral dimension x , $d\Phi/dx$).
3. Angular extension (in direction cosine s_x , $d\Phi/ds_x$).

These properties are summarized in the spectral "brightness" $B(x, s_x; \nu)$ defined in Appendix E; different sources (plasmas, synchrotrons, etc.) have different brightness functions. The illumination optics are defined in terms of their modifications to the brightness. In general, in the soft X-rays, optics are formed by glancing mirrors, although filters also are used. Their effect on the modeling is fully specified (including partial coherence) by the appropriate changes in brightness B .

Since low- Z (atomic number) materials have the lowest absorption coefficient ($\mu \approx 10^4 \text{ cm}^{-1}$), the mask is formed by a free-standing thin film of SiC, Si_3N_4 , or Si 1–2 μm thick. This film is used to support the absorber pattern (hence the name "carrier"), which is, in turn, fabricated from a film of high- Z material, usually 0.25–0.7 μm thick, depending on the spectrum of the X-rays. The goal is to provide a modulation in the range of 5:1 to 10:1. Au, W, and Pd (with $\mu \approx 5\text{--}8 \times 10^5 \text{ cm}^{-1}$) are acceptable optically and from a processing point of view.

As is discussed later, the input field cannot be obtained simply by multiplying the source field (illumination field) with the complex mask transmission function when the edge effect of mask absorbers is considered.

Propagation

Between the mask and wafer is a gap filled with a uniform medium (typically helium at normal conditions). The image formation, however, is more complex than that of a projection imaging system, because the system transfer function of proximity printing modulates the phase instead of the amplitude of the input field, as discussed in detail below.

Output field and recording

The resist is composed of an organic film whose dissolution properties are radically altered by exposure to the radiation [3]. The changes are brought on by a modification of the chemical bonding in some of the components of the resist. In the X-ray region, the photoelectrons created by the direct absorption process (primaries) are too energetic to effect these chemical changes, which typically involve energies around 3 eV. Each absorption event creates one photoelectron and one Auger electron, since the probability of fluorescence is so small that it can be neglected at low Z . The photoresist is exposed by the low-energy secondary electrons generated by the fast-moving primaries. The energy of a photoelectron, or an Auger electron, is gradually transferred to the resist during a sequence of scattering events, over a distance of a few tens of $\text{\AA}$. This results in a *blur* of the image formed by the radiation. The distribution of this energy in the resist is the so-called dose image.

It is important to note that, although the image is defined by both the magnitude and the phase of the

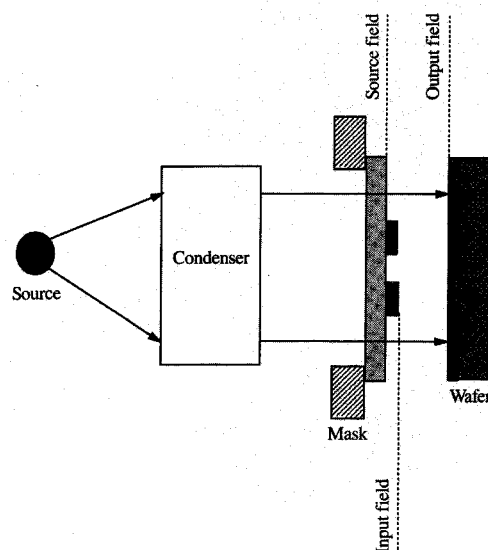


Figure 1

Illustration of a typical X-ray lithography system. Only grazing-incidence mirrors can be used as condensers.

electromagnetic wave, the exposure of photoresist is an electronic process. The probability of a photoelectron event is thus proportional to the integrated *intensity* over a period of time much longer than the oscillation period of the wave and is insensitive to the phase. Thus, the aerial image is defined as the intensity distribution of light (X-rays) in the resist. Rigorously speaking, this distribution is three-dimensional. However, since the diffraction effect and the decay of X-ray intensity in a distance of about 1 μm (single-layer resist thickness) are very small in comparison with the gap distance, we restrict the analysis to the two-dimensional distribution at the resist surface. The main effect is caused by the spectral dependence of the absorption coefficient on the wavelength, causing a differential attenuation of the image components. This effect may become noticeable at longer wavelengths, where $\mu(\nu)$ is larger. After exposure, the wafer is put into a developer to dissolve the exposed part for positive resists and the unexposed part for negative resists. The dissolution rate of resist in a developer is a function of absorbed energy distribution, and may further modify the profile of the image.

• Scope of calculations

It is important to distinguish between an abstract image fidelity and realistic lithographic criteria. In manufacturing, the important issue is that of the image (linewidth) control

over a range of exposure parameter variations, i.e., *exposure latitude*. In other words, the image printed in the resist must be, and must remain, within some specified limits from the nominal value. For comparison purposes one uses a 10% linewidth variation, although this value is a *total* value that includes all possible contributions. For instance, the actual budget allocated to intensity variations is much smaller, closer to 1 or 2%. In this discussion we concentrate on the study of the variation of the linewidth with *exposure dose* (D) and with the changes in the *gap* (g). This is equivalent to the dose-defocus case of optical lithography. It is important because of the topography on the underlying wafer and because of possible variations in the intensity across the exposure field. The domain in the subspace $\{D, g\}$ for which the changes in linewidth are less than 10% defines the exposure window. Clearly, the issue of the resolution of the lithography is implicitly taken into account—poor resolution means zero exposure window.

• Review of modeling approaches

The modeling of the image formation in X-ray lithography basically must answer two questions: What is the exposure window of X-ray lithography, and under what conditions can it be maximized? The implementation (and verification) of a successful model force an understanding of the physical processes involved, providing a firmer basis on which to build an optimization scheme.

Only two years ago, modeling could not explain the large exposure window and resolution demonstrated by the experiments by many groups worldwide [4–6]. Efforts were made to eliminate some factors appearing in the process which were later discovered to be positive. In the literature, it was generally assumed that the light propagation from the mask to the wafer is in the Fresnel region, so that a scalar diffraction theory with Kirchhoff boundary condition is valid in modeling the image formation process [5, 7, 8]. The broad-band nature was considered in simulations using synchrotron radiation as a light source [5, 8–10]. However, conclusions drawn from those calculations do not completely reflect reality, because important factors such as partial coherence, electron scattering, absorber sidewall profile, and fringing boundary conditions were not considered. As a result, unrealistically small exposure windows were predicted [9]. Later, the importance of blurring effects was discovered, and much better exposure windows were shown in several different studies [11–15]. Simplified models showed that high-contrast masks are desirable. A modulation of 10 dB (contrast 10:1) of the mask was generally considered necessary to produce a good print. It was later pointed out by Ku et al. [7] that a thickness corresponding to π phase shift, *while maintaining 10-dB modulation*, would give a higher exposure latitude. However, the simulation of [7] shows a large side lobe which would give rise to ghost

features which were not observed in experiments for the given conditions. The reason is that partial coherence was not considered in the simulation. The 10-dB modulation was later shown to be unnecessary [12]; a lower-contrast mask could actually give better exposure latitude, and only the phase shift required optimization. All of the papers cited above assumed clear-cut boundary conditions at the plane immediately after the mask. This *no-fringing* assumption produces very small error for large gaps with blur, but for small gaps the fringing effect is not negligible [16–18].

Models of image formation in proximity lithography

Image formation in proximity X-ray lithography can be described in the framework of the propagation of electromagnetic waves with appropriate boundary conditions. This implies that the interaction of light with materials is treated as a lumped parameter problem; i.e., a material is assumed to be uniform, and the interaction is manifested completely through the use of the complex dielectric constant $\epsilon(\omega)$ or the complex refractive index $n(\omega)$ at the frequency ω . These are typically obtained from the atomic scattering factors $f(\mathbf{k}, 0)$, which are well known from experiments and theory:

$$\epsilon(\omega) = 1 + \frac{4\pi N}{k^2} \mathbf{e}^* \cdot \mathbf{f}(k, 0), \quad (2)$$

via the use of the optical theorem [19], where k is the wave number, $\mathbf{e}^*$ is the polarization unit vector, and N is the density of atoms. The propagation can then be described with the readily available formalism of the propagation of partially coherent light. The use of lumped parameters requires a subsequent refinement of the models to include edge and other effects. Another approach is based on the use of the scattering theory. In this case, only atomic quantities are used and, in principle, the model is simpler. However, the computational level is much greater, justifying the use of the lumped optical models.

The following discussion is based on an axiomatic approach, where we follow the propagation of a beam of radiation through the optical system in the framework of partially coherent fields and stochasticity. The language is less intuitive than that of electromagnetic fields, but is considerably more succinct. The calculations thus yield the *expected value* of the output field, $\langle E^*E \rangle$, i.e., an observable quantity. The observed events closely reproduce the expected value if the number of events is large.

• Partially coherent light

Diffraction

We recall briefly that an electronic process records the optical intensity, which is determined by the

transformation of *mutual spectral density function* as polychromatic partially coherent light propagates through an optical system. The light intensity is defined as an expectation value [20]

$$I(\mathbf{r}) \equiv \langle E(\mathbf{r}; t)E^*(\mathbf{r}; t) \rangle = \int P(\mathbf{r}; \nu) d\nu, \quad (3)$$

where P is the power spectral density function, defined as

$$P(\mathbf{r}; \nu) \equiv G(\mathbf{r}_1, \mathbf{r}_2; \nu)|_{\mathbf{r}_1=\mathbf{r}_2=\mathbf{r}}. \quad (4)$$

$G(\mathbf{r}_1, \mathbf{r}_2; \nu)$ is the mutual spectral density function at points $\mathbf{r}_1$ and $\mathbf{r}_2$ and frequency ν [21]. From (3) we can see that the problem of finding the intensity is reduced to that of finding the power spectral density function.

Propagation of partially coherent light and cascaded systems

The propagation of monochromatic coherent light (represented by electric field amplitude E) through a linear system with point spread function $K(\mathbf{r}, \mathbf{r}'; \nu)$ is given as

$$E_o(\mathbf{r}, \nu) = \int E_i(\mathbf{r}', \nu)K(\mathbf{r}, \mathbf{r}'; \nu) d\mathbf{r}', \quad (5)$$

where E_o and E_i are the output and input electric fields, respectively. The propagation of partially coherent light through the same system is represented by a bilinear transformation of the mutual spectral density function of the form

$$G_o(\mathbf{r}_1, \mathbf{r}_2; \nu) = \iint G_i(\mathbf{r}'_1, \mathbf{r}'_2; \nu)K(\mathbf{r}_1, \mathbf{r}'_1; \nu)K^*(\mathbf{r}_2, \mathbf{r}'_2; \nu) d^2\mathbf{r}'_1 d^2\mathbf{r}'_2, \quad (6)$$

where G_o and G_i are the output and input mutual spectral density functions, respectively. A complex system separated by boundary regions can be divided into a sequence of subsystems; the output of each subsystem is given by an operation such as (6) and then used as the input for the following subsystem.

In X-ray lithography, the illumination system characteristics are usually obtained analytically except for the spectrum; the numerical simulation starts from the source field plane in Figure 1 in front of the mask absorber and proceeds to the calculation of the dose absorbed in the resist. The system can be divided into the mask subsystem, the gap subsystem, and the resist subsystem.

Boundary conditions In a scalar diffraction theory, it is assumed that the field is not disturbed by the presence of a screen (a perfect conductor) and that the field behind the screen is zero. In the soft X-ray regime, there is no perfect conductor; for instance, the field behind the metal absorber is as high as 20% of the field in the clear region. A simplistic way of accounting for the field behind the screen

is to propagate the incident light as a plane wave and to assume that the wave decays, as in bulk materials. This assumption leads to a step change in both the amplitude and the phase of the light immediately after the mask at the interface between the absorber region and the clear region. In reality, as the light propagates along the absorber surface, a ringing field is created at the edge. The plane wave simplification does not cause much difference in the bulk region far from the edge. The diffraction of the square wave in the near field differs from that of the real pattern, which leads to a discrepancy in the final image intensity distribution. The extent of the discrepancy depends on the light wavelength, the dimension of features, the material properties of the absorber, the thickness and sidewall slope of the absorber, and the gap. The thickness of the absorber can be greater than the lateral dimensions of the features on the mask. The light propagating from the membrane to the plane immediately after the absorber is confined by the metal absorber ($n < 1$) and thus constitutes a waveguide. For a given wavelength, the waveguide can support only a limited number of normal modes (guided modes with a discrete spectrum of wave numbers) and a continuous spectrum of radiation modes which decay much faster than normal modes. For X-rays, the metal absorbers (Au or W) cannot be considered as perfect conductors because of their low absorption coefficient. It is well known that the modes in a lossy waveguide are not orthogonal to one another, thus making it difficult to treat the problem analytically. From a user's point of view, a numerical solution is more desirable for *in situ* simulation.

Scalar diffraction theory As long as the boundary conditions are given correctly, scalar diffraction theory is very accurate in predicting intensity pattern. There are two versions of scalar theory for diffraction from a plane boundary; as discussed in Appendix A, one is the Fresnel-Kirchhoff theory, and the other is the Rayleigh-Sommerfeld theory. The latter is more accurate in the very near field, which in the spatial frequency domain is given as

$$\mathcal{H}(f_x, f_y) = \exp[ig\sqrt{k^2 - 4\pi^2(f_x^2 + f_y^2)}], \quad (7)$$

where $\mathcal{H}(f_x, f_y)$ is the Fourier transform of $K(x, y)$, g is the proximity gap, k is the wave number, and f_x and f_y are the spatial frequencies in the x and y directions, respectively.

• *Simulation algorithms*

The simulation of the imaging process usually begins with the illumination on the mask and ends with the computation of the final aerial image or energy deposition in the resist. If a resist profile is desired, a dissolution simulation is also needed. Since the dissolution process of

an X-ray resist does not differ from that of other resists, it is not covered here. If X-ray lithography is taken as a cascaded system of many subsystems, the brightness of light (mutual spectral density) at the mask plane can usually be obtained analytically; we treat the propagation of light from the front surface of the mask to the wafer surface with computer simulation, which can consist of two parts, the propagation of light in the mask (absorber) and the simulation of diffraction of partially coherent light.

Diffraction simulation

Given a mutual spectral density function on the mask surface, there are two routes which can be taken. The first is use of the Hopkins method, which divides the problem into the *system* part, including the illumination and imaging system transfer function, and the *input* part, which is the mask transmission function. The second route is decomposition of the partially coherent illumination into an equivalent distribution of incoherent point sources (*modes* or *channels* of individual coherent illumination), the final image being the incoherent summation of the intensities from each individual mode. The Hopkins method is efficient when the system and illumination are fixed and only the object changes each time the simulation program is run. In X-ray lithography, various illumination conditions must often be simulated, and the method is not efficient enough to be used in real time for polychromatic light. The modal expansion or decomposition method also is very time-consuming with polychromatic light, but approximations can be made with enough accuracy to make this approach more attractive in proximity imaging simulation.

The simulation starts with the mutual spectral density function of the illumination light given at the incident plane of the mask,

$$G(x_1, x_2, z_m; \nu) = A(x_1; \nu) A^*(x_2; \nu) \mu(x_1 - x_2, z_m; \nu), \quad (8)$$

where z is the propagating direction and z_m denotes the mask plane, $A(x; \nu)$ is the light field amplitude, and $\mu(x_1 - x_2; \nu)$ is the normalized spatial coherence at frequency ν . For simplicity of notation, only one lateral dimension x is used (the y direction is completely symmetric). The above form of illumination assumes homogeneous partial coherence over the field and an inhomogeneous intensity distribution; it is applicable to a wide variety of illuminations. Given the mutual spectral density function in the mask plane and the mask transmission function $T(x, \nu)$, the task is to find the power spectral density function in the wafer plane,

$$P(x, z_w; \nu) = \iint G(x'_1, x'_2, z_m; \nu) T(x'_1; \nu) T^*(x'_2; \nu) \times K(x - x'_1; \nu) K^*(x - x'_2; \nu) dx'_1 dx'_2, \quad (9)$$

where z_w denotes the wafer plane. The above formalism assumes a thin object; i.e., the output field at one position on the mask is affected by the input field at that position alone—in system terminology, a “no-spread” condition. In reality, the mask subsystem has a point spread function, and the thick absorber on the mask causes fringing of the output field. The *fringing boundary condition* makes the simulation problem much more complicated. With modal decomposition, the exact effect of the mask subsystem can be taken into account, but with a great sacrifice in computation time. However, in the domain of X-ray proximity lithography, approximating the mask subsystem with a modified mask transmission function $T(x, y, \nu)$ (thin object approximation but with fringing effect) causes very little error, as discussed later. If we expand partially coherent light into many modes with the weight of mode s (point source) denoted as $\xi(s)$, the power spectral density function in the wafer plane (see Appendix C) is given as

$$P(x, z_w; \nu) = \int \xi(s) N_p(x, s) ds, \quad (10)$$

where N_p is the normalized intensity of the diffraction pattern of mode s . Furthermore, if the diffraction is in the Fresnel region, (10) can be simplified (see Appendix C) to

$$P(x, z_w; \nu) = \int \xi(s) N_p(x - gs) ds = \xi\left(\frac{x}{g}\right) \otimes N_p(x). \quad (11)$$

Equation (11) becomes a convolution (a linear shift-invariant system), and it shows that a finite beam divergence (partial coherence) is equivalent to a low-pass filter in proximity imaging. It gives rise to a blur of the image of coherent diffraction. It is interesting to observe that Equation (51) in Appendix C is the usual form of geometric blurring, with scaled $\xi(s)$ as the point spread function. The importance of the above result also lies in the fact that partial coherence (or finite source) is equivalent to any other type of blur (incoherent process). For instance, a vibration between the mask and the wafer has the same effect as the finite source size. The 3σ width of function $\xi(s)$ is normally referred to as the blur angle; the effect of small blur angle (3 to 6 mrad) on the exposure window of X-ray proximity lithography is positive, because it eliminates the spurious edges of the coherent diffraction pattern and also reduces the sudden change of pattern as the gap changes, thereby increasing the depth of focus. However, if the blur is too large, it reduces the exposure latitude by reducing the contrast. An example of the simulation of an image of 0.1- μm feature is shown in Figure 2.

Boundary conditions—beam propagation method

In the above treatment, we have assumed a complex mask transmission function $T(x, y; \nu)$ for light propagating

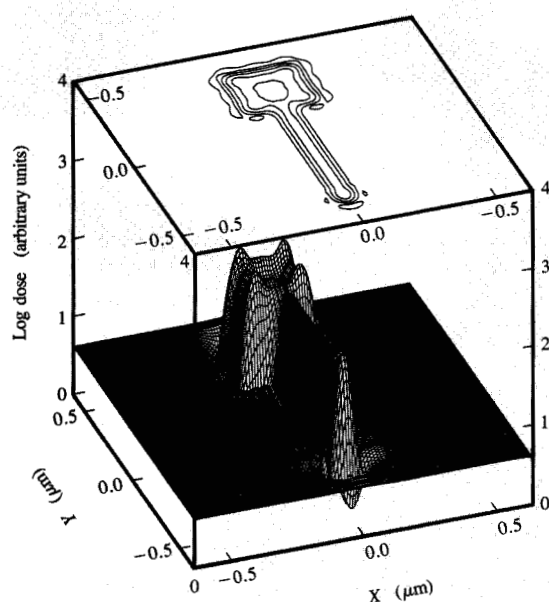


Figure 2

Simulation of a 0.1- μm gate structure with the double linear shift-invariant systems approach; Au absorber thickness 0.4 μm , 100-mrad slope sidewall, gap 10 μm , blur 3 mrad (3σ).

through the mask. The mask is actually a subsystem which has a point spread function. In this section, we formulate this subsystem by the beam propagation method (BPM). From BPM, we also can obtain the fringing effect and use it as the effective mask transmission function. BPM works best when the complex dielectric constant ϵ changes smoothly in the propagating direction. In X-ray lithography, ϵ is constant along z when the absorber has a vertical sidewall. Even with a sloped sidewall, the difference of ϵ between absorber and vacuum is still very small (10^{-3}). A brief description of the method is given in Appendix B; for a detailed discussion the reader is referred to [22].

Figure 3 shows the difference between the intensity distribution at the plane immediately behind the mask from assuming a plane wave propagation and that from BPM. A point source with a single wavelength (temporally and spatially coherent light) is used in the simulation. The modification to the boundary condition by the use of BPM calculation is significant when the gap is small, especially at the center of the feature and at the edge of the feature (nominal linewidth point), which determines the exposure latitude. When the gap is large or the absorber sidewall has a slope of a few degrees, however, the difference between

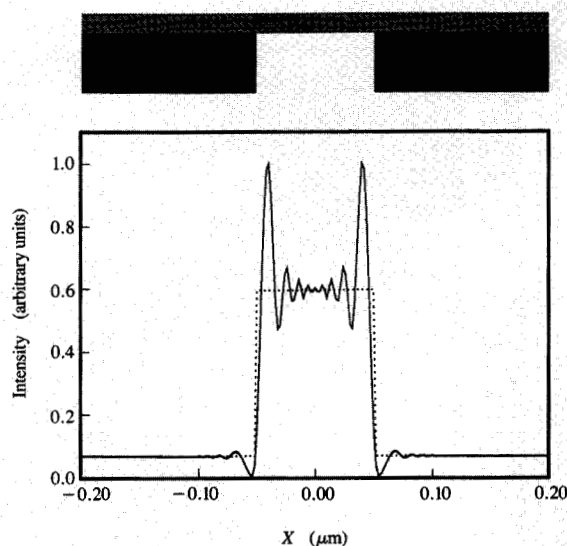


Figure 3

Comparison of two boundary patterns calculated with two different methods at the plane immediately behind the absorber. Solid line is from BPM, dashed line is from plane wave propagation. Au thickness 0.25 μm ; a single wavelength $\lambda = 1$ nm is used in the calculation.

the two calculations, or, in other words, the effect of wave-guiding on lithography, diminishes. To consider the fringing effect together with the partial coherence, the absorber can be approximated with a complex transmission function $T(x, y, \nu)$ calculated with BPM with normal incidence wave (or with the central ray in the case of oblique beam illumination). Because the blur angle seldom exceeds 10 mrad, the error caused by this approximation is negligible.

Physical effects

Modeling X-ray lithography, in the end, serves to define, or at least reduce, the parameters of the system and, more importantly, to optimize the system. The effects of various factors involved in the imaging process and optimization procedures are discussed here.

• Synchrotron radiation source

In the simulation of image formation, two goals are generally considered. One is to find the image intensity distribution as the input for the dissolution simulation of the resist profile. The other is to study the exposure latitude and the depth of focus for a given set of conditions. In either case, a uniform illumination can be

assumed. The illumination variation can be taken into account in the exposure latitude. Besides, only a small field (several microns) can be simulated because of the limitation of computational power. The task for modeling the source and illumination system is to find the power spectrum and spatial coherence. The illumination uniformity is left for a ray-tracing method such as SHADOW [23].

Synchrotron source spectrum

Synchrotron radiation is generated when an electron is accelerated along a circular path with a speed close to the speed of light. The bending of the path is realized with a Lorentz force provided by magnetic dipoles, so the radiation is also called bending magnet radiation. In the moving electron frame, the radiation looks like dipole radiation with the frequency given by Larmor frequency. When transformed to the laboratory frame, the Lorentz transformation shifts the radiation to much higher frequencies (blue shift). Let λ_c be the critical wavelength corresponding to the median wavelength in the radiation power spectrum given by

$$\lambda_c = \frac{4\pi\rho}{3\gamma^3} = 186/BE^2 \quad (\text{\AA}). \quad (12)$$

In the frequency range $\lambda - \lambda_c$, the power spectral density function is approximated [19] by

$$P(\lambda) = \frac{\partial\Phi}{\partial\lambda} = P_0 \sqrt{\frac{\lambda_c}{\lambda}} e^{-2\lambda_c/\lambda}, \quad (13)$$

where P_0 is a constant. The angular distribution of the radiation also is modified profoundly. At low electron velocities the radiation is emitted along the pattern typical of a radiating dipole, i.e., proportionally to $\cos^2 \theta$, where θ is the opening angle of the radiation measured from the electron velocity. At higher electron velocities, the lines of constant field are bent forward along the electron velocity to the point of assuming a strongly forward shape when $v \approx c$. The narrow cone has width: $\langle \theta^2 \rangle^{1/2} = 1/\gamma$, where $\gamma = E/m_0c^2 = 1957.0 E$ (GeV). The cone width is typically less than 0.5 mrad.

Synchrotron source brightness

When the light frequency is around the critical frequency, the angular distribution of this cone is approximated as a Gaussian [19, 24],

$$\begin{aligned} \xi(s_x, s_y, \nu) &= \frac{\partial^2\Phi}{\partial\nu\partial\Omega} \\ &= \xi_0 e^{-3\gamma^2\theta^2\lambda_c/\lambda} = \xi_0 \exp\left(-\frac{1}{2} \frac{s_x^2 + s_y^2}{\sigma_s^2}\right), \end{aligned} \quad (14)$$

where ξ_0 is a constant, Ω is a solid angle, $\sigma_s^2 = \lambda/(6\gamma^2\lambda_c)$,

$\theta^2 \approx \sin^2 \theta = 1 - \cos^2 \theta = s_x^2 + s_y^2$. In the orbit plane, the circular motion of the electrons sweeps a large angle. This sweeping essentially gives uniform angular distribution (s_y) in the horizontal direction, which is eventually limited by a slit in applications. The limit of s_y is called the acceptance angle, which corresponds to a direction cosine denoted by s_0 . The electron beam in the storage ring has a Gaussian distribution in both density and momentum and also is best described by a phase space quantity B_0 . Since the electrons radiate independently, the total brightness of the source B_s is obtained by summing up brightness B_e from all the radiating electrons:

$$B_s(x, y, s_x, s_y; \nu) = B_e(s_x, s_y; \nu) \otimes B_0(x, y, s_x, s_y). \quad (15)$$

The brightness in the equivalent source plane can be approximated by

$$\begin{aligned} B_s(x, y, s_x, s_y; \nu) &= \Phi_0 \exp\left\{-\frac{1}{2} \left[\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{s_x}{\sigma_{sx}}\right)^2\right]\right\} \\ &\times \exp\left[-\frac{1}{2} \left(\frac{y}{\sigma_y}\right)^2\right] \text{rect}\left(\frac{s_y}{s_0}\right), \end{aligned} \quad (16)$$

where Φ_0 is a constant, $\text{rect}(x/b) = 1$ when $|x| < b$ and $\text{rect}(x/b) = 0$ otherwise. Typical values are $\sigma_x = 0.5$ mm, $\sigma_y = 0.5$ mm, and $\sigma_{sx} = 0.5$ mrad. It is important to point out that σ_{sx} is a function of ν , where different wavelengths have different divergence angles in the vertical direction.

• Illumination optics for a synchrotron-radiation-based system

The function of the illumination system is to deliver light into the imaging system and provide the required illumination uniformity, spectral bandwidth, and spatial coherence. For X-ray proximity lithography, the wavelength should not be longer than 16 Å to keep the high resolution; for the short-wavelength part, light with a wavelength shorter than 6 Å should also be eliminated because 1) photoresists have low absorption coefficients, 2) it causes heat in the mask and may cause device damage, and 3) for nanostructure fabrication, photoelectron scattering becomes a major concern for features of 70 nm and beyond with energetic photons (energy higher than 2 keV). Synchrotron radiation has a broad-band spectrum. The long-wavelength light is usually cut off by a window fabricated from a Be membrane 10–25 μm thick. A 2-μm-thick Si membrane is often used to cut shorter-wavelength soft X-rays because of the Si edge near 1900 eV. Hard X-rays are lost in the grazing-incidence mirror because they do not satisfy critical angle conditions. Uniformity of illumination had been the only concern in designing an X-ray lithography illumination system [25] until the issue of spatial coherence was raised [11]. Both the uniformity and partial coherence are important in

image formation and so also should be in system design. There are several types of practical exposure systems for X-ray lithography, but there is a general consensus that one or more mirrors should be used, if for no other reason than to reduce the amount of nonimaging hard radiation delivered to the mask-wafer assembly. A further distinction is based on the use of condensing or noncondensing optics. The first corresponds to an increase in the angle of acceptance of the beamline, thus providing power gain, while the others merely relay the beam to the mask-wafer assembly. Finally, a distinction exists between full-field and scanning exposure, where the illumination is provided as a uniform distribution at the mask in the former or in the form of a vertically scanning line in the latter.

Scanning mirror system

A scanning system often uses different focusing in the horizontal and the vertical directions, forming an anamorphic image. The system is equivalent to two different lens systems pointing in different directions (Figure 4). As the figure shows, the two systems are really the two classical illumination systems. The horizontal is equivalent to a Kohler illumination, and the vertical is equivalent to a critical illumination. The nonuniformity of the critical illumination is eliminated by the scanning in the vertical direction. In the vertical direction, the aperture is usually large enough not to limit the light. For 1:1 imaging, the coherence in the mask plane is the same as the coherence in the source plane. The generalized brightness is then

$$B(x, y, s_x, s_y, z_m; \nu) = \Phi_0 \exp \left\{ -\frac{1}{2} \left[\left(\frac{x}{\sigma_x} \right)^2 + \left(\frac{s_x}{\sigma_{sx}} \right)^2 \right] \right\} \times \text{rect} \left(\frac{y}{s_0 f_c} \right) \exp \left[-\frac{1}{2} \left(\frac{s_y}{\sigma_y f_c} \right)^2 \right]. \quad (17)$$

The angular divergence of the brightness is equivalent to spatial coherence; the smaller the divergence angle, the more coherent the light. The horizontal direction has a uniform power spectral density and is highly coherent, since the blur angle is σ_y/f_c , which is typically 0.1 mrad. The divergence in the vertical direction is the same as in the source plane (in the range of 0.1–1.0 mrad). The uniform illumination in the vertical direction is obtained by scanning the mirror in the vertical direction. The equivalent brightness is an integration of the brightness given in (17) over the scanning angle. Usually a small overscan is needed to ensure uniformity at the edge of the field. For the simplicity of mathematical treatment, we use an infinite scan range. Reference [8] treated the case in which the source is the focal point and the beam diverges to the mask, whereas in most current designs the beam is focused to the mask plane and the generalized brightness is

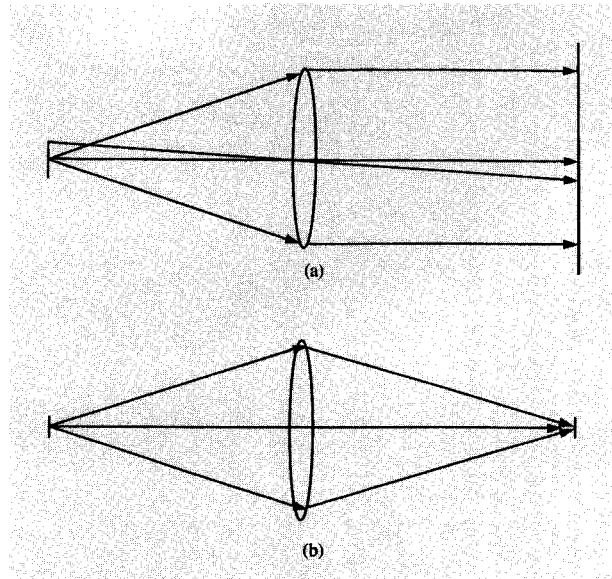


Figure 4

Equivalent optical diagram for scanning illumination system: (a) horizontal direction; (b) vertical direction.

given as

$$B(x, y, s_x, s_y, z_m; \nu) = \Phi_0 \text{rect} \left(\frac{y}{s_0 f_c} \right) \exp \left[-\frac{1}{2} \left(\frac{s_y}{\sigma_y f_c} \right)^2 \right] \times \exp \left[-\frac{1}{2} \left(\frac{s_x}{\sigma'_{sx}} \right)^2 \right], \quad (18)$$

where σ'_{sx} is the new divergence angle in the vertical direction,

$$\sigma'_{sx} = \sqrt{\sigma_{sx}^2 + (\sigma_x/q)^2}, \quad (19)$$

where q is the distance between the mirror and the mask. In the horizontal direction,

$$\sigma'_{sy} = \sigma_y/q. \quad (20)$$

For example, the IBM/Oxford storage ring Helios 1 in the vertical direction has $\sigma_x = 0.7$ mm, $\sigma_{sx} = 0.3$ mrad, and $q = 3.5$ m, so we get $\sigma'_{sx} = 0.36$ mrad in the vertical direction. Smaller q gives a larger blur angle, but the run-out error, given by $Dg/2q$, increases as q decreases. The scanning exposure system implemented at CXrL at the University of Wisconsin has been demonstrated to be versatile and flexible, and a power uniformity of better than $\pm 1\%$ has recently been measured [25]. Scanning the mask-wafer assembly is another choice, which gives the same divergence angle but no run-out error. However, it complicates stepper design.

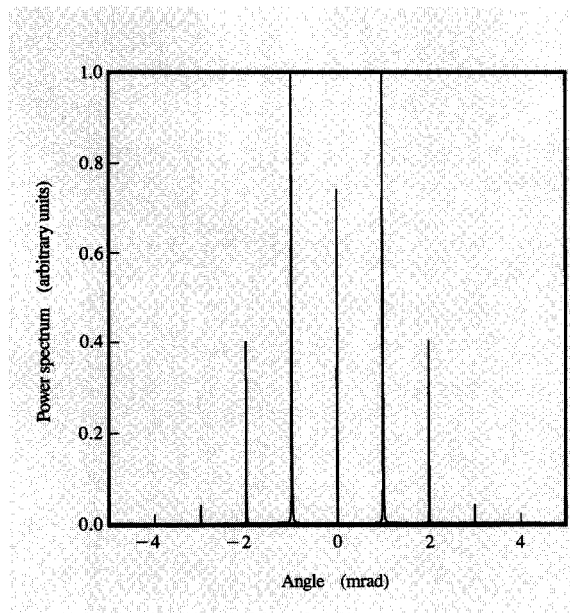


Figure 5

Virtual sources created by phase grating of pitch $1.0 \mu\text{m}$ and $1.0\text{-}\mu\text{m}$ -thick Si at wavelength $\lambda = 1 \text{ nm}$.

Full-field illumination

A full-field illumination system provides a uniform illumination of a whole field in the mask plane ($25 \times 25 \text{ cm}^2$) without lateral motion of either the mask-wafer assembly or the mirror. In the vertical direction, both the source intensity distribution and the source divergence distribution are Gaussian. Therefore, it is impossible to obtain uniform illumination by either imaging (source intensity distribution becomes mask plane intensity distribution) or collimation (source divergence distribution becomes mask plane intensity distribution). A novel approach based on a faceted mirror which is similar to a fly's-eye lens has been proposed to overcome the difficulty [26]. In this design, a two-dimensional array of toroidal facets are positioned on top of a base toroidal surface. Each facet illuminates the whole field, and the base toroid images the source onto the center of the field. The uniformity of this system is largely determined by the number of facets used. The partial coherence of this design is determined by the acceptance angle in the horizontal direction, the source divergence angle in the vertical direction, the source distance p , and the image distance q . In the vertical direction,

$$\sigma'_{\text{sx}} = \sigma_{\text{sx}} \frac{p}{q}. \quad (21)$$

In the horizontal direction, the angular distribution is

uniform within $\pm\alpha_y$, where $\pm\alpha_y$ is given by

$$\alpha_y = s_0 \frac{p}{q}. \quad (22)$$

The blur angles in the horizontal direction and in the vertical direction are different. Two mirrors are required to provide separate control over the coherence.

Coherence reduction

It is difficult to control the spatial coherence in the above designs. One method of achieving an optimum blur angle of 3–6 mrad is by using a random phase plate. A vibrating rough mirror also can reduce the coherence, but the correlation length, which determines the blur angle, is difficult to control. A transmitting phase grating, composed of a membrane of a transparent material, can be used effectively to reduce coherence. If the incident mutual spectral density function on the front surface z_f is

$$G\left(x + \frac{\Delta x}{2}, x - \frac{\Delta x}{2}, z_f; \nu\right) = \mu(\Delta x; \nu)P(x; \nu), \quad (23)$$

and the complex transmittance of the phase plate is given as $T(x) = e^{i\phi(x)}$, the mutual spectral density behind the moving phaser is given by

$$G\left(x + \frac{\Delta x}{2}, x - \frac{\Delta x}{2}, z_b; \nu\right) = \mu(\Delta x; \nu)P(x; \nu)C(\Delta x), \quad (24)$$

where the correlation of the phase plate, which can be characterized by a stationary stochastic process moving with a velocity of v , is

$$C(\Delta x) = \langle e^{i[\phi(x_1) - \phi(x_2)]} \rangle = \int e^{i[\phi(x + \Delta x + v\tau) - \phi(x)]} d\tau. \quad (25)$$

Because $C(\Delta x)$ is much narrower than $\mu(\Delta x)$ (the reason for introducing the diffuser), the coherence is basically determined by $C(\Delta x)$. The amount of light scattered is determined by the RMS variation of phase σ_ϕ . If $\sigma_\phi \sim \pi$, almost all of the light is scattered. The correlation C must be controlled so that the divergence of the beam matches the blur requirement. The phase plate does not have to be random, because a regular phase grating creates discrete virtual sources. As an example, assuming a sinusoidal phase grating with a pitch of $1.0 \mu\text{m}$ ($0.5 \mu\text{m}$ linewidth and $1.0 \mu\text{m}$ depth), two significant orders at 1 mrad and 2 mrad are created besides the zeroth order for $\lambda = 10 \text{ \AA}$, as shown in Figure 5. Because a lateral vibration faster than the recording time randomizes the phase between the orders, independent sources result. The natural vibration on the beamline is probably enough, since only $1 \mu\text{m}$ magnitude is needed. The fabrication of the phase grating in the X-ray regime can be based on Si membrane, with plasma-etched grooves $1.0 \mu\text{m}$ deep. Since a sinusoidal

profile gives a larger first-order diffraction than a vertical trench, no strict vertical profile control is necessary in etching the deep trench.

☛ *Plasma X-ray sources*

Because of their relative portability and low initial capital investment, plasma sources are explored as alternative light sources for X-ray lithography. Plasma X-ray sources differ from synchrotron sources in three respects in their application to X-ray lithography. First, the spectrum from a plasma source has a narrow-band wavelength, while synchrotron radiation is broad-band. Broad-band light has a smoothing effect on the diffraction pattern, which helps to produce a larger exposure window. Second, the central wavelength of a plasma source is longer, since it is difficult to obtain wavelengths shorter than 12 Å with adequate conversion efficiency. In terms of image formation, longer wavelengths mean a slightly reduced working distance because of diffraction. Third, a plasma source produces a diverging beam, while synchrotron radiation is collimated. With a noncollimated beam, run-out error is always present; this can be corrected by using a pre-distorted mask, or it can just be taken as a systematic error. More serious is the variation in run-out error when the gap changes, which is discussed later. Although the lateral source size of a plasma source is smaller than that of a synchrotron source, the spatial coherence in the mask plane is lower for plasma source because of its much shorter distance from the source to the mask. Lower spatial coherence is equivalent to larger blur, which is a positive factor in image formation.

Power spectrum and spatial coherence

We focus on laser-induced plasma sources only. The radiation of soft X-rays by plasma sources is basically black-body radiation. The efficient emission of wavelengths below 16 Å requires a plasma of very high density and temperature. The plasmas are produced by plasma focus, plasma pinch, or laser irradiation. Thus far, laser-induced plasma sources have been most successful. The laser intensity required to produce plasma electrons with a temperature around 600 eV is of the order of 10^{15} W/cm². The laser pulse width must be longer than 0.5×10^{-9} s to produce efficient emission in the 12–16-Å range [27]. The conversion efficiency is between 5% and 10%, depending on the targets. The spectra are determined by the choice of atomic species for ionization and the plasma temperature. Higher photon energies can be achieved by choosing heavier materials such as copper, but conversion efficiency, which determines throughput, may be adversely affected [28]. The spectral band of plasma sources is much narrower than that of synchrotron sources. For example, the Hampshire Instruments source with a stainless steel target has a triangular-shaped spectrum with central peak

14 Å and bandwidth 4 Å (FWHM).¹ A typical source size is 200 μm, determined by the laser spot size. There is currently no collimation system for a laser plasma source; this limits the distance from the source to the mask, because a greater distance means a smaller collecting angle for the same illumination area. If the distance from the source to the mask is 100 mm, the blur half angle is 1 mrad.

Effect of oblique beam on the image

To expose a 20×20 -mm² field with a source spot size of 200 μm and distance from source to mask of 100 mm, the edge of the field must have an obliquity angle δ of 100 mrad. The first problem that has arisen from this is the well-known run-out error given by

$$r = \delta g. \quad (26)$$

If all levels are exposed with X-ray lithography, a systematic error does not cause serious problems. If there is a gap variation between levels, the variation of run-out is given by

$$\Delta r = \delta \Delta g. \quad (27)$$

For $\Delta g = 2$ μm, $\delta = 1$ mrad and $\Delta r = 0.1$ μm, which is unacceptable for 0.25-μm lithography. The gap and the wafer topography must be controlled tightly; otherwise, collimation is necessary. Another question associated with oblique beam illumination is the so-called shadowing effect of the thick absorber layer on the mask. In the shadow of the edge of the absorber, the intensity distribution is asymmetric, as shown in **Figure 6**. The ringing on the right-hand side of the intensity pattern immediately behind the mask can be explained physically from the interference of incident light with light reflected from the wall (as in Lloyd's mirror interference). The asymmetry disappeared almost entirely after a path length of 20 μm (**Figure 6**). The difference in intensity profiles does not have a significant impact on linewidth, because the place where they differ is not at the nominal linewidth position.

Exposure window

The optimized thickness of the Au absorber to correspond to a π phase shift for the plasma source is 0.29 μm, which is thinner compared to 0.45 μm for synchrotron radiation because of the longer wavelength of the plasma source. **Figure 7** shows the exposure window of a plasma source with a blur angle of 4 mrad and various biases.

☛ *Exposure process*

The aerial image is not useful unless the energy is deposited in the photoresist, which results in a differential dissolution rate of the resist upon development. The

¹ J. Forsyth, private communication.

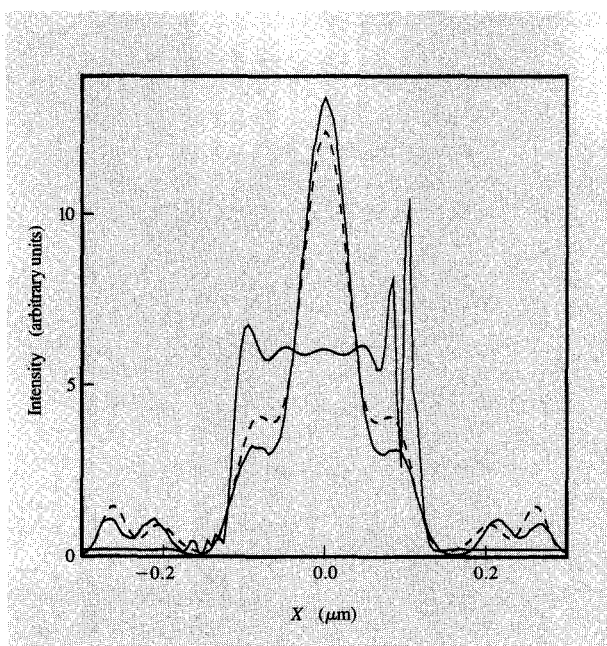


Figure 6

Difference between normal incidence and oblique beam illumination; Au thickness $0.3\ \mu\text{m}$, gap $15\ \mu\text{m}$, plasma source central wavelength $14\ \text{\AA}$. The asymmetric curve is the light intensity behind the absorber. The dotted line indicates normal incidence; the solid, oblique beam illumination. The pattern of the oblique beam illumination is shifted to the center for comparison.

energy deposition process is fairly well understood from the Monte Carlo simulation of electron scattering in photoresist [29, 30]. We do not go into detail about the latent image formation, but discuss only one phenomenon, photoelectron blurring, and its effects on overall resolution and exposure latitude.

The absorption of one energetic X-ray photon generates one high-energy photoelectron, together with one Auger electron, which are called primary electrons. The two primary electrons are scattered by the atoms 10–100 times before they come to rest. In the scattering process, elastic scattering changes only their directions, and the energy is transferred to secondary electrons only through inelastic scattering. In the simulation, the deflection is usually calculated on the basis of the elastic scattering cross section, and the energy is transferred through continuous slowing down (Bethe model) [29]. The simulation of electron scattering is a very slow process with the Monte Carlo method because of the large number of electrons involved. A simplified way of accounting for the smearing due to scattering is approximating the system as a linear shift-invariant system with aerial image (light intensity distribution) as the input and dose image (energy

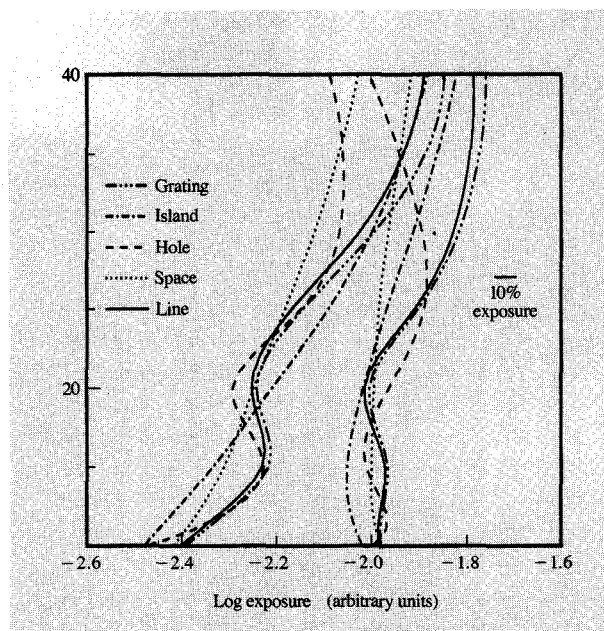


Figure 7

Log exposure-gap diagram for five $0.25\text{-}\mu\text{m}$ features for a plasma source; Au thickness $0.29\ \mu\text{m}$, sidewall slope 60/450, plasma source central wavelength $14\ \text{\AA}$. Nominal size on the mask is measured at the 50% thickness of absorber.

deposited) as the output. The point spread function can be computed by Monte Carlo techniques for a given resist and photon energy. A line spread function (averaged over the thickness of resist) of 2.29-keV photons in PMMA² is shown in **Figure 8**. The blur function is best fitted by three Gaussians with different weights. The sigmas are in the range of 2–40 nm for photo energies in the range of 700–2000 eV. Since a blur width of 1/2 to 1/3 of the feature dimension is helpful in obtaining large exposure windows, electron scattering does not become a significant blur source until feature sizes reach 70 nm. However, for features of that dimension, a less sensitive resist should be used to avoid shot noise effects and to obtain a statistical average.

Applications

• Process optimization

X-ray proximity lithography is well recognized as a lithography tool with high resolution and a large exposure window. Because of the large exposure window and depth of focus associated with X-ray lithography, the need for optimizing the system has largely been ignored. However,

² L. Ocola, private communication.

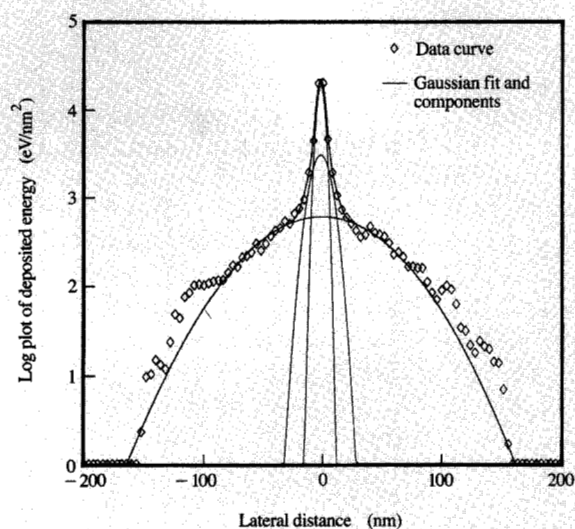


Figure 8

Line spread function of electron scattering in PMMA generated by 2.29-keV photons averaged over the thickness of the resist. The diamonds are simulation data, and the solid lines are three Gaussian curves and their weighted summation. The weights are 4.38, 0.68, and 0.14, corresponding to $\sigma = 2.68$ nm, 7.25 nm, and 39.84 nm.

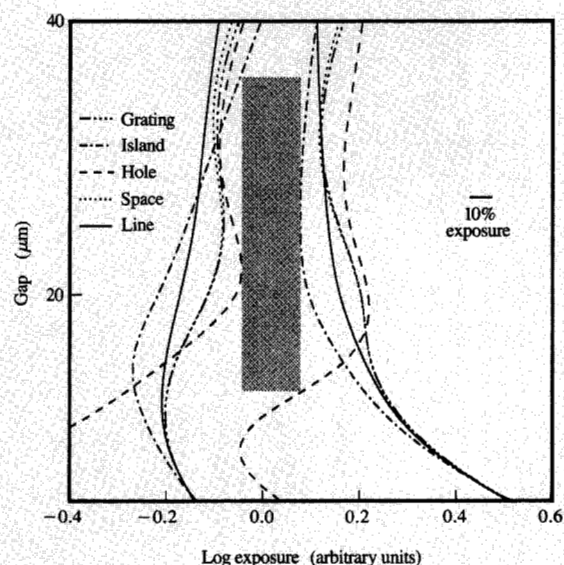


Figure 9

Log exposure-gap diagram for five 0.25- μ m features; synchrotron radiation source centered around 8.5 Å, Au thickness 0.45 μ m, sidewall slope 60/450. Contact hole biased +10 nm per edge, all other features biased +30 nm per edge. Nominal size on the mask is measured at the 50% thickness of absorber.

when critical dimensions shrink down to 0.25 μ m and smaller, the budget for each component of the lithography system becomes very small. For example, the $\pm 10\%$ linewidth variation tolerance allocated to pattern transfer also includes process variation of the resist. Optimizing the system in the image formation process would limit the budget to less controllable factors such as stochastic noise in the resist process. In proximity lithography, phase plays a very important role; the system transfer function is a pure phase function under coherent illumination, in contrast to a low-pass filter (only magnitude change) in the projection case. Proximity imaging is not diffraction-limited in the same sense as in projection lithography. The diffraction reduces image quality only when there is a gap between the mask and the wafer. The resolution is far from the diffraction limit of $\lambda/2 = 5$ Å. In projection lithography, reducing spatial coherence shifts the spatial frequency spectrum so that higher-frequency components get through the filter and thus enhance the resolution. In proximity lithography, lower spatial coherence is desired, not to enhance the resolution but to reduce spurious patterns in order to increase the depth of focus and even broaden the exposure latitude in some cases. An

attenuated phase shifter is a natural choice for proximity X-ray lithography because of the low absorption coefficient of metals in the soft X-ray regime. In X-ray lithography, π phase shift provided by low-contrast ($\sim 7:1$) masks produces artifacts which are basically contributed by the high-frequency components in the input field and can appear in the center of opaque features if no precautions are taken. One way of avoiding the ghost features is by using an absorber with a sloped sidewall which does not create unnecessary and even harmful high-frequency components while providing the required modulation. The slope also smoothes the intensity profile so that the exposure level does not change abruptly when the gap changes. Reduced spatial coherence in illumination also helps to smooth out the Poisson spot. When low-contrast masks and an absorber with a sloped sidewall are used, the exposure latitude at the nominal feature size is usually not the largest. Different types of features also show different diffraction effects, which may cause bias from nominal width of different types of features on the same mask level, because only one exposure dose can be chosen. It is necessary, then, to bias the features on the mask in advance to form a common exposure window. **Figure 9**

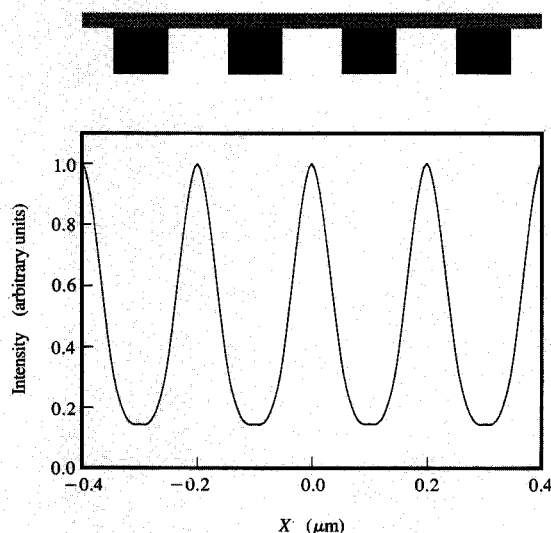


Figure 10

Intensity of line/space pairs; synchrotron source, absorber mask thickness 170 nm, gap 20 μm . Electron scattering is considered.

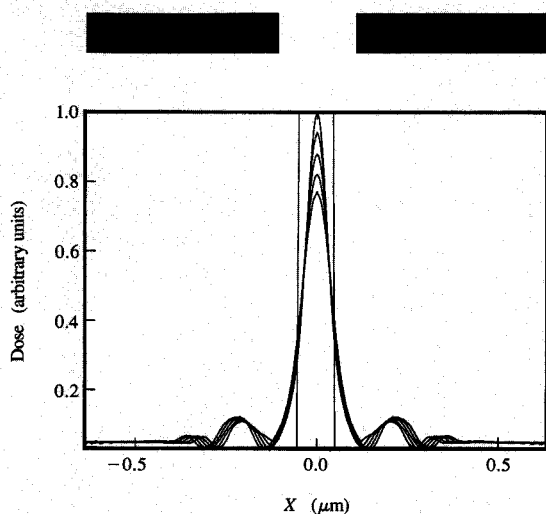


Figure 11

Reduction printing with proximity X-ray lithography. Only clear patterns can be printed this way; negative resist may be needed for printing lines. In the figure, the mask dimension is 0.2 μm and the printed pattern is 0.1 μm ; gap range 20–30 μm .

shows an optimized situation with partially coherent illumination, sloped sidewall of absorber, thickness corresponding to π phase shift, and biased mask.

• Phase-shifting mask

In the last section, we discussed attenuated phase-shifting masks which do not need special effort. There are other types of phase-shifting technology which can be used to improve depth of focus and especially working distance for feature sizes below 0.25 μm . As discussed previously, the resolution of X-ray proximity lithography is far from the diffraction limit. Even for features as small as 0.1 μm , the depth of focus is still very large if the gap is small enough. However, very small gaps may cause damage to the mask if particles are present because of the membrane structure of the mask. To increase the working distance, two kinds of phase-shifting techniques can be devised for particular features. One is the “super-resolution” phase mask; the other is a low-contrast mask with very thin Au absorber suitable for line/space pairs. We also discuss a clear phase-shifting mask which is very similar to chromeless phase masks used in optical lithography.

Very low-contrast mask

To print features of 0.1- μm size, gaps in the range of 10–15 μm are normally needed. However, if the phase shift between clear area and absorber area is reduced by using a very thin absorber thickness (170 nm Au), the proximity gap can actually be increased to 20 μm , as shown in Figure 10. However, this works only for lines/spaces.

“Super-resolution” mask

There is another way of printing high-resolution pattern at much larger distances than expected. In the Introduction, we pointed out that the diffraction shifts the intensity peak away from the nominal edge. The shift becomes larger as the gap increases, as shown by Equation (1). We can make use of this effect by printing clear features smaller than the nominal features on the mask by adjusting the exposure level. For a given wavelength, feature size, and gap, the opening on the mask can be obtained by solving a reverse problem for a given system. The resulting pattern is shown in Figure 11. The exposure window of such a mask is shown in Figure 12. The mask size is twice the size of the printed pattern, so that a reduction is realized. This kind of mask can be very useful in printing quantum devices which are not densely packed.

“Clear” phase-shifting mask

The clear phase-shifting mask is very popular in optical projection lithography (chromeless PSM) to extend the resolution limit of optical lithography. In the X-ray regime, the same phase-shifting technology can also be used to

print a line at the transition position from 0 to π phase shift [31]. The phase shift of π in X-ray lithography is more difficult to obtain because of the small α in electric permittivity ϵ , as was discussed in the Introduction. The typical thickness needed to obtain a π phase shift for light materials such as Si or PMMA is 1.5–3 μm . Maintaining the edge acuity of the shifting layer is very important because the diffraction pattern depends on the sharp transition of the phase from one region to another. An example of the diffraction pattern with 1.5 μm Si as the shifting layer is shown in Figure 13. The broad-band nature of synchrotron light would give a range of phase shift for shifting layer of the same thickness; however, the effect is only a small blurring in the edge because Be and Si membrane windows limit the spectrum to within a range of 1200 eV. The diffraction pattern is not symmetric because of the absorption in the Si shifting layer; the nominal center of the line should be shifted from the edge in advance to avoid an overlay error.

Summary

The modeling of image formation has come a long way in the last couple of years. A much clearer picture of the imaging capability of proximity X-ray lithography has emerged from the simulation work done at CXrL [11–13] and by other groups [14, 15, 18]. As for the present models, most factors known to have an effect on image formation have been taken into account. Up to the latent image formation, the models are based on quite rigorous physical principles. However, the ultimate verification is still by means of controlled experiments. The absolute measurement of image is not possible because of the lack of availability of high-resolution linear detecting devices. The modeling results and experimental results generally agree with one another in terms of the exposure latitude and depth of focus.

Modeling has shown that the large exposure windows of X-ray lithography are available only when the system is optimized. Fortunately, many positive factors exist in current systems which were not realized, or were even taken as drawbacks; this is the reason for many good experimental results. Modeling also has pointed a way to the optimization of X-ray lithography. Special attention, however, should be paid to the optimization of the whole system instead of optimizing component by component.

The dissolution process of chemically amplified resists still needs more understanding, especially the acid diffusion process, which may have a significant effect, either positive or negative, on linewidth control.

Appendix A: Scalar diffraction formula

Fresnel-Kirchoff diffraction formulation By choosing a spherical Green function and Kirchoff boundary

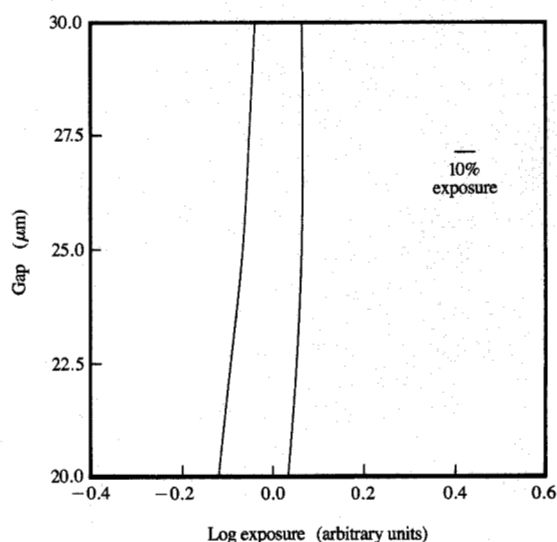


Figure 12

Log exposure-gap diagram for reduction printing of lines; synchrotron source, Au thickness 0.45 μm , blur 3 mrad.

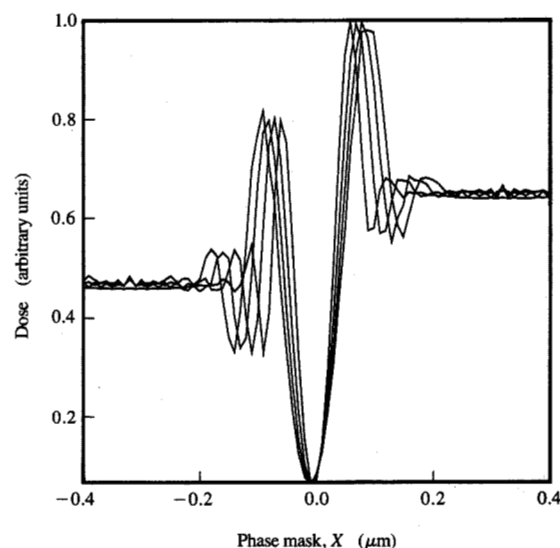


Figure 13

Clear phase-shifting mask with 1.5 μm Si as the shifting layer; gap range 5–12.5 μm .

conditions, the diffraction of incident light E_i can be described [32] as

$$E_o(\mathbf{r}) = \int \frac{1}{4\pi} \frac{e^{ikR}}{R} \left[\frac{\partial E_i(\mathbf{r}')}{\partial n} - \left(ik - \frac{1}{R} \right) \cos \theta E_i(\mathbf{r}') \right] d^2\mathbf{r}', \quad (28)$$

where $\hat{n}$ is the outward normal to the screen, $\mathbf{r}$ is the coordinate of the observation point, $\mathbf{r}'$ is the coordinate of the screen plane, $\partial E_i(\mathbf{r}')/\partial n$ is the directional derivative of E_i in the direction $\hat{n}$, and

$$\cos \theta = \cos(\hat{n}, \hat{\mathbf{R}}), \quad \mathbf{R} = \mathbf{r} - \mathbf{r}', \quad R = |\mathbf{R}|.$$

If we assume the incident light to be a plane wave propagating in the z direction, $A(x, y)e^{ikz}$, neglecting $1/R$ compared to $1/\lambda$ and assuming that the field and the derivative are zero in the shadow region, the system point spread function is

$$K(\mathbf{r}, \mathbf{r}') = \left(\frac{1}{i\lambda} \frac{\cos \theta + 1}{2} \right) \frac{\exp(ikR)}{R}. \quad (29)$$

Equation (28) is obtained by setting to zero both the field and its derivative on a plane behind the opaque screen (Kirchhoff boundary conditions), that leads to a trivial solution everywhere mathematically. This theory is not mathematically consistent.

Rayleigh-Sommerfeld diffraction formulation A better choice of Green function yields the following Rayleigh-Sommerfeld formula [32]:

$$E_o(\mathbf{r}) = \int E_i(\mathbf{r}') \frac{e^{ikR}}{R} \left(\frac{1}{i\lambda} - \frac{1}{2\pi R} \right) \cos \theta d\mathbf{r}'. \quad (30)$$

We can write the point spread function as

$$K(\mathbf{r}, \mathbf{r}') = \cos \theta \left(\frac{1}{i\lambda} - \frac{1}{2\pi R} \right) \frac{\exp(ikR)}{R}. \quad (31)$$

If we write (31) in the spatial frequency domain, a much simpler relation exists:

$$\mathcal{H}(f_x, f_y) = \exp(ig\sqrt{k^2 - k_x^2 - k_y^2}), \quad (32)$$

where $k_x = 2\pi f_x$, $k_y = 2\pi f_y$, and g is the gap between the object plane (mask) and image plane (wafer). It is very important to compare (32) with the system transfer function of a projection imaging system of numerical aperture NA at a wavelength λ :

$$\mathcal{H}(f_x, f_y) = \text{Circ} \left(\frac{\sqrt{f_x^2 + f_y^2}}{\text{NA}/\lambda} \right), \quad (33)$$

where $\text{Circ}(r/a) = 1$ when radius $r < a$ and $\text{Circ}(r/a) = 0$ otherwise. In proximity imaging, the system alters the phase only; in projection systems, the transfer function is a low-pass filter; thus, it is called diffraction-limited.

Appendix B: Beam propagation method

Let E denote the electric field of monochromatic light; the propagation in a region of weakly varying complex refractive index $n(\mathbf{r})$ is described by the Helmholtz equation

$$\nabla^2 E + k^2 n^2(\mathbf{r}) E = 0, \quad (34)$$

where $\mathbf{r}$ is the coordinate and k is the wave number. Given $n^2 = n_0^2 + \Delta n^2$, we can write $E(\mathbf{r}) = \psi(\mathbf{r})e^{\Gamma(\mathbf{r})}$, where $\Gamma(\mathbf{r})$ is a slowly varying phase modification term and ψ satisfies

$$\nabla^2 \psi + k^2 n_0^2(\mathbf{r}) \psi = 0. \quad (35)$$

The term n_0 is chosen such that the propagation of ψ along z can be obtained easily. In X-ray lithography we choose a constant n_0 to be the index of refraction of the vacuum in order to obtain the solution

$$E(x, y, z_0 + \Delta z) = \psi(x, y, z_0 + \Delta z) \exp \left(ik \frac{\Delta n^2}{2n_0} \Delta z \right). \quad (36)$$

The propagation of ψ is simply a summation of plane wave propagation,

$$\psi(x, y, z_0 + \Delta z) = \int A(k_x, k_y) \exp[i(k_x x + k_y y)] \times \exp(in_0 k_z \Delta z) \frac{1}{4\pi^2} dk_x dk_y, \quad (37)$$

where $A(k_x, k_y)$ is a decomposition of the total field of the previous step,

$$A(k_x, k_y) = \int E(x, y, z_0) \exp[-i(k_x x + k_y y)] dx dy. \quad (38)$$

The above solution can be realized quite easily with a numerical method by simply replacing the Fourier transform with the fast Fourier transform (FFT). The error of the above method is of third order in Δz .

Appendix C: Modal expansion

The idea of expanding partially coherent light into channels of coherent light can be carried out in a continuous way. The process involves finding an equivalent incoherent source, each source point of which emits spatially coherent light (a mode). Integrating over the whole source surface incoherently (summing up the intensity of each mode) gives the total intensity.

Expand $\mu(\Delta x)$ from (8) in a Fourier integral

$$\mu(\Delta x) = \int \xi(s) e^{iks\Delta x} ds, \quad (39)$$

where $\Delta x = x_1 - x_2$; then, $\xi(s)$ is equivalent to the angular divergence part of spectral brightness $B(x, s_x; \nu)$ when the illumination is uniform or slowly varying. By

substituting $\mu(\Delta x)$ in (9) and simplifying the result, we obtain

$$P(x, z_w; \nu) = \int \xi(s) \left| \int_{\text{mask}} T(x') A(x') e^{ikx'x} K(x - x') dx' \right|^2 ds, \quad (40)$$

where $\xi(s)$ is the effective source intensity far from the mask at an angle s , and the diffraction of a coherent mode is given by a convolution,

$$V_p(x, s) = \int_{\text{mask}} T(x') A(x') e^{ikx'x} K(x - x') dx'. \quad (41)$$

Let N_p be the normalized light intensity (dimensionless) of coherent diffraction due to a single source point (a mode),

$$N_p(x, s) = |V_p(x, s)|^2. \quad (42)$$

The image spectral density function at the wafer plane is an incoherent summation of all the mode intensities:

$$P(x, z_w; \nu) = \int \xi(s) N_p(x, s) ds. \quad (43)$$

It is important to analyze this result. From (41) and (42) we can see that the system is composed of *two cascaded linear systems* with a *square operation* after the first one. The first linear system encountered in the computation for V_p (41) is a *linear shift-invariant* (LSI) system, so that an FFT algorithm can be used. The second one (42) is not, since $N_p(x, s)$ does *not* represent a convolution. Direct numerical integration must be performed in order to obtain the exact results, which may be very time-consuming.

In computer simulation, only a finite area can be computed, so a discrete Fourier transform has to be used. In using the discrete Fourier transform of $\mu(\Delta x)$, periodicity is assumed [33]. Let the field size be $[-D/2, D/2]$; $\mu(\Delta x)$ is defined within $[-D, D]$. We can assume a periodic μ with a period $2D$ without changing its value within $[-D, D]$. Let $s_n = n\lambda/2D$; then,

$$\mu(\Delta x) = \sum_{n=0}^{\infty} \xi(s_n) e^{iks_n \Delta x} \quad (44)$$

and

$$V_p(x, s_n) = [T(x) A(x) e^{iks_n x}] \otimes K(x), \quad (45)$$

where $\otimes$ denotes convolution. The coherent diffraction can be calculated with an FFT method in the frequency domain, resulting in much faster computation. In the frequency domain, the Fourier transform of the complex field amplitude $U_0(x) = T(x) A(x)$, denoted by $\mathcal{U}_0(f)$, needs only to be shifted for different modes, thus saving about one half of the computation:

$$\mathcal{V}_p(f, s_n) = \mathcal{U}_0\left(f - \frac{s_n}{\lambda}\right) \mathcal{H}(f). \quad (46)$$

The inverse FFT must be done for each mode. The final spectral density function is given as

$$P(x, z_w; \nu) = \sum_n \xi(s_n) \left| \int \mathcal{V}_p(f, s_n) e^{i2\pi f x} df \right|^2. \quad (47)$$

Appendix D: Double linear shift-invariant system approximation

With light diffraction in the Fresnel zone, Equation (40) can be further simplified. If we retain only the second power in the expansion of distance r (gap much larger than the feature size) and neglect the obliquity factor $\cos \theta$,

$$r = \sqrt{g^2 + (x - x')^2} \approx g + \frac{1}{2g} [x^2 + (x')^2 - 2xx'],$$

the intensity field amplitude illuminated by a point source is simplified to

$$V_p(x, s) = \exp\left(ik \frac{x^2}{2g}\right) \int_{\text{mask}} T(x_m) A(x_m) \times \exp\left[ikx' \left(\frac{x}{g} - s\right)\right] \exp\left[ik \frac{(x')^2}{2g}\right] dx'. \quad (48)$$

If we scale s according to gs and use U_p to denote the integral over the mask plane, we can write

$$V_p(x, s) = \exp\left(ik \frac{x^2}{2g}\right) U_p(x - gs). \quad (49)$$

The complex amplitude V_p is not shift-invariant with respect to the source because of the phase term, but the intensity is a shift-invariant system with a point spread function

$$N_p(x - gs) = |U_p(x - gs)|^2. \quad (50)$$

From (40) we have

$$P(x, z_w; \nu) = \int \xi(s) N_p(x - gs) ds = \xi\left(\frac{x}{g}\right) \otimes N_p(x). \quad (51)$$

Appendix E: Definition of generalized brightness

A useful quantity in studying synchrotron beam propagation, *generalized brightness (radiance)* on a plane normal to z , can be defined as the following [34, 35]:

$$B(x, y, s_x, s_y; \nu) = \frac{1}{\lambda^2} \int G\left(x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2}, x - \frac{\Delta x}{2}, y - \frac{\Delta y}{2}; \nu\right) \times e^{-ik(\Delta x s_x + \Delta y s_y)} d\Delta x d\Delta y, \quad (52)$$

where s_x and s_y are directional cosines defined as $s_x = \cos \alpha$, $s_y = \cos \beta$, $s_z = \cos \theta = (1 - s_x^2 - s_y^2)^{1/2}$, solid angle element $d\Omega = ds_x ds_y / ds_z$. B is the phase space representation of the power flux density per unit area per unit solid angle per unit frequency interval. The total power flux through this plane can be written as

$$\begin{aligned}\Phi(\nu) &= \iint B(x, y, s_x, s_y; \nu) \cos \theta dx dy ds_x ds_y d\nu \\ &= \iint B(x, y, s_x, s_y; \nu) \cos^2(\theta) dx dy d\Omega d\nu.\end{aligned}\quad (53)$$

Acknowledgments

This work was supported in part by the National Science Foundation under Contract No. ECS-892116 through the Center for X-ray Lithography at the University of Wisconsin-Madison. The Center for X-ray Lithography is supported by the Defense Advanced Research Projects Agency under Contract Nos. 00014-89-J-2017 and SEMATECH/SRC 88-MC-507.

RISC System/6000 is a registered trademark of International Business Machines Corporation.

References

1. D. L. Spears and H. I. Smith, "High-Resolution Pattern Replication Using Soft X-Rays," *Electron. Lett.* **8**, 102 (1972).
2. E. D. Palik, *Handbook of Optical Constants of Solids*, Academic Press, Inc., Orlando, FL, 1985.
3. W. Moreau, *Semiconductor Lithography*, Plenum Press, New York, 1988.
4. M. L. Schattenburg, K. Early, Y.-C. Ku, W. Chu, M. I. Shepard, S.-C. The, and H. I. Smith, "Fabrication and Testing of 0.1 μm Linewidth Microgap X-Ray Masks," *J. Vac. Sci. Technol. B* **8**, 1604 (1990).
5. A. Betz, K. Heinrich, A. Heuberger, H. Heuber, and H. Oertel, "Resolution Limits in X-Ray Lithography Calculated by Means of X-Ray Lithography Simulator XMAS," *J. Vac. Sci. Technol. B* **4**, 248 (1986).
6. M. Gentili, R. Kumer, L. Luciani, L. Grela, D. Plumb, and Q. Leonard, "0.1 μm X-Ray Mask Replication," *J. Vac. Sci. Technol. B* **9**, 3319 (1991).
7. Y. C. Ku, E. H. Anderson, M. L. Schattenburg, and H. I. Smith, "Use of a Pi-Phase Shifting X-Ray Mask to Increase the Intensity Slope at Feature Edge," *J. Vac. Sci. Technol. B* **6**, 150 (1988).
8. D. So, B. Lai, G. Wells, and F. Cerrina, "The Effects of Beam Emittance on X-Ray Lithography Exposure Line Resolution," *J. Vac. Sci. Technol. A* **5**, 1537 (1987).
9. B. J. Lin, "A Comparison of Projection and Proximity Printings—from UV to X-Rays," *Microelectron. Eng.* **11**, 137 (1990).
10. A. Heuberger, "Comparison of Different X-Ray Sources: X-Ray Tubes, Laser Induced Plasma Sources, Compact and Conventional Storage Rings," *Proc. SPIE* **771**, 8 (1987).
11. J. Z. Y. Guo, G. Chen, M. Khan, P. Anderson, and F. Cerrina, "Aerial Image Formation in Synchrotron-Radiation-Based X-Ray Lithography: The Whole Picture," *J. Vac. Sci. Technol. B* **8**, 1537 (1990).
12. J. Z. Y. Guo and F. Cerrina, "Verification of Partially Coherent Light Diffraction Models in X-Ray Lithography," *J. Vac. Sci. Technol. B* **9**, 3207 (1991).
13. F. Cerrina and J. Z. Y. Guo, "Optimization of Partially Coherent Illumination in X-Ray Lithography," *Proc. SPIE* **1671**, 442 (1992).
14. Y. Vladimirovsky and J. R. Maldonado, "Illumination Effects on Image Formation in X-Ray Proximity Printing," *Microelectron. Eng.* **13**, 343 (1991).
15. H. K. Oertel, M. Weib, and H.-L. Huber, "Investigation of the Process Latitude for Sub-Half-Micron Pattern Replication in X-Ray Lithography," *Microelectron. Eng.* **13**, 339 (1991).
16. M. L. Schattenburg, K. Li, R. T. Shin, J. A. Kong, and H. I. Smith, "Electromagnetic Calculation of Soft X-Ray Diffraction from 0.1 μm Scale Gold Structure," *J. Vac. Sci. Technol. B* **9**, 3232 (1991).
17. J. Z. Y. Guo and F. Cerrina, "Experimental and Theoretical Study of Image Bias in X-Ray Lithography," *J. Vac. Sci. Technol. B* **10**, 3150 (1992).
18. S. Hector, H. I. Smith, and M. L. Schattenburg, "Modeling and Experimental Verification of Illumination and Diffraction Effects on Image Quality in X-Ray Lithography," *J. Vac. Sci. Technol. B* **10**, 3164 (1992).
19. W. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, Inc., New York, 1975.
20. J. Goodman, *Statistical Optics*, John Wiley & Sons, Inc., New York, 1985, p. 201.
21. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, New York, 1980, p. 534.
22. J. van Roey, J. van der Donk, and P. E. Lagasse, "Beam-Propagation Method: Analysis and Assessment," *J. Opt. Soc. Amer.* **71**, 803 (1981).
23. B. Lai and F. Cerrina, "SHADOW: A Synchrotron Radiation Ray Tracing Program," *Nucl. Instrum. & Meth. Phys. Res. A* **246**, 337 (1986).
24. K. Green, *Internal Report No. 50522*, Brookhaven National Laboratory, Upton, NY, 1977.
25. R. K. Cole and F. Cerrina, "Novel Toroidal Mirror Enhances X-Ray Lithography Beamline at CXRL," *Proc. SPIE* **1465**, 111 (1991).
26. R. J. Rosser and P. M. J. Wormell, "Saddle Toroid Arrays: Novel Grazing Incidence Optics for Synchrotron X-Ray Optics," Topical meeting on short wavelength coherent radiation: Generation and Application, N. Falmouth, MA, Sept. 26, 1988; *Proc. Opt. Soc. Amer.*, p. 325.
27. R. Hollman, "X-ray Lithography Using Broadband Sources," *J. Vac. Sci. Technol. B* **6**, 186 (1988).
28. A. L. Hoffman, G. F. Albrecht, and E. A. Crawford, "High Brightness Laser/Plasma Source for High Throughput Submicron X-Ray Lithography," *J. Vac. Sci. Technol. B* **3**, 258 (1985).
29. K. Murata and D. Kayser, "Monte Carlo Methods and Microlithography Simulation for Electron and X-Ray Beams," *Advances in Microelectronics and Electron Physics*, Vol. 69, Academic Press, Inc., New York, 1987, p. 175.
30. K. Murata, M. Kotera, K. Nagami, and S. Namba, "Monte Carlo Modeling of the Photo and Auger Electron Production in X-Ray Lithography with Synchrotron Radiation," *IEEE Trans. Electron Devices* **ED-32**, 1694 (1985).
31. V. White and F. Cerrina, "Metal-less X-Ray Phase Shifting Masks for Nanolithography," *J. Vac. Sci. Technol. B* **10**, 3141 (1992).
32. J. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Book Co., Inc., San Francisco, 1968, p. 54.
33. B. E. A. Saleh and M. Rabbani, "Simulation of Partially Coherent Imagery in the Space and Frequency Domain and by Modal Expansion," *Appl. Opt.* **21**, 2770 (1982).

34. A. Walther, "Radiometry and Coherence," *J. Opt. Soc. Amer.* **58**, 1256 (1968).
35. K. J. Kim, "Brightness, Coherence and Propagation Characteristics of Synchrotron Radiation," *Nucl. Instrum. & Meth. Phys. Res. A* **246**, 71 (1986).

Received July 27, 1992; accepted for publication December 10, 1992

Jerry Z. Y. Guo *Center for X-ray Lithography and Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 3731 Schneider Drive, Stoughton, Wisconsin 53589 (jguo@thor.xraylith.wisc.edu).* Mr. Guo is currently a graduate student and expects to receive his Ph.D. from the University of Wisconsin-Madison in May 1993. He received his B.S. and M.S. degrees in optical engineering from Tsinghua University in Beijing in 1979 and 1984, respectively. Mr. Guo has been working on X-ray lithography modeling and system optimization, condenser optics design for synchrotron-based X-ray lithography, and experimental evaluation and optimization of exposure windows of X-ray lithography. His interests are in advanced lithography and optical systems. He has published over twenty papers.

Franco Cerrina *Center for X-ray Lithography and Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 3731 Schneider Drive, Stoughton, Wisconsin 53589 (cerrina@thor.xraylith.wisc.edu).* Dr. Cerrina is Professor of Electrical Engineering and Director of the Center for X-Ray Lithography at the University of Wisconsin-Madison. He received his Ph.D. degree from the University of Rome in 1974. Before joining the UW-Madison faculty in 1984, he worked as a member of the research staff at the Italian National Research Council and as a postdoctoral research associate at Montana State University. Dr. Cerrina's current interests are centered around novel applications of synchrotron radiation, specifically X-ray lithography, X-ray optics, X-ray microscopy, and X-ray-assisted processing. He is a Senior Member of IEEE, and a member of the American Vacuum Society, the American Optical Society, the American Physical Society, and the Optical Society of America.

