Jpn. J. Appl. Phys. Vol. 31 (1992) pp. 4110–4119
Part 1, No. 12B, December 1992

Derivation and Simulation of Higher Numerical Aperture Scalar Aerial Images

Daniel C. COLE, Eytan BAROUCH[1], Uwe HOLLERBACH[1]
and Steven A. ORSZAG[1]

IBM Corporation, Essex Junction, Vermont 05452-4299, USA
[1]Program in Applied and Computational Mathematics,
Princeton University, Princeton, New Jersey 08544-1000, USA

(Received July 17, 1992; accepted for publication October 17, 1992)

The application of scalar diffraction theory to a projection optics system is examined here for somewhat higher numerical aperture conditions upon removing the paraxial, or small angle, approximation that is typically made. A detailed derivation is given that notes the key physical assumptions originally contained in work by others on vector imaging theory. The asymptotic limit of large lenses and focal length sizes to object and image sizes is explicitly carried out, while keeping numerical apertures and magnification fixed. Numerical results of the resulting equations are presented for a variety of imaging conditions, including phase shift masks and modified illumination.

KEYWORDS: optics, aerial, image, numerical aperture, microlithography

§1. Introduction

The aerial image of a mask, or the light intensity in a plane of an optical projection system's image space, is a critically important quantity in microlithography for governing how well a developed photoresist structure replicates a mask design. The aerial image represents, not fully, but to a large extent, the optical information about the mask that enters into the photoresist structure. Provided that resist development technologies strive to achieve an energy threshold development model,[1] then examining only the aerial image can yield reasonable predictions on mask printability.

Of course, a full detailed calculation of the resist structure certainly needs to take into account other factors than the aerial image, such as the propagation of the projected light within the resist, the exposure of the resist, any subsequent thermal or other such processes, plus the dissolution of the resist. Indeed, much of the past and present work by some of us has aimed precisely at developing the detailed simulation capability for predicting two- and three-dimensional resist structures.[2–4] Nevertheless, despite this important qualification, and excluding the very high numerical aperture regime, we emphasize that the aerial image is still usually the single most important predictor for mask printability.

Besides being important in microlithography, the aerial image is also of theoretical interest in optics because it dictates how the intensity of light is distributed in the image space of a projection lens system. We shall concentrate here on using *scalar* diffraction theory to deduce the aerial image, as is usually done,[5–8] but where the paraxial approximation will not be imposed. The paraxial approximation assumes that the angles of the light rays to the optic axis are sufficiently small that small angle approximations can be made.[6,7] By *not* imposing this approximation, the full limits of the capabilities of scalar diffraction theory can be examined and contrasted to the more accurate physical description of *vector* diffraction theory.[8,9] We believe that this extension to scalar diffraction theory is not merely an academic exercise, since it should do a reasonably good job of extending the range of applicability of scalar diffraction theory to the next generation of microlithography projection systems.[10]

Our other aim for extending the limits of scalar diffraction theory is that many of the same methods and physical assumptions employed here also apply to vector diffraction theory.[11–13] Nevertheless, the mathematics is sufficiently simpler here that key areas are more clearly revealed where important physical assumptions enter the theory, and where future refinements and investigations might be made. In particular, we have used the scalar aerial image formulation presented here as a test vehicle for improving the algorithm for computing an aerial image, while recognizing that most of these algorithmic improvements can immediately be carried over to vector aerial image computations. The result has been a very significant speed-up in computation over traditional methods.

The applicable range of our nonparaxial scalar results should adequately extend the range of scalar diffraction theory for numerical apertures (NAs) up to about 0.6, and perhaps somewhat higher with some anticipated decrease in accuracy. Above this range, vector diffraction theory is clearly needed to account for polarization effects and the inherent coupling between the vector components of the electromagnetic fields.[14–16]

The present analysis accounts for typical differences up to 5%–15%, or so, from paraxial scalar aerial image calculations at NAs of about 0.6. These differences decrease as NA decreases. The effects of magnification are clearly revealed in our final scalar imaging equation. In terms of the complexity in numerical computation of our final nonparaxial equation, it is of the same order as the paraxial equation. This result provides incentive to researchers to use the nonparaxial formulation when $\mathrm{NA} \leq 0.6$, particularly when examining images out of focus.

As for the outline of this article, §2 contains our deriva-

tion of the scalar aerial image for an aplanatic lens projection system. Section 2 is divided into several parts that contain the main derivation, as well as discussions on the optics of the condenser, the normalization constant, and the paraxial limit for low NAs. In §3 we discuss several numerical results, while §4 contains concluding remarks.

§2. Nonparaxial Scalar Aerial Image

2.1 *General*

Let $V^{(r)}(\boldsymbol{x}; t)$ be one polarization component of the electric field associated with a quasi monochromatic optical disturbance at position $\boldsymbol{x}$ and at time t. Following the usual approach of scalar theory, we assume that each component can be treated independently, so that the coupling can be ignored between the electromagnetic field components resulting from a rigorous solution of Maxwell's equations.[8] For a sufficiently high numerical aperture, such as NA greater than about 0.6, this assumption is expected to break down.

Let $V^{(r)}(\boldsymbol{x}; t) = \mathrm{Re}\,[V(\boldsymbol{x}, t)]$. Assuming the light to have a narrow band in frequency, let $A(\boldsymbol{x}, t)$ represent the complex envelope of $V(\boldsymbol{x}, t)$, where

$$V(\boldsymbol{x}, t) \equiv A(\boldsymbol{x}, t)\,\mathrm{e}^{-i2\pi\bar{\nu}t}, \tag{1}$$

and $\bar{\nu}$ is the center frequency of the wave. Hence, $A(\boldsymbol{x}, t)$ will be a slowly varying function of time that manifests itself in the coherence properties of the projection system.

Consider Fig. 1. We will assume that the object to be imaged is a planar reticle, sufficiently thin that its effect on the incident field can be approximated by a complex transmission function $T(\boldsymbol{x}_O)$. Points with subscript o (i.e. $\boldsymbol{x}_O$, $\boldsymbol{x}'_O$, etc.) will be assumed to lie in a plane immediately to the left of this mask in Fig. 1. Likewise, points with subscripts E and X will be assumed to lie in the indicated entrance and exit pupil planes, respectively, while points with subscript I lie within the image space to the right of the exit pupil plane.

Assuming the optical system to be linear, as occurs with the Huygens-Fresnel principle, then we can write that

$$A(\boldsymbol{x}_I; t) = \iint_{-\infty}^{\infty} dx_O\, dy_O K(\boldsymbol{x}_I; \boldsymbol{x}_O) T(\boldsymbol{x}_O) A(\boldsymbol{x}_O; t-\delta), \tag{2}$$

where $K(\boldsymbol{x}_I; \boldsymbol{x}_O)$ is an amplitude transmission function from the mask to the image (i.e., an amplitude spread function or impulse response function), and δ is a time delay that depends on $\boldsymbol{x}_I$ and $\boldsymbol{x}_O$. (See ref. 7, pp. 287–305, for a more detailed understanding of δ.)

2.2 *Derivation of the amplitude transmission function*

Here we introduce a detailed derivation of our main result for the function $K(\boldsymbol{x}_I; \boldsymbol{x}_O)$. Assuming an aplanatic lens, ~~a low Strehl ratio~~, and conservation of energy, we obtain an expression that does not scale with magnification. It is fundamentally different than the paraxial result which remains invariant under magnification scaling.

Consider the case where the object and image sizes are assumed to be small, while the dimensions of the optical system are taken to be enormously larger than the wavelength of the light and of the object and image sizes. Thus, we will be interested here in the asymptotic limit where the size of the lenses, their focal lengths, and the distances between the lenses become quite large, while maintaining the required NAs and magnification of the optical projection system.

If $T(\boldsymbol{x}_O)$ vanishes in the object plane outside a small area $\mathscr{D}$ around $\boldsymbol{x}_O$, where it equals unity, then eq. (2) may be estimated by

$$K(\boldsymbol{x}_I; \boldsymbol{x}_O) \approx A(\boldsymbol{x}_I; t)/[A(\boldsymbol{x}_O; t-\delta)\mathscr{D}]. \tag{3}$$

We can deduce K by considering the special case where the complex amplitude A is independent of time. Moreover, let us assume that the normal to the wave front at the object is also approximately normal to the ob-

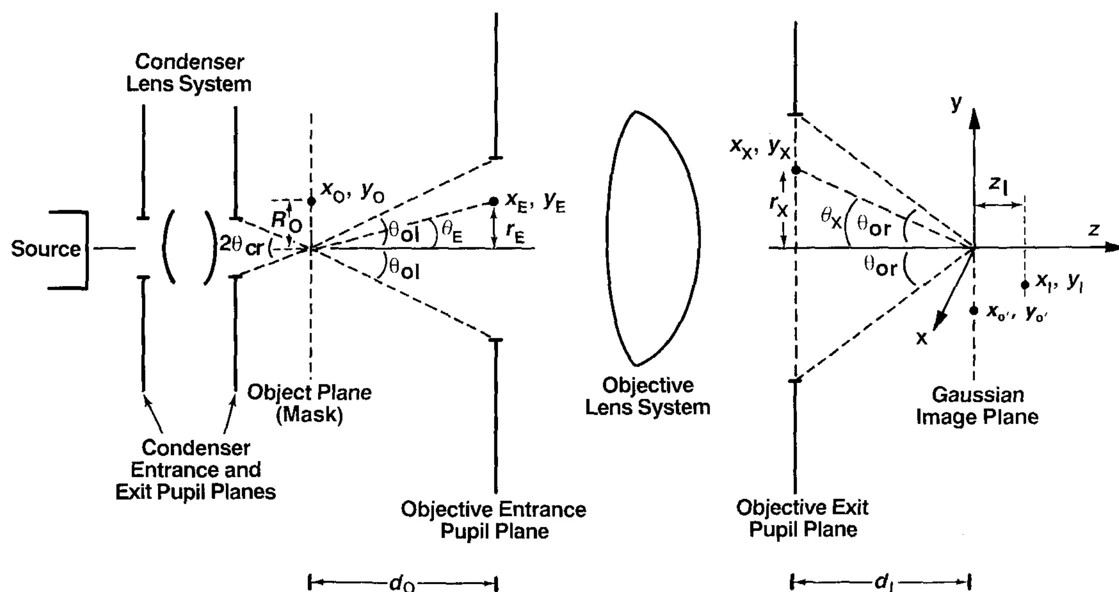

Fig. 1. Schematic outline of optical projection system.

ject plane, which will be a very good approximation if the numerical aperture of the condenser lens is small. (We comment on this point more later.) When A is independent of time, it satisfies the Helmholtz wave equation, since $V(\boldsymbol{x}, t)$ satisfies the wave equation. Consequently, we can use the integral theorem of Helmholtz and Kirchoff to deduce A at a point, knowing A and its normal derivative on a surrounding surface.

We will use three steps to deduce $A(\boldsymbol{x}_I)$ in eq. (3) in terms of $A(\boldsymbol{x}_O)\mathscr{D}$, thereby enabling $K(\boldsymbol{x}_I; \boldsymbol{x}_O)$ to be found. First, knowing A on the object plane, then A can be obtained in the plane of the entrance pupil. Assuming d_O in Fig. 1 to be large compared to the wavelength, and assuming the Rayleigh-Sommerfeld form of the diffraction formula due to an aperture (see, for example, ref. 6, pp. 44–45, and note that in the asymptotic regime, ref. 9, p. 249, indicates that this form of the diffraction formula yields essentially equivalent results to the usual application of Kirchoff boundary conditions), then

$$A(\boldsymbol{x}_E) = \frac{1}{i\lambda} \iint_{-\infty}^{\infty} dx'_O \, dy'_O T(\boldsymbol{x}'_O) A(\boldsymbol{x}'_O) \frac{e^{ikr'_{OE}}}{r'_{OE}} \cos(\Theta'_{OE})$$
$$\approx \frac{1}{i\lambda} \mathscr{D} A(\boldsymbol{x}_O) \frac{e^{ikr_{OE}}}{r_{OE}} \cos(\Theta_{OE}). \quad (4)$$

Here, $k = 2\pi/\lambda$, and r'_{OE} and r_{OE} are the distances between the object points $\boldsymbol{x}'_O$ and $\boldsymbol{x}_O$, and the point $\boldsymbol{x}_E$ in the entrance pupil plane, respectively, while Θ'_{OE} and Θ_{OE} represent the angles between these rays and the optic axis.

The second step expresses the amplitude A near the Gaussian image plane in terms of its value in the plane of the exit pupil:

$$A(\boldsymbol{x}_I) = \frac{1}{i\lambda} \iint_{-\infty}^{\infty} dx_X \, dy_X A(\boldsymbol{x}_X) \frac{e^{ikr_{XI}}}{r_{XI}} \cos(\Theta_{XI}), \quad (5)$$

where r_{XI} is the distance between $\boldsymbol{x}_X$ and $\boldsymbol{x}_I$, and Θ_{XI} is the angle between this ray and the optic axis.

Now comes the critical third step of this derivation. We need to determine $A(\boldsymbol{x}_X)$ in the exit pupil plane given $A(\boldsymbol{x}_E)$ in the entrance pupil. Once this relationship is established, then $K(\boldsymbol{x}_I, \boldsymbol{x}_O)$ can be deduced in eq. (3). Here is where key approximations are made that evidently cannot be improved upon (i.e., either further validated or falsified) without going through a detailed simulation of a specific optical system, such as by integrating over the individual surfaces of lenses within the objective lens system. For typical projection systems, this task would be an extremely impractical computational task.

The first key assumption was used by Wolf in the important work of ref. 11. Let us assume that

$$A(\boldsymbol{x}_X) = a(\boldsymbol{x}_X) \frac{e^{ik(\Phi - r_{XO'})}}{r_{XO'}}, \quad (6)$$

where $r_{XO'}$ is the distance from $\boldsymbol{x}_X$ to the Gaussian image point $\boldsymbol{x}_{O'}$ of $\boldsymbol{x}_O$ (see Fig. 1), $a(\boldsymbol{x}_X)$ is an amplitude that needs to be determined, and Φ is an aberration function due to a wave front converging on $\boldsymbol{x}_{O'}$ that is not spherical. Of course, if no assumptions are imposed on $a(\boldsymbol{x}_X)$ and Φ other than that they are real quantities with $a(\boldsymbol{x}_X) \geq 0$, then eq. (6) is perfectly general. However, once we identify Φ as an "aberration function," the implication is that for a "well corrected lens," then $\Phi = 0$. The result of this assumption is that the complex amplitude $A(\boldsymbol{x}_X)$ due to a point source at $\boldsymbol{x}_O$ is such as to give rise to a "spherical" wave front converging on the Gaussian image point $\boldsymbol{x}_{O'}$, but modified in amplitude due to the presence of $a(\boldsymbol{x}_X)$. The physical assumption introduced in eq. (6) then consists of the assumption about the meaning of a "well corrected lens": namely, that $\Phi = 0$.

The second key assumption enters upon deducing $a(\boldsymbol{x}_X)$ in eq. (6). Here, we will assume that the lens system is sufficiently corrected that the surfaces of the lenses obey the aplanatic condition. More specifically, we assume that if a ray trace is carried out through the system, starting along the ray at angle Θ_E from the optic axis, as indicated in Fig. 1, and ending along the ray at angle Θ_X in Fig. 1, then the surfaces of the lenses yield the following, due only to Snell's law dictating the directions of the lines traced through the surface interfaces:

$$\sin(\Theta_E) = M \sin(\Theta_X). \quad (7)$$

Here, M is the "magnification", or more precisely, the reduction given by M=(object height)/(image height).

Finally, to deduce $a(\boldsymbol{x}_X)$ we follow the assumption of Richards and Wolf in ref. 12, which was later used by Yeung in ref. 13: namely, that the intensity law of geometrical optics (§3.1.2 in ref. 8) holds for rays connecting the entrance and exit pupils. Thus, the assumption will be made that reflections from optical surfaces can be neglected. Again, only by a detailed calculation of a specific lens system can this factor be taken into account. The closer the "Strehl ratio" of a lens system is to unity, the more accurate this assumption will be in practice.

Consequently,

$$|A(\boldsymbol{x}_E)|^2 r_{OE}^2 \sin(\Theta_{OE}) \, d\Theta_{OE} = |A(\boldsymbol{x}_X)|^2 r_{XI}^2 \sin(\Theta_{XI}) \, d\Theta_{XI}. \quad (8)$$

Hence,

$$a(\boldsymbol{x}_X) = \frac{1}{\lambda} \mathscr{D} |A(\boldsymbol{x}_O)| \frac{r_{XO'}}{r_{XI}} \cos(\Theta_{OE}) \times \left| \frac{\sin(\Theta_{OE}) \, d\Theta_{OE}}{\sin(\Theta_{XI}) \, d\Theta_{XI}} \right|^{1/2}. \quad (9)$$

Equations (9) and (6) enable $A(\boldsymbol{x}_X)$ in eq. (5) to be determined. Substituting into eq. (3) then yields, aside from an unimportant constant phase factor,

$$K(\boldsymbol{x}_I; \boldsymbol{x}_O) = \frac{1}{\lambda^2} \iint_{S_X} dx_X \, dy_X \frac{1}{r_{XI}^2} \cos(\Theta_{OE}) \times \left| \frac{\sin(\Theta_{OE}) \, d\Theta_{OE}}{\sin(\Theta_{XI}) \, d\Theta_{XI}} \right|^{1/2} e^{ik(\Phi - r_{XO'} + r_{XI})} \cos(\Theta_{XI}), \quad (10)$$

where S_X is the area in the geometric optical opening of the exit pupil plane.

We can now substitute in the following asymptotic expansion (see Fig. 1):

$$r_{XO'}=[(x_{O'}-x_X)^2+(y_{O'}-y_X)^2+d_I^2]^{1/2}$$
$$\approx d_I[1+(r_X/d_I)^2]^{1/2}+\frac{-\frac{x_X x_{O'}}{d_I}-\frac{y_X y_{O'}}{d_I}}{[1+(r_X/d_I)^2]^{1/2}}, \tag{11}$$

where, from Gaussian optics, $x_{O'}=-x_O/M$ and $y_{O'}=-y_O/M$. Similarly,

$$r_{XI}=[(x_I-x_X)^2+(y_I-y_X)^2+(d_I+z_I)^2]^{1/2}$$
$$\approx d_I[1+(r_X/d_I)^2]^{1/2}+\frac{z_I-\frac{x_X x_I}{d_I}-\frac{y_X y_I}{d_I}}{[1+(r_X/d_I)^2]^{1/2}}, \tag{12}$$

where z_I is the displacement from the best focus position according to Gaussian optics. Hence,

$$-r_{XO'}+r_{XI}\approx\frac{-\frac{x_X x_O}{d_I M}-\frac{y_X y_O}{d_I M}}{[1+(r_X/d_I)^2]^{1/2}}+\frac{z_I-\frac{x_X x_I}{d_I}-\frac{y_X y_I}{d_I}}{[1+(r_X/d_I)^2]^{1/2}}, \tag{13}$$

which enters the argument of the exponential in eq. (10).

As for the other quantities in eq. (10), we note that, for example,

$$\cos(\Theta_{OE})\equiv d_O/[d_O^2+(r_E-R_O)^2]^{1/2}$$
$$\rightarrow d_O/[d_O^2+r_E]^{1/2}\equiv\cos(\Theta_E), \tag{14}$$

in the limit where d_O and r_E in Fig. 1 are scaled to ∞, while R_O is held fixed. Consequently, in Fig. 1 the angle Θ_E between the line connecting the point of intersection of the horizontal axis and the object plane to the point $\boldsymbol{x}_E$ is equivalent, in the indicated limit of $d_O\rightarrow\infty$, to the angle Θ_{OE} of the ray from $\boldsymbol{x}_O$ to $\boldsymbol{x}_E$. Similarly, angles Θ_{XI} and Θ_X are equivalent in the limit of $d_I\rightarrow\infty$, where d_O and d_I are scaled proportionately. Having noted this equivalence, the aplanatic condition of eq. (7) can be used to express quantities in eq. (10) in terms of $\sin(\Theta_I)$.

Letting $r_X/d_I=\rho$, $x_X/d_I=\rho\cos(\Theta)$, and $y_X/d_I=\rho\sin(\Theta)$, and restricting Θ_E and Θ_I to lie between $-\pi/2$ and $\pi/2$, we obtain that:

$$|\sin(\Theta_I)|=\rho/[1+\rho^2]^{1/2}, \tag{15}$$

$$\cos(\Theta_I)=\frac{1}{(1+\rho^2)^{1/2}}, \tag{16}$$

$$\cos(\Theta_E)=\left[1-\frac{1}{M^2}\sin^2(\Theta_I)\right]^{1/2}=\frac{(1+\rho^2-\rho^2/M^2)^{1/2}}{(1+\rho^2)^{1/2}}, \tag{17}$$

$$\left[\frac{\sin(\Theta_O)\,d\Theta_O}{\sin(\Theta_I)\,d\Theta_I}\right]^{1/2}=\frac{1}{M}\left[\frac{1-\sin^2(\Theta_I)}{1-\frac{1}{M^2}\sin^2(\Theta_I)}\right]^{1/4}$$
$$=\frac{1}{[1+\rho^2-\rho^2/M^2]^{1/4}}. \tag{18}$$

Let N_{or} and N_{ol} represent the NA on the right- and left-hand sides of the objective lens in Fig. 1, respectively, and similarly let N_{cr} and N_{cl} be the corresponding NAs for the condenser lens. Hence, $N_{or}=\sin\Theta_{or}$, $N_{ol}=\sin\Theta_{ol}$, and similarly for N_{cr} and N_{cl}. Thus,

$$K(\boldsymbol{x}_I;\boldsymbol{x}_O)\approx\int_0^{N_{or}/(1-N_{or}^2)^{1/2}}\rho\,d\rho\int_0^{2\pi}d\Theta$$
$$\times\exp\left[ik\left(\Phi+\frac{-x_O\rho\cos(\Theta)-y_O\rho\sin(\Theta)}{M(1+\rho^2)^{1/2}}+\frac{z_I-x_I\rho\cos(\Theta)-y_I\rho\sin(\Theta)}{(1+\rho^2)^{1/2}}\right)\right]\times F, \tag{19}$$

where

$$F=\frac{1}{\lambda^2}d_I^2\frac{1}{r_{XI}^2}\cos(\Theta_{OE})\cos(\Theta_{XI})\left[\frac{\sin(\Theta_{OE})\,d\Theta_{OE}}{\sin(\Theta_{XI})\,d\Theta_{XI}}\right]^{1/2}$$
$$\approx\frac{1}{\lambda^2}\frac{1}{M}\frac{(1+\rho^2-\rho^2/M^2)^{1/4}}{(1+\rho^2)^2}. \tag{20}$$

The above result is exact in the asymptotic limit of $d_I\rightarrow\infty$. Finally, we can express this result in a form more familiar to optical researchers by the following change of variables to direction cosines. Here, $\hat{s}$ is a unit vector that points in the direction from the point in the Gaussian image plane on the optic axis to a point on the exit pupil plane:

$$s_x\equiv\frac{\rho\cos\Theta}{(1+\rho^2)^{1/2}},\quad s_y\equiv\frac{\rho\sin\Theta}{(1+\rho^2)^{1/2}},\quad s_z=-(1-s_x^2-s_y^2)^{1/2}. \tag{21}$$

We finally obtain our main result as,

$$K(\boldsymbol{x}_I;\boldsymbol{x}_O)=\iint_{s_x^2+s_y^2\leq N_{0I}^2}ds_x\,ds_y$$
$$\times\exp\left[ik\left(\Phi-\left(\frac{x_O}{M}+x_I\right)s_x-\left(\frac{y_O}{M}+y_I\right)s_y-z_Is_z\right)\right]$$
$$\times\frac{1}{\lambda^2}\frac{1}{M}\left[\frac{1-(1-s_z^2)/M^2}{s_z^2}\right]^{1/4}, \tag{22}$$

thereby converting our result into the form of the Debye integral (ref. 8, p. 436). This result was obtained here by explicitly taking the limit of $d_I\rightarrow\infty$, while holding the magnification and N_{or} fixed. The lens objective system was assumed to satisfy the aplanatic condition of eq. (7) and the condition of negligible light reflections off lens surfaces. Equation (22) takes into account image reversal and the reduction of the object size, since the substitution $x_{O'}=-x_O/M$ and $y_{O'}=-y_O/M$ converts eq. (22) into a form involving the difference of the image and scaled object coordinates.

Ignoring polarization effects, the observable intensity will be proportional to the time average of $[V^{(r)}(\boldsymbol{x},t)]^2$. Hence, we may substitute our result of eq. (22) into the classical Hopkins formula[5,7]

$$I(\boldsymbol{x}_I)=\frac{1}{I_F}\iiiint_{-\infty}^{\infty}dx_O\,dy_O\,dx_O'\,dy_O'K(\boldsymbol{x}_I;\boldsymbol{x}_O)$$
$$\times K^*(\boldsymbol{x}_I;\boldsymbol{x}_O')T(\boldsymbol{x}_O)T^*(\boldsymbol{x}_O')J(\boldsymbol{x}_O;\boldsymbol{x}_O'), \tag{23}$$

where I is the intensity, $J(\boldsymbol{x}_O;\boldsymbol{x}_O')$ is the mutual intensity distribution,[7,8] and I_F is a normalization factor. The use

of this formula implies that a quasi monochromatic assumption holds for all nonzero transmission points on the mask (see ref. 7, §7.2.2). Upon specifying the mutual intensity distribution $J(x_O; x'_O)$, then eqs. (22) and (23) enable one to compute the aerial image of a mask.

Finally, we should note a regime in which the above results break down and need to be modified. In particular, if $N_{ol}=(1/M)N_{or}$ is *not* small and if $\sigma \gtrsim 1$ [eq. (26)], then $N_{cr}=\sigma N_{ol}$ will not be small either. The effect of the mask upon the light from the condenser will then *not* be well approximated by a transmission function acting upon a wavefront that lies approximately in the same plane as the mask. In this case, the precise field at the mask critically depends on how the imaging in the condenser lens is done. We note, however, that for condenser lens systems presently used in microlithography manufacturing, N_{cr} is typically small, so that the above concern should not be a problem.

2.3 *Condenser*

The mutual intensity distribution $J(x_O; x'_O)$ arises from the properties of the light source and condenser lens system in Fig. 1. Usually the source is constructed so as to provide incoherent light over some aperture opening. More specifically, for two points x_S and x'_S within the plane of this aperture, the mutual intensity distribution $J(x_S; x'_S)$ is approximated by being proportional to $I(x_S)\delta(x_S-x'_S)\delta(y_S-y'_S)$, where $I(x_S)$ is the intensity at x_S. Computing $J(x_O; x'_O)$ then entails imaging this incoherent illumination onto the object plane of the objective lens.

A few research groups recently reported improved imaging of small mask shapes by altering aspects of the condenser illumination system involving light sources that lie off the optical axis.[17-21] In addition, spatial filters have also been found to be helpful in improving resolution.[22] Provided that N_{cl} and N_{cr} are still small, however, as has been reported to date, then the basic paraxial approximations will still be valid for this part of the optical projection system. From a simulation point of view, the complication lies in accurately computing $J(x_O; x'_O)$ by numerically integrating over the aperture opening of the incoherent light source, while using the appropriate amplitude transmission function of the condenser lens system. In general, analytic solutions do not exist and numerical computations are necessary.

However, a special case does exist that is often used in practice. Under relatively relaxed conditions for Köhler illumination and under more special conditions for critical illumination (ref. 8, §10.5.1-10.5.2), $J(x_O; x'_O)$ reduces to, aside from a proportionality constant,

$$J(x_O; x'_O)=\frac{2J_1(v)}{v}, \tag{24}$$

where $J_1(v)$ is the first order Bessel function of the first kind,

$$v=\frac{2\pi\sigma N_{or}}{\lambda M}[(x_O-x'_O)^2+(y_O-y'_O)^2]^{1/2}, \tag{25}$$

and

$$\sigma \equiv N_{cr}/N_{ol} \tag{26}$$

is defined to be the partial coherence parameter for this type of condenser system. A circular aperture for the light source was assumed. Noting that $\sigma N_{or}/M=N_{cr}$, then eq. (24) depends only on the condenser imaging properties, as should be expected.

2.4 *Normalization*

The intensity $I(x_I)$ in eq. (23) can be calculated for different points x_I to within the normalization factor I_F, which we now address. A convenient normalization for $I(x_I)$ is when $I=1$ for the case of "flood exposure", where the mask is removed, or alternatively and equivalently, when x_I lies well within the Gaussian image region of a very wide mask opening. For different condenser illumination systems, I_F will vary, as it depends directly on the choice for $J(x_O; x'_O)$. In general I_F needs to be computed numerically, although in the special case of eq. (24), an analytic result can be calculated as follows.

We let $I(x_I)=1$ when $T(x_O)$ in eq. (23) equals unity for all x_O in the object plane. When the condenser optics is specified by eq. (24), one can use the identity

$$\frac{J_1(v)}{v}=\frac{1}{2\pi}\iint_{\xi^2+\eta^2\le 1} d\xi\, d\eta \exp[-i(\xi v\cos\phi+\eta v\sin\phi)]. \tag{27}$$

The integrations over x_O, y_O, x'_O, and y'_O in eq. (23) can then be explicitly carried out, yielding four Dirac delta functions. The four integrations over s_x, s_y, s'_x, and s'_y due to the presence of $K(x_I; x_O)$ and $K^*(x_I; x'_O)$ in eq. (23), and due to the form of eq. (22), can then be readily evaluated. The remaining integrations are over ξ and η from eq. (27):

$$I_F=\frac{M^2}{\pi}\iint_{\xi^2+\eta^2\le\min(1,1/\sigma^2)} d\xi\, d\eta \left[\frac{1-\frac{\sigma^2N_{or}^2}{M^2}(\xi^2+\eta^2)}{1-\sigma^2N_{or}^2(\xi^2+\eta^2)}\right]^{1/2}. \tag{28}$$

This integration can be carried out by (1) converting to polar coordinates r and ϑ, (2) integrating over ϑ, (3) changing variables to $u=r^2$, and (4) finally using integrals #144 and #151 in ref. 23. We obtain:

$$I_F=\frac{M^2}{(\sigma N_{or})^2}\left([1-(1-t^2)^{1/2}\cdot(1-t^2/M^2)^{1/2}] +\frac{M^2-1}{M}\log\left[\frac{M(1-t^2/M^2)^{1/2}-(1-t^2)^{1/2}}{M-1}\right]\right), \tag{29}$$

where

$$t=\begin{Bmatrix}\sigma N_{or} & \text{if} & \sigma\le 1\\ N_{or} & \text{if} & \sigma>1\end{Bmatrix}. \tag{30}$$

The above normalization factor does not depend on z_I, where $z_I \ll d_I$, as is the usual case for projection systems in microlithography.

Figure 2 shows a set of plots of I_F/M^2 versus $t=\sigma N_{or}$, when $\sigma<1$. For a fixed value of $\sigma>1$, then I_F/M^2 versus N_{or} can be deduced from Fig. 2 by dividing I_F/M^2 by $1/\sigma^2$, and replacing t in the x-axis by N_{or}. Some interesting

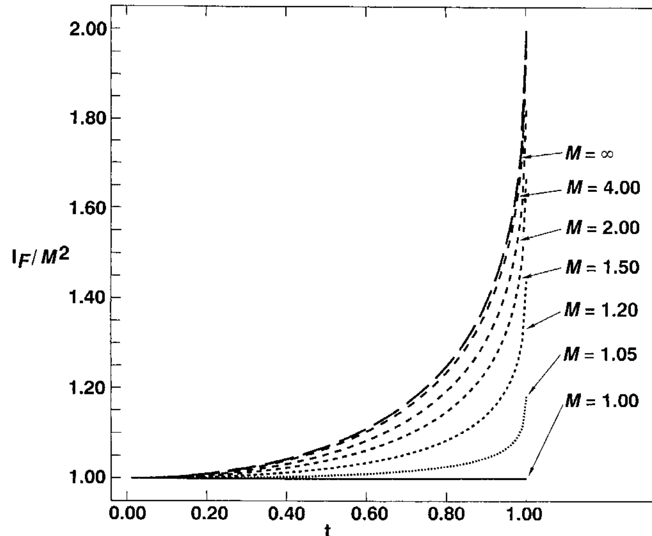

Fig. 2. Normalization factor of I_F/M^2 versus $t=\sigma N_{or}$, when $\sigma<1$.

special cases of eqs. (29) and (30) are that (a) $I_F=\min(1, 1/\sigma^2)$ when $M=1$, (b) $I_F/M^2\to 1$ when $\sigma\to 0$, (c) $I_F/M^2\to\min(1, 1/\sigma^2)$ when $N_{or}\to 0$, and (d)

$$I_F/M^2\to\frac{2}{(\sigma N_{or})^2}\{1-[1-(\sigma N_{or})^2]^{1/2}\},\quad \text{when } M\to\infty \text{ and } \sigma\leq 1. \tag{31}$$

We also note from Fig. 2 that I_F/M^2 lies quite close to case (d) for $M\gtrsim 4$.

2.5 *Reduction to paraxial result*

In ref. 10, we briefly described how the above formulae reduced to the paraxial results when $N_{or}\ll 1$, resulting in:

$$K(x_I;x_O)\approx\iint_{s_x^2+s_y^2\leq N_{or}^2} ds_x\,ds_y \times\exp\left[ik\left(\Phi-\left(\frac{x_O}{M}+x_I\right)s_x-\left(\frac{y_O}{M}+y_I\right)s_y +z_I\left(1+\frac{1}{2}(s_x^2+s_y^2)\right)\right)\right]\frac{1}{\lambda^2}\frac{1}{M}, \tag{32}$$

and

$$I_F\approx\begin{Bmatrix} M^2 & \text{if} \quad \sigma\leq 1\\ M^2/\sigma^2 & \text{if} \quad \sigma>1\end{Bmatrix}. \tag{33}$$

Except for the transformation $x_{O'}=-x_O/M$ and $y_{O'}=-y_O/M$, which one expects based solely on Gaussian optics, the image $I(x_I)$ is then independent of M. We note that for $4\lesssim M$, the nonparaxial result is also essentially independent of M. However, for $1\leq M\lesssim 4$, the nonparaxial formulae predict a noticeable dependence on M.

The familiar transformation to lithographers of the image being invariant under changes of wavelength, NA, and defocus, such that the scaled product of the mask coordinates times N_{or}/λ remains fixed, as well as the scaled product of defocus times N_{or}^2/λ remains fixed, can be shown to hold only in the case of the paraxial approximation. When $N_{or}\gtrsim 0.5$, this invariance breaks down and is most noticeable for larger values of defocus.

§3. Numerical Results

Turning to some specific simulation results, Fig. 3 shows the aerial image due to a 0.5 μm × 0.5 μm contact hole, with $\lambda=0.365$ μm, $N_{or}=0.6$, and $\sigma=0.2$, and with a circular aperture for the light source in the condenser. As with all the results shown here, the intensity is normalized to the intensity of flood exposure. Figure 4 compares the results for the paraxial [eqs. (32) and (33)] versus nonparaxial formulae [eqs. (22) and (29)] for a vertical slice taken across the center of the square contact imaged in Fig. 3. Here, the peak differences are 8% and 14% for the $z_I=0$ μm and $z_I=1$ μm cases, respectively.

As discussed in ref. 10, for small values of N_{or}, such as $N_{or}\lesssim 0.4$, the difference between the paraxial and nonparaxial scalar results is negligibly small. The difference becomes noticeable for $N_{or}>0.4$, particularly when $N_{or}\approx 0.6$ and the image is viewed from out of focus. Magnification effects are noticeable for $1\leq M\lesssim 4$. However, for $4\lesssim M$, the scalar aerial image becomes nearly independent of M, aside from the expected scaling of the mask to object dimensions. For $N_{or}\gtrsim 0.7$, scalar aerial image predictions are unreliable and vector theory calculations should be used. Indeed, for $0.6\lesssim N_{or}\lesssim 0.7$, then scalar aerial images will undoubtedly only be approximately valid.

Figure 5 shows the differences at the center of contact holes between the paraxial and nonparaxial formulae when imaging different sized contacts holes, as defocus is varied. Figures 5(a) and 5(b) illustrate this difference for $\lambda=0.365$ μm and $\lambda=0.248$ μm, respectively. As can be seen, in general the numerical differences computed from these formulae is difficult to anticipate, particularly when $z_I\neq 0$. When z_I is large, the relative differences in intensity can become quite large. For example, when $z_I=1.5$ μm,

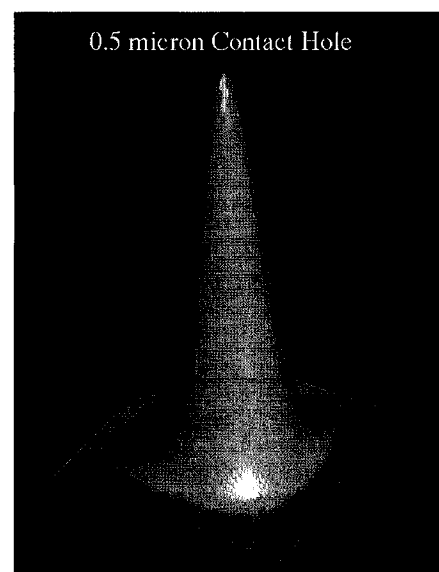

Fig. 3. Surface plot of intensity versus position over a 1.0 μm × 1.0 μm region in the plane of best focus in image space, due to an isolated, nonrepeating, M × 0.5 μm square contact hole, with $\lambda=0.365$ μm, $N_{or}=0.6$, $\sigma=0.2$, and $M=5$. The peak intensity is 1.52.

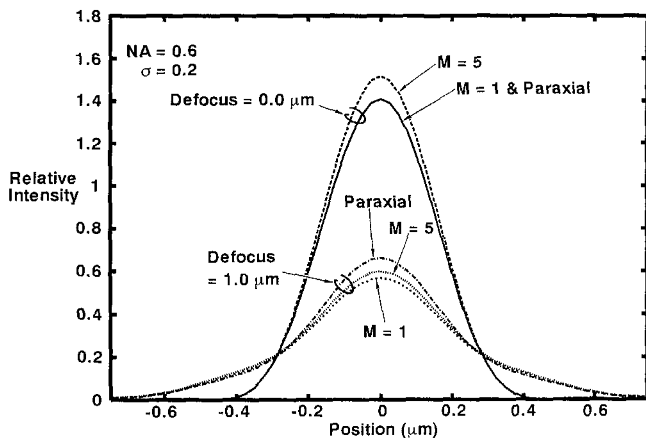

Fig. 4. Intensity across the center, vertical slice of a square contact hole, for $\lambda=0.365\,\mu$m, $\sigma=0.2$, and $N_{or}=0.6$. A $0.5\,\mu$m $\times$ $0.5\,\mu$m mask opening was assumed for the $M=1$ case, with an $M\times M$ larger mask opening for the $M=5$ case.

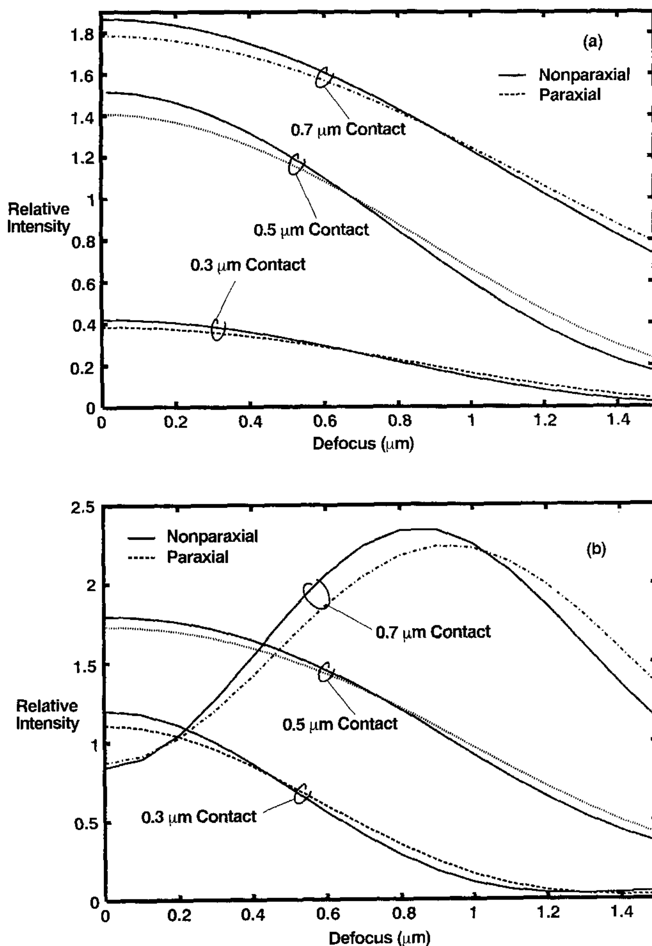

Fig. 5. Intensity at the center of different sized contact holes as defocus is varied. Here, $\sigma=0.2$, $N_{or}=0.6$, and $M=5$ for the nonparaxial case. (a) $\lambda=0.365\,\mu$m; (b) $\lambda=0.248\,\mu$m.

the relative differences for the 0.7 μm, 0.5 μm, and 0.3 μm sized contact holes in Fig. 5(a) are 7%, 26%, and 46%, respectively.

Nevertheless, there does exist a general pattern in Figs. 5(a) and 5(b) at $z_I=0.0\,\mu$m that can be understood roughly as follows. As the contact hole size goes to zero, the relative intensity will decrease to zero. As the contact hole size increases, the intensity will increase until it reaches a maximum value, then it will lower and oscillate around $I=1.0$, eventually reaching $I=1.0$ for the center of very large contact holes. Although the 0.7 μm contact hole curves in Fig. 5(b) undoubtedly seems surprising, they actually follow a general trend; indeed, a larger contact hole than 0.7 μm in Fig. 5(a) with $\lambda=0.365\,\mu$m, will yield a similar behavior. What is happening in the 0.7 μm case is that the relatively large "ringing", or overshoots in intensity at $\sigma=0.2$, that occurs at the edges of the contact hole when in focus, merge to the middle of the image when going out of focus, thereby producing a surprisingly high intensity at the center of the hole. At the same time, the intensity at the edges of the hole dimenishes below unity as z_I increases. At yet higher out of focus conditions, the center intensity will also fall well below unity.

Due to the differences noted above between the paraxial versus nonparaxial results, clearly there is a need to pin down when these differences are important for making judgments about imaging in microlithography. Hence, we now turn to the exposure-defocus (E-D) measure for judging the printability of masks, as discussed in ref. 1, and as elaborated upon subsequently in the work of Lin (*e.g.*, ref. 24). Here, exposure is taken to be equal to 1/intensity, where the intensity is calculated at plus and minus some tolerance of the desired feature width, over a range of defocus values.

The curves on the right in Fig. 6 represent the E-D plots for the 0.5 μm $\times$ 0.5 μm contact hole, calculated using the paraxial and nonparaxial formulae. A $\pm10\%$ tolerance was assumed in the contact hole width, as well as a $\pm10\%$ width in exposure. The depth of focus (DOF) computed using the two formulae differed by 6%. Hence, even though larger relative differences than this result existed in intensity at most points for this contact hole, still the computed DOF was only moderately different for the paraxial versus nonparaxial formulae.

The curves on the left in Fig. 6 were computed for a contact hole designed using phase shifted elements to print a 0.5 μm square contact hole. As expected for well designed phase shifted structures,[24,25] the DOF is dramatically improved for this contact hole structure; here, DOF is improved by nearly a factor of two. The computed DOF now varies by 9% between the paraxial and nonparaxial results. This increased relative difference is due to sampling of regions farther out in focus, where differences generally increase between the two formulae. Other contact hole sizes simulated at $N_{or}=0.6$ and $\sigma=0.2$, yielded similar results, with maximum differences ranging typically between 5%–15% when phase shifting was included.

When properly used, advanced lithography methods such as (1) phase shifts in the mask,[24,25] (2) modified illumination in the condenser,[17-22] (3) spatial filters in the entrance and exit pupil planes of the objective lens system,[27] and (4) averaging over different defocus

planes,[27,28] can resolve significantly smaller feature sizes at larger DOFs. These techniques generally result in larger differences between the paraxial and nonparaxial formulae than when the techniques are not applied, simply because of the gain in DOF.

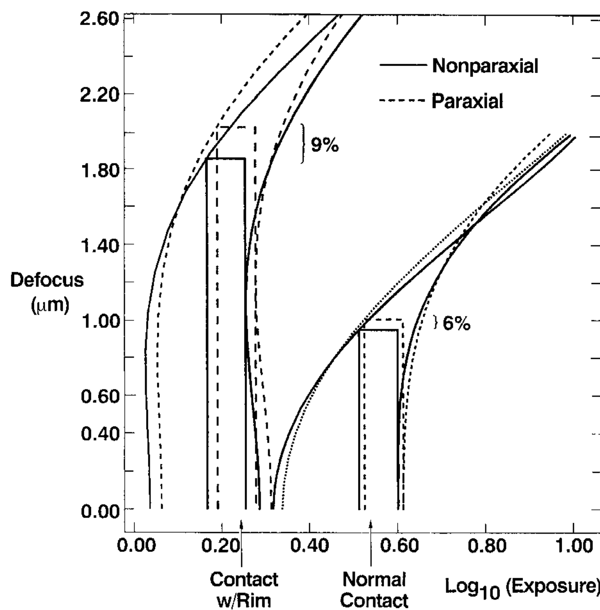

Fig. 6. Exposure-defocus (E-D) plots taken from the vertical center slice across contact holes. Here, $\lambda=0.365$ μm, $\sigma=0.2$, and $N_{or}=0.6$. For the nonparaxial case, $M=5$. Plots are shown for a normal square hole with width 0.5 μm, and for a contact hole with an inner clear region 0.3 μm × 0.3 μm, with a 0.2 μm opaque border, and a 0.2 μm clear outside border shifted in phase by π from the inner clear region.

Finally, we conclude this section on numerical results by illustrating the aerial images calculated for situations involving these advanced lithography methods. Figures 7 and 8 illustrate the use of an alternating phase shifting method to improve the aerial image resolution of periodic, rectangular mask shapes. Figures 9 and 10 illustrate modified illumination methods to image contact hole structures with and without phase shifting methods. Images using modified illumination methods can also be computed for fairly large mask regions, as shown in Fig. 11. (This mask design was illustrated in ref. 20.) Recent advances in computational methods by some of us have significantly sped up the computation of aerial images to enable images, such as in Fig. 11, to be computed in well under 1 minute on an IBM RS/6000, model 540 workstation.

§4. Concluding Remarks

A detailed derivation was given for extending scalar diffraction theory for imaging in a projection lens systems to the case where the paraxial, or small angle, approximation is not invoked. This derivation was carried out in the asymptotic limit of large lens and focal length sizes, while keeping NAs and magnification fixed. Physical assumptions used here, that were originally contained in work by others on vector imaging theory,[11-13] were highlighted.

The differences between the paraxial and nonparaxial formulae are highly dependent on spatial coherence, wavelength, NA, and defocus. For $N_{or}\lesssim 0.4$, the differences are negligible, but become more significant for $N_{or}\approx 0.6$. For $N_{or}\approx 0.6$, relative differences vary up to

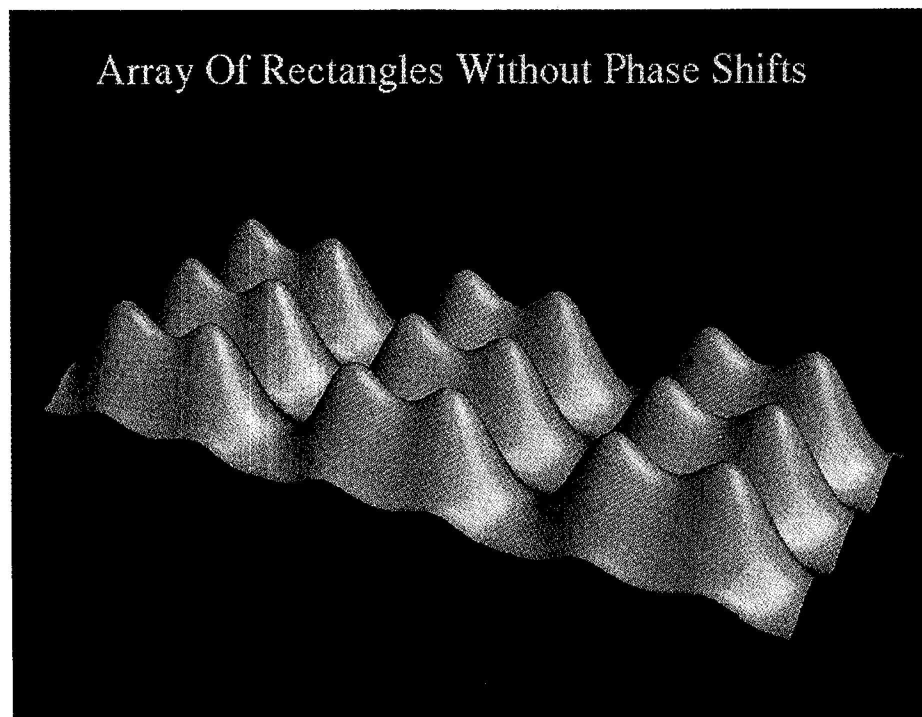

Fig. 7. Aerial image of a periodic pattern of clear rectangles with dimensions 0.3 μm × 1.3 μm, separated by 0.3 μm and 0.2 μm in the width and length directions, respectively. Here, $\lambda=0.365$ μm, $N_{or}=0.6$, and $\sigma=0.2$. The aerial image poorly resolves the rectangular mask pattern. A 1.8 μm × 1.5 μm region is shown. The peak intensity in the paraxial case is 0.68.

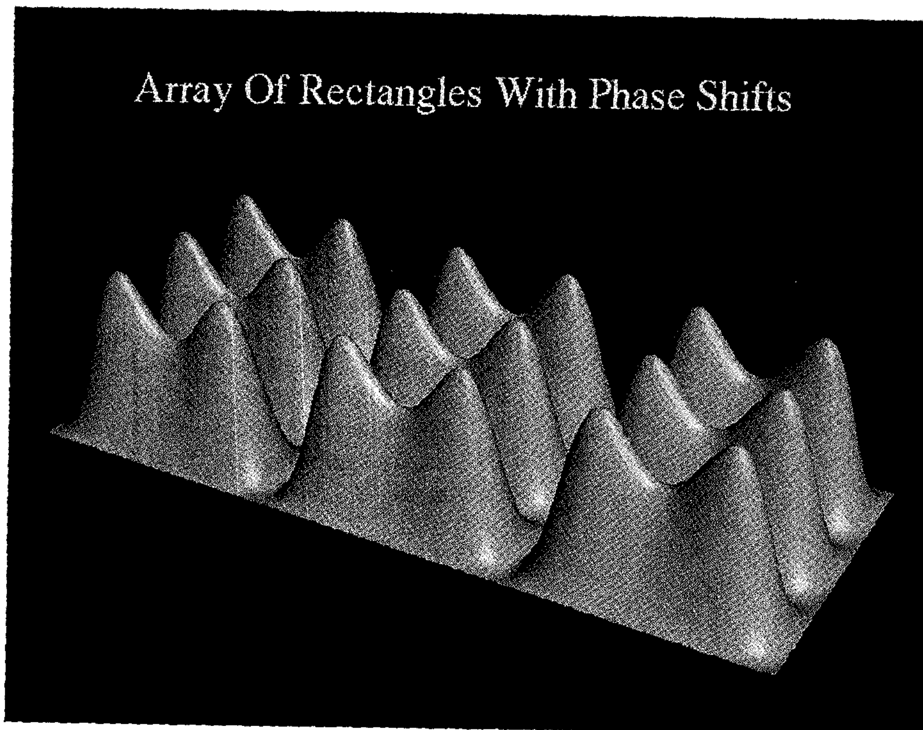

Fig. 8. Aerial image of the same periodic array of rectangles as in Fig. 7, but here every rectangle is shifted in phase by π from its neighbors. Phase shifting the features improves the replication of the rectangular mask pattern by the aerial image.

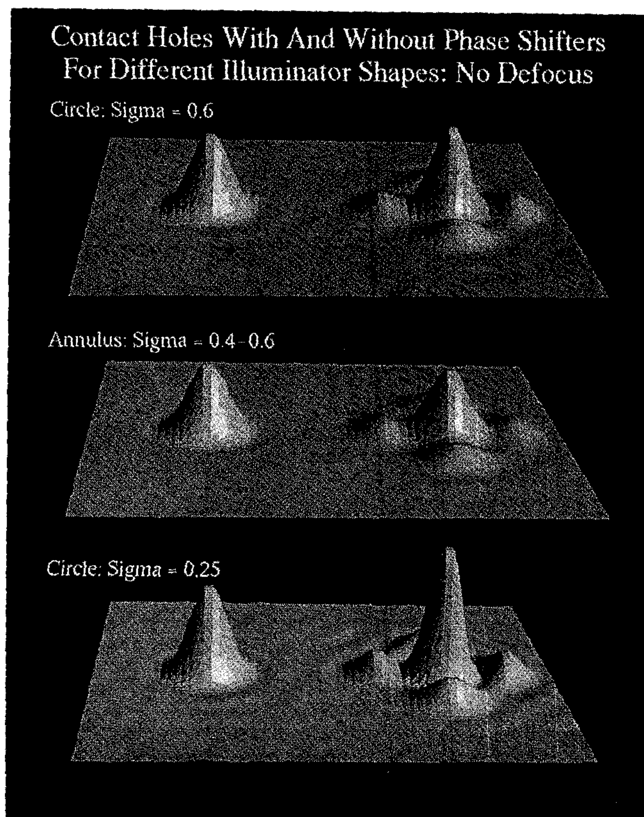

Fig. 9. Aerial images of two contact holes with dimensions 0.4 μm × 0.4 μm, for three different illuminator shapes. The left contact hole is normal, while the right one has outrigger phase shifters at a phase of π. Here, $\lambda = 0.365$ μm, $N_{or} = 0.6$, $M = 5$. The images are computed at the best focus plane.

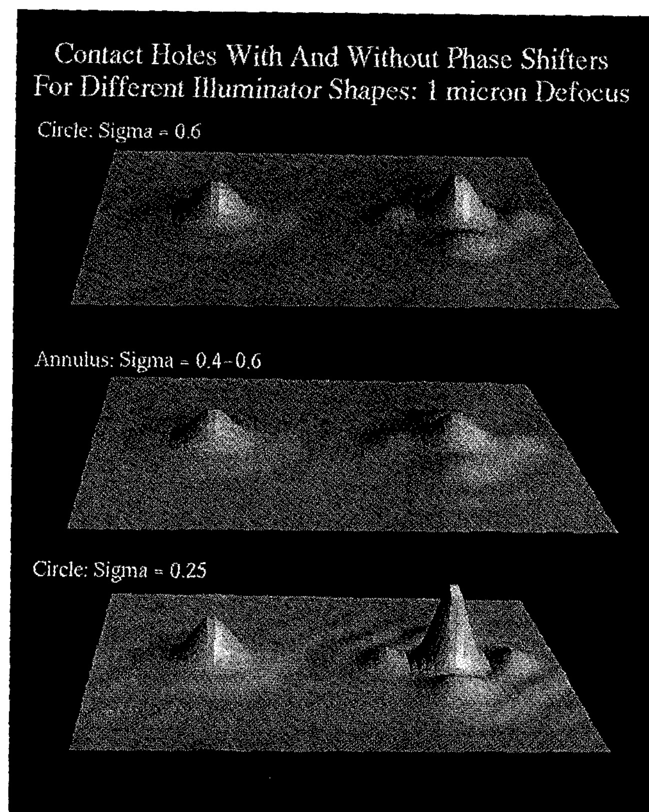

Fig. 10. Aerial images of the same contact holes as in the previous figure, but at a defocus of 1.0 μm.

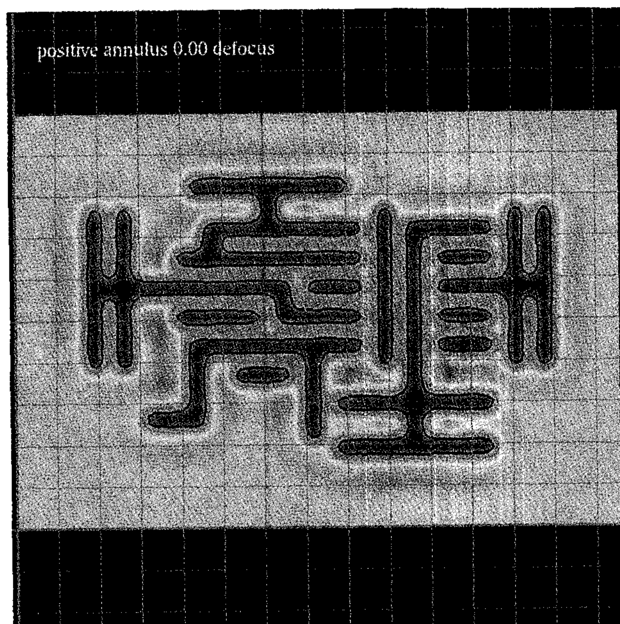

Fig. 11. Contour plot of the aerial image of a relatively large mask region (10.0 μm × 14.5 μm) for the mask reported at the end of ref. 20. Here, $\lambda=0.365$ μm, $N_{or}=0.55$, $M=4$, $\sigma_{inner}=0.4$, and $\sigma_{outer}=0.6$ (annular illumination). The dark contour level is chosen at a normalized intensity of 0.3, where unity represents the intensity at the center of a large open mask region.

about a 10% effect, although stronger differences up to about 15% can be found if the DOF is increased significantly via advanced imaging methods.

We note that the numerical complexity of our formulae is essentially the same as that for the paraxial case. Hence, we recommend the use of the scalar nonparaxial result for imaging, provided of course that the NA of the objective lens is not made so large that polarization effects demand that a full vector diffraction theory approach be used.

Acknowledgements

Three of the authors (E. Barouch, U. Hollerbach, and S. A. Orszag) were supported in part by the Air Force Office of Scientific Research (AFOSR), the Defense Advanced Research Project Agency (DARPA), the Naval Research Laboratory (NRL), and the U.S. Army Labcom. The authors thank D. G. Flagello, A. E. Rosenbluth, T. Brunner, and M. S. Yeung for helpful discussions.

References

1) A. E. Rosenbluth, D. Goodman and B. J. Lin: J. Vac. Sci. & Technol. B **1** (1983) 1190.
2) E. Barouch, U. Hollerbach, S. A. Orszag, B. Bradie and M. Peckerar: IEEE Electron Device Lett. **EDL-12** (1991) 513.
3) E. Barouch, U. Hollerbach and S. A. Orszag: Proc. SPIE **1465** (1991) 254.
4) E. Barouch, B. Bradie, U. Hollerbach and S. A. Orszag: J. Vac. Sci. & Technol. B **8** (1990) 1432.
5) H. H. Hopkins: Proc. Roy. Soc. A **217** (1953) 408.
6) J. W. Goodman: *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
7) J. W. Goodman: *Statistical Optics* (Wiley, New York, 1985).
8) M. Born and E. Wolf: *Principles of Optics* (Pergamon, New York, 1975) 5th ed.
9) J. J. Stamnes: *Waves in Focal Regions* (Adam Hilger, Boston, 1986).
10) D. C. Cole, E. Barouch, U. Hollerbach and S. A. Orszag: J. Vac. Sci. & Technol. B **10** (1992) 3037.
11) E. Wolf: Proc. Roy. Soc. A **253** (1959) 349.
12) B. Richards and E. Wolf: Proc. Roy. Soc. **253** (1959) 358.
13) M. S. Yeung: Proc. SPIE **922** (1988) 149.
14) D. G. Flagello and T. Milster: Proc. SPIE **1625** (1992).
15) D. G. Flagello and A. E. Rosenbluth: J. Vac. Sci. & Technol. B **10** (1992) 2997.
16) B. Küyel, E. Barouch, U. Hollerbach and S. A. Orszag: Proc. SPIE **1674** (1992).
17) K. Kamon, T. Miyamoto, Y. Myoi, H. Nagata, M. Tanaka and K. Horie: Jpn. J. Appl. Phys. **30** (1991) 3021.
18) K. Tounai, H. Tanabe, H. Nozue and K. Kasama: Proc. SPIE **1674** (1992) 753.
19) M. Noguchi, M. Muraki, Y. Iwasaki and A. Suzuki: Proc. SPIE **1674** (1992) 92.
20) N. Shiraishi, S. Hirukawa, Y. Takeuchi and N. Magome: Proc. SPIE **1674** (1992) 741.
21) S. Asai, I. Hanyu and K. Hikosaka: Microelectron. Eng. **17** (1992) 109.
22) E. Tamechika, S. Matsuo, K. Komatsu, Y. Takeuchi, Y. Mimura and K. Harada: J. Vac. Sci. & Technol. B **10** (1992) 3027.
23) *CRC Handbook of Chemistry and Physics*, ed. R. C. Weast (The Chemical Rubber Co., Cleveland, Ohio, USA, 1972) 53rd ed.
24) B. J. Lin: Proc. SPIE **1496** (1991) 54.
25) M. D. Levenson, N. S. Viswanathan and R. A. Simpson: IEEE Trans. Electron Devices **ED-29** (1982) 1828.
26) H. Fukuda, T. Terasawa and S. Okazaki: J. Vac. Sci. & Technol. B **9** (1991) 3113.
27) H. Fukuda, N. Hasegawa, T. Tanaka and T. Hayashida: IEEE Electron Device Lett. **EDL-8** (1987) 179.
28) C. Spence, D. Cole and B. Peck: Proc. SPIE **1674** (1992).