Extending scalar aerial image calculations to higher numerical apertures

Daniel C. Cole
IBM Corporation, Essex Junction, Vermont 05452

Eytan Barouch, Uwe Hollerbach, and Steven A. Orszag
Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544

(Received 27 May 1992; accepted 6 August 1992)

The usual formula for the scalar aerial image of an isolated object due to a projection lens system has been generalized beyond the paraxial approximation in an attempt to extend scalar diffraction theory to include numerical aperture (NA) values up to about 0.6. Beyond this regime, or certainly beyond NA=0.7, polarization effects need to be included, thereby demanding a full vector treatment and invalidating the present scalar formulation. A key point to the present scalar result without the paraxial approximation is the predicted functional dependence of the aerial image on magnification as NA increases. A second key point is that the usual scaling of λ/NA for the object dimensions and λ/NA^2 for defocus become invalid for high NA systems. Numerical results of illustrative test cases are shown.

I. INTRODUCTION

In the usual derivation of the scalar aerial image for optical projection systems, as in, for example, Ref. 1 and Secs. 7.1–7.2 in Ref. 2, a number of approximations are made that restrict the range of applicability of the scalar result to numerical aperture (NA) values that are not "too large." More precisely, although the applicable range of the scalar formula is not fully known, since full vector calculations for partially coherent light in projection systems have only been relatively recently pursued,[3] generally the feeling by most researchers has been that the accuracy of scalar aerial image calculations becomes questionable when inclination angles exceed about 30°, or NA=0.5 (see, e.g., Refs. 4-8).

A number of researchers have made substantial contributions to the required vector theory and computation of imaging of light for high numerical aperture systems. Particularly important advances on coherent light systems are Refs. 9–12. Stamnes' book[13] contains an exhaustive study of more recent information on coherent systems. Recent simulation work is contained in Refs. 8, 14, and 15 for coherent light, and Refs. 3, 16, and 17 for the more general case of partially coherent light.

While working on the theory for vector aerial image calculations, we observed that the range of applicability of the less complicated *scalar* aerial image for partially coherent projection lens systems might be extendable. Our interest in doing so is threefold. First, the generalized scalar aerial image may yield reasonably accurate scalar aerial images up to NAs as high as 0.6 to, with less accuracy, even 0.7, before polarization effects become critically important. (Indeed, if one considers an imaging system with light largely polarized such that its electric field is perpendicular to the plane of incidence, where the light is being imaged onto a plane perpendicular to the optic axis, then the present scalar diffraction theory *without* the paraxial, or small angle, approximation should agree quite closely with vector diffraction theory. This result should hold even for NAs higher than 0.7.) Hence, this approach may address the probable next generation of microlithography projection systems designed for achieving 256-Mb DRAM technologies. Second, improving the approximations made in the usual scalar approach have a theoretical interest, since the intrinsic simpler nature of the scalar approach helps to clearly identify many of the usual physical assumptions in such calculations, as well as to bring out one physical parameter not treated in the usual scalar approach: namely, the magnification M. Third, more clearly understanding the key physical assumptions in the simpler scalar case may yield insights into improving the physics typically inserted into the treatment of vector aerial image calculations.

Despite the simpler nature of scalar aerial image theory as compared with the vector case, the full derivation is still relatively long. Consequently, this article will concentrate here on the description of our final formulas, as well as on some numerical results to illustrate the new physical points of interest. The full details of our derivation will be presented elsewhere.[18]

As for the outline of our article, in Sec. II we briefly discuss what key physical points need to be included to address higher NA regimes than the usual scalar image approach. In Sec. III, the quantities in the formula for our more exact scalar aerial image are discussed. Section III then turns to numerical results, while Sec. IV contains some concluding remarks.

II. PHYSICAL COMPLICATIONS OF HIGH NA SYSTEMS

Unless one is relatively deeply involved in the physical theory for imaging upon and exposing photoresist for microlithography systems, the different physical corrections reported in the literature[3,8,16,17,19–23] during recent years for high NA systems must seem quite complicated. Nevertheless, one can readily divide these reported high NA corrections into two categories: namely, (1) the imaging due to high NA projection systems, which is addressed in

 ©1992 American Vacuum Society

Refs. 3, 8, 16, and 17, and (2) the nonvertical propagation of the light within the resist structure, which is largely addressed in Refs. 3, 8, 16, 17, and 19–23. Some of these articles are restricted to (i) one-dimensional (1D) as opposed to two-dimensional (2D) masks, (ii) coherent illumination rather than the more general case of partial coherence, and (iii) planar rather than nonplanar resist and substrate structures. Also, most of the approaches employ different numerical computational schemes while striving for accuracy versus speed tradeoffs.

This present work addresses *only* (1) above: the high, or actually middle, regime of NA correction due to the imaging of light in air, prior to possible subsequent light propagation within a photoresist structure. Even within this subcategory of high NA corrections, there exist different levels of physical accuracy. One such level is quite apparent from our title, since the article is purposefully not including the vector character of electromagnetic propagation. Here, the two main physical approximations we improve upon, as compared with previous scalar treatments, is to not introduce the paraxial approximation, nor the thin lens approximation (see, e.g., Sec. 5-1 in Ref. 24). Assumptions that we do make, are the following: (1) we assume, as discussed by Wolf in Ref. 10, that the intensity law of geometrical optics (Sec. 3.1.2 in Ref. 25) holds for rays of light connecting the entrance and exit pupil of our optical projection system; (2) we are dealing with a sufficiently corrected lens system that the surfaces of the lenses result in the sine theorem being obeyed (Sec. 4.5.1 in Ref. 25) for ray-traces through the optical system; and (3) that the Strehl ratio of the lens system is sufficiently high that we can neglect reflections off lens surfaces. To go beyond any of these three assumptions appears to require that the details of the lens surfaces be taken into account in the computations, which for typical projection systems entails a very impractical computational task.

A fair part of our work can be viewed as the scalar complement of the imaging part of Yeung's key work in Ref. 3. Other differences do exist, however, such as we calculate results for 2D masks, we employ different numerical methods, and our treatment of the condenser lens system is simplified. Also, in our derivation, an explicit asymptotic limit is taken for infinitely large lenses and focal lengths, as compared with image and object sizes.[18] For most practical purposes this regime is what is typically found in microlithography projection systems. Because of our scalar treatment, and because of the asymptotic limit taken, the final formulas are relatively simple, thereby enabling one to clearly see how magnification enters into the intensity expression, and how the paraxial limit can be readily obtained. A similar derivation and analysis can be carried out for the more complicated, but physically more accurate, vector aerial image situation.

III. DISCUSSION OF SCALAR AERIAL IMAGE FORMULAS

The expression we recently derived[18] for the relative scalar intensity at point $\mathbf{x}_I$ in the image space of a projection system indicated schematically in Fig. 1, is

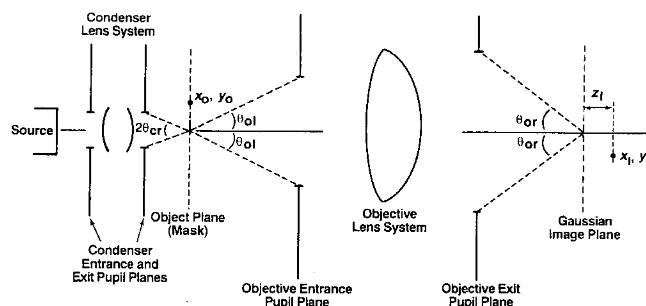

FIG. 1. Schematic outline of optical projection system.

$$I(\mathbf{x}_I)=\frac{1}{I_F}\iiint_{-\infty}^{\infty} dx_O\, dy_O\, dx'_O\, dy'_O\, K(\mathbf{x}_I;\mathbf{x}_O) \times K^*(\mathbf{x}_I;\mathbf{x}'_O)T(\mathbf{x}_O)T^*(\mathbf{x}'_O)J(\mathbf{x}_O;\mathbf{x}'_O), \quad (1)$$

where $\mathbf{x}_O$ and $\mathbf{x}'_O$ are points on the mask in the object space, $T(\mathbf{x}_O)$ equals the complex transmission function of the mask at the point $\mathbf{x}_O$, $K(\mathbf{x}_I;\mathbf{x}_O)$ is an amplitude transmission function from object point $\mathbf{x}_O$ to image point $\mathbf{x}_I$, $J(\mathbf{x}_O;\mathbf{x}'_O)$ represents the mutual intensity distribution[2,25] incident on the mask, and I_F is a normalization constant chosen such that $I(\mathbf{x}_I)=1$ if no mask is present, thereby allowing a "flood exposure." Under relatively relaxed conditions for (circular) Köhler illumination, and under more special conditions for critical illumination, $J(\mathbf{x}_O;\mathbf{x}'_O)$ reduces to, aside from a proportionality constant,

$$J(\mathbf{x}_O;\mathbf{x}'_O)\approx\frac{2J_1(\nu)}{\nu}, \quad (2)$$

where

$$\nu=\frac{2\pi\sigma N_{or}}{\lambda M}\left[(x_O-x'_O)^2+(y_O-y'_O)^2\right]^{1/2}. \quad (3)$$

Here, M is the "magnification," or more precisely the reduction given by $M=$ (object height)/(image height), λ is the wavelength, N_{or} is the NA on the right-hand side of the objective lens in Fig. 1 ($N_{or}=\sin\Theta_{or}$), and σ is the partial coherence parameter given here by

$$\sigma=N_{cr}/N_{ol}. \quad (4)$$

In Eq. (4), N_{ol} and N_{cr} refer to, respectively, the NAs on the left- and right-hand sides of the objective and condenser lenses in Fig. 1 ($N_{ol}=\sin\Theta_{ol}$ and $N_{cr}=\sin\Theta_{cr}$). The conditions referred to above for the application of Eqs. (2)–(4) to Köhler and critical illumination are largely satisfied if N_{cr} is small, which is typically true. For more specifics, one should refer to Ref. 25, Secs. 10.5.1–10.5.2.

Actually, so far the formulas listed here apply for both the paraxial and nonparaxial situations. The difference between these two cases arises once $K(\mathbf{x}_I;\mathbf{x}_O)$ is specified. For the nonparaxial scalar case, we obtained:

$$K(\mathbf{x}_I;\mathbf{x}_O)=\iint_{s_x^2+s_y^2<N_{or}^2} ds_x\,ds_y\left(\exp\left\{ik\left[\Phi-\left(\frac{x_O}{M}+x_I\right)s_x-\left(\frac{y_O}{M}+y_I\right)s_y+z_I(1-s_x^2-s_y^2)^{1/2}\right]\right\}\right)\times\frac{1}{\lambda^2}\frac{1}{M}\left(\frac{1-(s_x^2+s_y^2)/M^2}{1-s_x^2-s_y^2}\right)^{1/4}. \quad (5)$$

Here, $k=2\pi/\lambda$; $\hat{\mathbf{s}}$ is a unit vector that points from the point in the plane of focus on the optic axis to a point on the exit pupil plane; $N_{or}=\sin\Theta_{or}$ (Fig. 1), where Θ_{ol} and Θ_{or} are related by the sine theorem,

$$M\sin\Theta_{ol}=\sin\Theta_{or}\,; \quad (6)$$

Φ is the aberration function; and z_I is the displacement from the best focus position according to Gaussian optics. Equation (5) takes into account image reversal and the reduction of the object size; the substitution $u_O=-x_O/M$ and $v_O=-y_O/M$ converts Eq. (5) into a form involving the difference of the image and scaled object coordinates.

Finally, the normalization factor I_F in Eq. (1) is given by

$$I_F=\frac{M^2}{(\sigma N_{or})^2}\left[[1-(1-t^2)^{1/2}(1-t^2/M^2)^{1/2}]+\frac{M^2-1}{M}\log\left(\frac{M(1-t^2/M^2)^{1/2}-(1-t^2)^{1/2}}{M-1}\right)\right], \quad (7)$$

where

$$t=\begin{cases}\sigma N_{or}, & \text{if } \sigma\leqslant 1\\ N_{or}, & \text{if } \sigma>1\end{cases}. \quad (8)$$

To see how Eq. (5) reduces to the usual paraxial case, one needs to impose that $N_{or}\ll 1$, which implies $(s_x^2+s_y^2)\ll 1$:

$$K(\mathbf{x}_I;\mathbf{x}_O)\approx\iint_{s_x^2+s_y^2<N_{or}^2} ds_x\,ds_y\left(\exp\left\{ik\left[\Phi-\left(\frac{x_O}{M}+x_I\right)s_x-\left(\frac{y_O}{M}+y_I\right)s_y+z_I\left(1+\frac{1}{2}(s_x^2+s_y^2)\right)\right]\right\}\right)\frac{1}{\lambda^2}\frac{1}{M}. \quad (9)$$

Letting $u_O=-x_O/M$ and $v_O=-y_O/M$ then yields the usual paraxial image approximation due to a nonrepeating object [see, e.g., Eq. (5-34) and pp. 123–124 in Ref. 24, with $s_x=x_I/d_I$, $s_y=y_I/d_I$, and $M\to 1/M$]. To further reduce the above result to the form that is at present typically used in most scalar aerial image simulators, one needs to introduce a periodic repeating mask, so that the mask transmission function $T(\mathbf{x}_O)$ in Eq. (1) is expressed in terms of a Fourier series (see, e.g., Refs. 1, 26, and 27). Finally, for the paraxial case of $N_{or}\ll 1$, Eqs. (7) and (8) yields

$$I_F\approx\begin{cases}M^2, & \text{if } \sigma\leqslant 1\\ M^2/\sigma^2, & \text{if } \sigma>1\end{cases}. \quad (10)$$

Hence, from Eqs. (1), (9), and (10), $I(\mathbf{x}_I)$ is *independent* of M in the paraxial limit of $N_{or}\ll 1$ (except for $u_0 = -x_O/M$ and $v_0 = -y_O/M$). As far as we know, this independence of M is what is assumed in all scalar image simulators reported to date in the literature.

As a point of interest, we note that in the nonparaxial, aplanatic case of Eqs. (5) and (7), with $M=1$ and zero defocus ($z_I=0$), we obtain exact agreement with the usual paraxial approximation of Eqs. (9) and (10). (When $z_I\neq 0$, these two cases are *not* the same.) Evidently, here we have the interesting situation that compensating errors in the paraxial approximation cancel out when treating the $M=1$, $z_I=0$, nonparaxial, aplanatic scalar situation. As a word of caution in applying our formulas, we note that our restriction of low N_{cr} for Eq. (2) to hold requires that $N_{cr}=\sigma N_{or}/M$ should be small compared to unity, or at least smaller than about 0.4.

IV. NUMERICAL RESULTS

We evaluated the fourfold integral over the mask in Eq. (1) by the use of a Gauss–Legendre quadrature scheme, where the regions of integration were restricted to the illuminating regions of the mask. Figures 2(a)–2(d) contain plots of $I(\mathbf{x}_I)$ across a vertical slice of a square contact hole; here, $\lambda=0.365$ μm, and several values of M and σ were chosen. The $M=1$ curves also indicate the paraxial results for Figs. 2(a)–2(c). For the defocus $=1$ μm case in Fig. 2(d), with $N_{or}=0.6$ and $\sigma=0.2$, the paraxial and $M=1$ curves are *not* the same, as anticipated earlier. Indeed, here the paraxial curve lies *above* the nonparaxial curves, with the peak being about 14% higher than the $M=1$ curve. In contrast, for the $z_I=0$ μm curves in Figs. 2(a)–2(d), the paraxial curves lie *below* the nonparaxial curves, with a difference existing of about 8% at the peak of the $N_{or}=0.6$, $\sigma=0.2$, and $M=5$ curve in Fig. 2(d). Comparison with the simulator (SPLAT) described in Ref. 27, taken for a large periodic background, yielded nearly identical results to the paraxial curves calculated here from Eq. (9).

Thus, distinct changes in intensity from the paraxial case are predicted by the scalar, nonparaxial results, particularly when $z_I\neq 0$. It was found that when $N_{or}=0.4$, 0.5, 0.6, and 0.7, differences of about 3% and 2%, 5% and 3%, 8% and 5%, and 10% and 6% were observed in the peak intensities for the $\sigma=0.2$ and $\sigma=0.6$ cases, respectively, of the $z_I=0$ μm, nonparaxial versus paraxial curves in Fig. 2. {We note that $N_{or}\approx 0.7$ is probably in the regime where the vector nature of the light is significant; however, we present our scalar results [Fig. 2(c)] for this case as well for future comparison purposes.} Thus, the differences increase as N_{or} increases, and are larger when the coherency of the light is larger (σ smaller). As for reduction, plots of aerial images were found to visibly change with M for $M=1$ up to about $M\approx 5$; however, for $M\gtrsim 5$, further changes from the $M=5$ case were quite small, as can be expected upon examining Eq. (5).

Figures 3 and 4 show the application of our formulas to a more complicated mask structure containing five transparent regions, with two of the features shifted in phase by

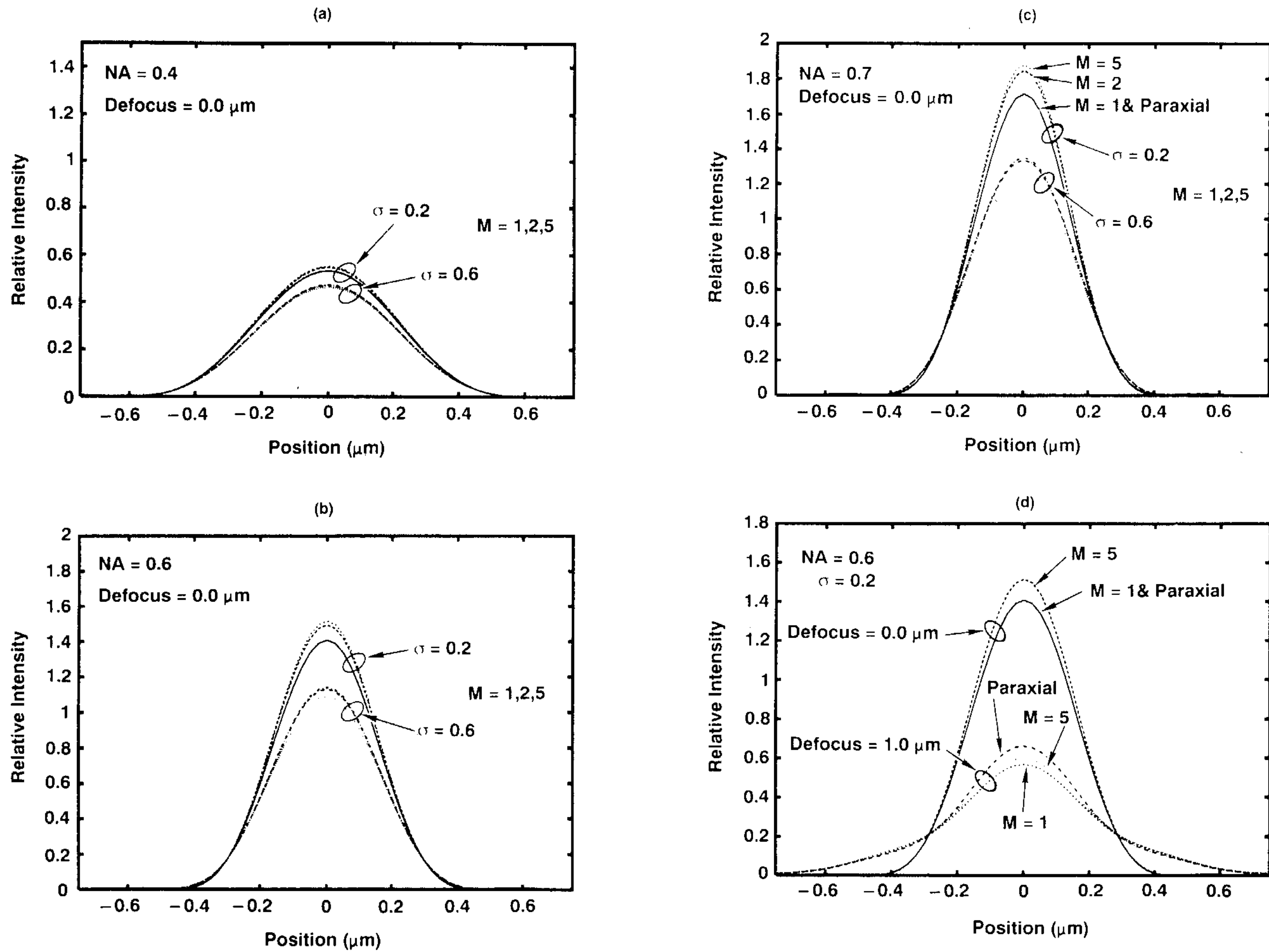

FIG. 2. Relative intensity, normalized to the intensity due to flood exposure (no mask), for $\lambda=0.365$ μm. A 0.5 μm$\times$0.5 μm square mask opening was assumed for the $M=1$ case, with $M\times M$ larger mask openings for the $M=2$ and $M=5$ cases. A vertical slice through the center was taken. (a) $N_{or}=0.4$, $z_I=0$ μm; (b) $N_{or}=0.6$, $z_I=0$ μm; (c) $N_{or}=0.7$, $z_I=0$ μm; and (d) $N_{or}=0.6$, $z_I=(0,1)$ μm. For each value of σ in (a)–(c), the curves with the highest to lowest peak intensities, respectively, correspond to the $M=5$, $M=2$, and $M=1$ magnifications.

π. The entire mask is treated as completely isolated and nonrepeating. We note that low σ values are precisely the regime where phase-shifted masks typically work best, which is also the regime where the larger difference was observed in Fig. 2 between the paraxial and nonparaxial scalar results. Hence, this observation provides incentive for accounting for the nonparaxial effects deduced here when designing phase shifted masks.

V. CONCLUDING REMARKS

Scalar diffraction theory is an approximation to the actual physical decription of diffraction phenomena. Imposing the paraxial approximation represents an additional approximation imposed on this already approximate theory. Certainly in the regime of small NA, both these approximations are well justified. However, for higher values of NA, they become less accurate, and eventually simply inappropriate.

Our present formulation of the imaging of partially coherent light in an optical projection system removes the paraxial approximation, but still the approximation of scalar diffraction remains. Possibly this approach enables a

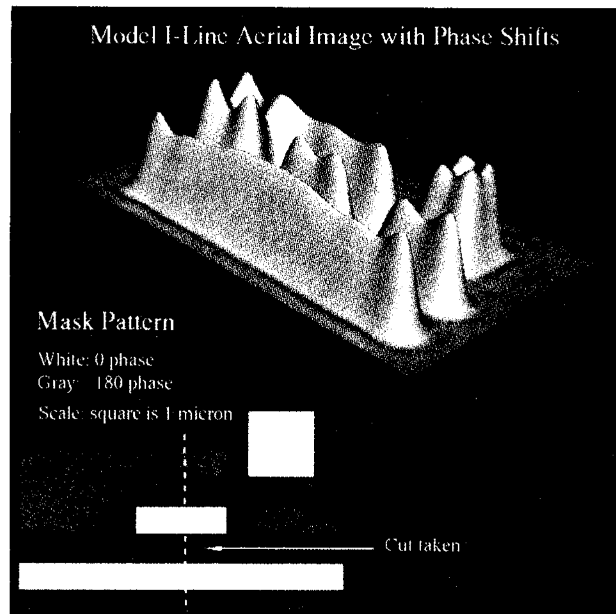

FIG. 3. A 2D aerial image of a relatively large phase shifted mask structure (3.5 μm$\times$6.0 μm). Here, $N_{or}=0.6$, $\sigma=0.2$, and $\lambda=0.365$ μm.

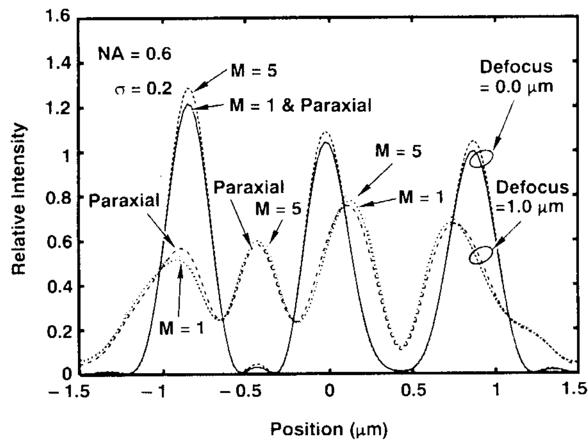

FIG. 4. Light intensity vs position is shown here for the cut in the mask indicated in Fig. 3. The grey (π) region of the mask in the indicated cut in Fig. 3 lies on the right-hand side of the plot shown here. Between the two 0° regions, at $x \approx -0.5$ μm, lies a small constructive interference region, which grows considerably upon going out of focus. Note that the $M=1$, $M=5$, and paraxial curves shift above and below each other across the defocus = 1 μm curve.

somewhat higher domain of NA values to be obtained via scalar diffraction theory; however, a full confirmation must wait until a detailed comparison is carried out with vector diffraction theory.

The easier analysis of scalar calculations *without* the paraxial approximation, versus vector calculations, bring several features to the forefront, which qualitatively at least are also predicted by vector aerial image calculations. In particular, we note that changes in magnification will alter the aerial image of diffraction limited, high NA systems. The largest difference was predicted here by scalar theory for the smaller partial coherence regime, precisely where the technology of phase-shift masks are most useful.

We should also note the critically important issue that simple scaling arguments become invalid for larger NA systems. More precisely, within the limits of the paraxial approximation, λ, $N_{\rm or}$, $z_{\rm I}$, and the object coordinates L_i can be altered to another set of values λ', $N'_{\rm or}$, $z'_{\rm I}$, and L'_i, and still yield precisely the same aerial image, provided that

$$L_i N_{\rm or}/\lambda = L'_i N'_{\rm or}/\lambda', \quad \forall \; L_i, L'_i, \tag{11}$$

and

$$z_{\rm I} N^2_{\rm or}/\lambda = z'_{\rm I} N'^2_{\rm or}/\lambda'. \tag{12}$$

This scaling argument is often used by lithographers in evaluating the probable printing effects of a projection system with a higher NA and (or) smaller λ, than what they presently have available. In this way, one can make predictions about projection systems not yet developed, by simply altering the mask dimensions.

Unfortunately, when higher NA effects are included, the above simple scaling transformation can be shown to no longer hold [compare Eqs. (5) and (9)], even before the vector character of the light is taken into account. Although the above scaling will still undoubtedly be used by lithographers as a first pass estimate of effects, clearly the true reliability of this argument holds only if one restricts $N_{\rm or}$ and $N'_{\rm or}$ to be less than about 0.4, [Fig. 2(a)], where first-order high NA effects were found to be negligible.

ACKNOWLEDGMENTS

Three of the authors (E. B., U. H., and S. A. O.) 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.

[1]H. H. Hopkins, Proc. R. Soc. London Ser. A **217**, 408 (1953).
[2]J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
[3]M. S. Yeung, Proc. SPIE **922**, 149 (1988).
[4]H. H. Hopkins, Proc. Phys. Soc. **55**, 116 (1943).
[5]B. Richards and E. Wolf, Proc. Phys. Soc. B **69**, 854 (1956).
[6]B. M. Watrasiewicz, Opt. Act. **12**, 391 (1965).
[7]V. B. Jipson, and C. C. Williams, Appl. Opt. **22**, 2202 (1983).
[8]D. G. Flagello and T. Milster, Proc. SPIE **1625**, 246 (1992).
[9]W. Ignatowski, Trans. Opt. Inst. Petrograd. **1**, paper IV (1919).
[10]E. Wolf, Proc. R. Soc. London Ser. A **253**, 349 (1959).
[11]B. Richards and E. Wolf, Proc. R. Soc. London Ser. A **253**, 358 (1959).
[12]A. Boivin and E. Wolf, Phys. Rev. B **138**, 1561 (1965).
[13]J. J. Stamnes, *Waves in Focal Regions* (Adam Hilger, Boston, 1986).
[14]M. Mansuripur, J. Opt. Soc. Am. A **3**, 2086 (1986).
[15]M. Mansuripur, J. Opt. Soc. Am. A **6**, 786 (1989).
[16]B. Küyel, E. Barouch, U. Hollerbach, and S. A. Orszag, Proc. SPIE **1674**, 376 (1992).
[17]D. G. Flagello and A. E. Rosenbluth, J. Vac. Sci. Technol. B **10**, 2997 (1992).
[18]D. C. Cole, E. Barouch, U. Hollerbach, and S. A. Orszag, Jpn. J. Applied Phys. (to be published).
[19]H. P. Urbach and D. A. Bernard, J. Opt. Soc. Am. A **6**, 1343 (1989).
[20]D. A. Bernard and H. P. Urbach, J. Opt. Soc. Am. A **8**, 123 (1991).
[21]C. M. Yuan and A. J. Strojwas, J. Opt. Soc. Am. A **8**, 778 (1991).
[22]E. Barouch, B. Bradie, U. Hollerbach, and S. A. Orszag, J. Vac. Sci. Technol. B **8**, 1432 (1990).
[23]E. Barouch, U. Hollerbach, S. A. Orszag, and M. Peckerar, IEEE Electron Device Lett. **12**, 513 (1991).
[24]J. W. Goodman, *Introduction to Fourier Optics* (McGraw–Hill, New York, 1968).
[25]M. Born and E. Wolf, *Principles of Optics*, 5th ed. (Pergamon, New York, 1975).
[26]E. C. Kintner, Appl. Opt. **17**, 2747 (1978).
[27]K. K. H. Toh, M.S. thesis, University of California at Berkeley, 1988.