

Lithographic Image Simulation for the 21st Century with 19th-Century Tools

Ronald L. Gordon^a, Alan E. Rosenbluth^b

^aIBM Microelectronics SRDC, 2070 Rt. 52, Hopewell Jct., NY 12533

^bIBM T.J. Watson Research Center, Yorktown Heights, NY 10598

ABSTRACT

Simulation of lithographic processes in semiconductor manufacturing has gone from a crude learning tool 20 years ago to a critical part of yield enhancement strategy today. Although many disparate models, championed by equally disparate communities, exist to describe various photoresist development phenomena, these communities would all agree that the one piece of the simulation picture that can, and must, be computed accurately is the image intensity in the photoresist.

The imaging of a photomask onto a thin-film stack is the only phenomenon in the lithographic process that is described fully by well-known, definitive physical laws. Although many approximations are made in the derivation of the Fourier transform relations between the mask object, the pupil, and the image, these and their impacts are well-understood and need little further investigation.

The imaging process in optical lithography is modeled as a partially-coherent, Köhler illumination system. As Hopkins has shown, we can separate the computation into 2 pieces: one that takes information about the illumination source, the projection lens pupil, the resist stack, and the mask size or pitch, and the other that only needs the details of the mask structure. As the latter piece of the calculation can be expressed as a fast Fourier transform, it is the first piece that dominates.

This piece involves computation of a potentially large number of complex quantities called Transmission Cross Coefficients (TCCs), which are correlations of the pupil function weighted with the illumination intensity distribution. The advantage of performing the image calculations this way is that the computation of these TCCs represents an up-front cost, not to be repeated if one is only interested in changing the mask features, which is the case in Model-Based Optical Proximity Correction (MBOPC).¹ The down side, however, is that the number of these expensive double integrals that must be performed increases as the square of the mask unit cell area; this number can cause even the fastest computers to balk if one needs to study medium- or long-range effects. One can reduce this computational burden by approximating with a smaller area, but accuracy is a concern, especially when building a model that will purportedly represent a manufacturing process.

This work will review the current methodologies used to simulate the intensity distribution in air above the resist and address the above problems. More to the point, a methodology has been developed to eliminate the expensive numerical integrations in the TCC calculations, as the resulting integrals in many cases of interest can be either evaluated analytically, or replaced by analytical functions accurate to within machine precision. With the burden of computing these numbers lightened, more accurate representations of the image field can be realized, and better overall models are then possible.

1. INTRODUCTION

Simulation of optical lithographic processes has been in development for nearly 30 years, although general acceptance of simulation tools within the lithography community has only happened over the past 8-10 years. Most of this is due to the inherent complexity of the resist chemistry and the resist development process. Although there have been great strides in understanding the physics of such phenomena as photoacid generation and migration,² practical lithography simulation demands that simplistic models be used. Further, parameters for such models may not represent truly physical entities, but are obtained through some best fit to one or more data sets.

rlgordon@us.ibm.com; Phone: (845) 892 2432

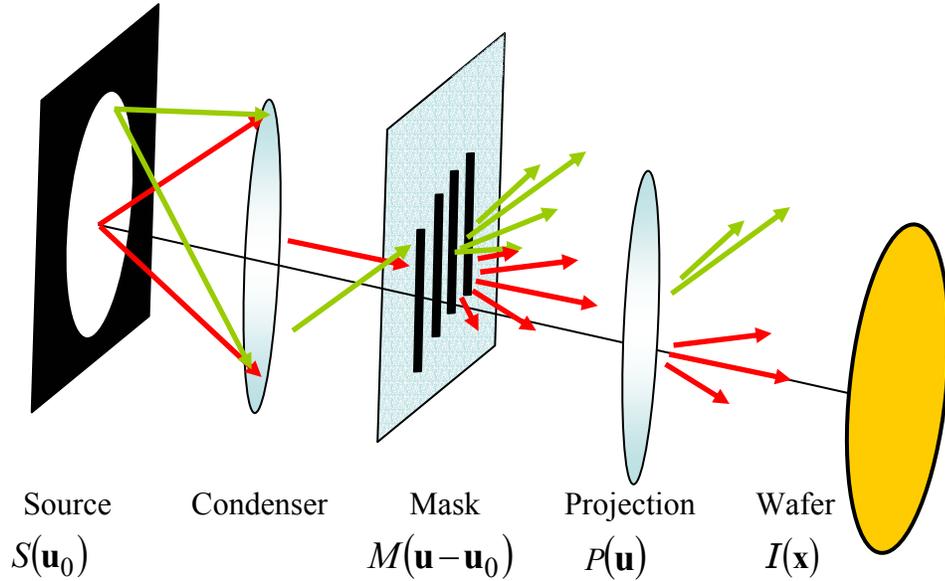


Figure 1. Schematic of a Köhler illumination system used in optical lithography.

Nevertheless, one piece of the overall model, the optical imaging model, has far more solid physics behind it.³ Because the steppers and scanners are designed as Köhler illumination systems, Hopkins's theory of imaging under partially coherent illumination⁴ applies. Consider the Köhler illumination system in Fig. 1. The output image only depends on the composition of the source, exit pupil, and photomask. The (secondary) source is typically some circular shape, most commonly a disk or an annulus, but can take on arcs and dipoles. The pupil contains all phase information, such as defocus and aberrations, but can also have amplitude variation caused by power conservation, vector and thin-film effects. Finally, the photomask is, for UV and Deep UV wavelengths, a transmissive structure with a bright (glass) or dark (chrome, or other attenuating material) background, composed of arbitrary polygons of opposite tone from the background that define some level (gate, metal, diffusion, etc.) in a semiconductor device to be manufactured.

The main premise in lithography is that the rate at which the photoresist develops is uniquely linked to the light intensity in the resist. Once the develop rate is known, one can predict the resist profile as a function of development time. Unfortunately, there is no relationship between the intensity in the resist and the develop rate derived from fundamental physical principles; all known models are empirical. Nevertheless, because such models, which require careful calibration against experimental data, must be extended to predict resist behavior of arbitrary patterns, accurate aerial image models are crucial to simulation.

How accurate? There are two types of accuracy: physical accuracy and numerical accuracy, and in most situations today, the typical lithographer, whose goal is to predict a priori experimental results under manufacturing conditions, cares more about the former than the latter (although they can both be equal in magnitude). For systems in air whose numerical aperture (NA) is close to unity, the depth of focus of an image predicted by the paraxial approximation (where the spherical phase curvature in the defocus term is replaced by a paraboloid) differs from that for an image computed without such an approximation by a factor of two.⁵ Worse, simpler models such as scalar and paraxial approximations tend toward optimistic predictions, so that worst-case scenario prediction is only made possible by the inclusion of more complicated models.

On the other hand, there are some practical scenarios where the simple scalar, paraxial approximation may be desirable. One must keep in mind that in lithographic applications, air is never the imaging medium; rather, it is a photoresist that has an index of refraction value ranging from 1.5 - 1.8. The angles in such a medium will always be smaller than in air, making paraxial approximations that much more valid. Furthermore, the aerial image is hardly ever the final output: some model, as discussed above, must be applied to this image in order to extract a linewidth or a resist profile. These models, even those most finely calibrated, possess some degree of error that can be on the order of that from employing the paraxial approximation.⁶ This is not to say that there is no value in applying more complex and accurate physical models to a simulator; there are times, however, when a ‘quick-and-dirty’ approximation will be enough.

The benefits of using the paraxial approximation may be somewhat unclear in the face of faster and cheaper computation. The main purpose of this paper, then, is to show that these benefits are real and dramatic. It will be shown that, for partially coherent, Köhler illumination systems in lithography, the paraxial approximation allows models to be expressed in terms of simple, analytical functions that require no time-consuming integrations in their evaluations. The impact of eliminating integrations from the simulation process cannot be overstated: it is common to see factors of 1000 or more in computational speed increases. Such speed increases enable explorations over much larger areas than possible with conventional methods. Furthermore, the paraxial approximation may be a foundation on which faster and more precise physical⁷ models may be built.

2. FINITE FORMULATION OF HOPKINS’ MODEL

2.1. Fundamentals

Consider a modern microlithographic imaging system, comprised of a (finite) source, condenser optics, a semi-transparent mask to be imaged, objective optics, and a silicon wafer in the image plane. In an ideal system, the source is imaged onto the entrance pupil of the objective lens by the condenser, and the mask is in the Fourier transform (focal) plane of the condenser. The electric field E from a coherent, on-axis source at the point \mathbf{x} is

$$E(\mathbf{x}) = \int_{\mathbb{R}^2} d^2\mathbf{u} M(\mathbf{u})P(\mathbf{u}) \exp\left(+i2\pi\frac{NA}{\lambda}\mathbf{u} \cdot \mathbf{x}\right), \quad (1)$$

where M is the mask diffraction spectrum (the Fourier transform of the mask transmission function), P is the exit pupil distribution (the Fourier transform of the point spread function), NA is the numerical aperture of the objective lens, and λ is the illumination wavelength. The vector quantity \mathbf{u} represents a scaled frequency coordinate in that the pupil function is zero when $|\mathbf{u}| > 1$. The idea here is that each point of the source is then mapped onto a unique plane wave, incident at some angle upon the mask. Further, a uniform source is mapped onto a uniform distribution of plane wave amplitudes upon the mask. To this end, the electric field from an off-axis point source σ is

$$E(\mathbf{x}, \boldsymbol{\sigma}) = \int_{\mathbb{R}^2} d^2\mathbf{u} M(\mathbf{u} - \boldsymbol{\sigma})P(\mathbf{u}) \exp\left(+i2\pi\frac{NA}{\lambda}\mathbf{u} \cdot \mathbf{x}\right). \quad (2)$$

The vector quantity $NA\boldsymbol{\sigma}$ represents the projection of a point on the unit sphere onto the mask plane; the notation is intentionally suggestive for a Köhler illumination system. That is, as NA denotes the numerical aperture, $\boldsymbol{\sigma}$ represents a point in the illuminator.

To simulate the effects of a partially coherent Köhler illumination system on the electric field intensity, the off-axis intensity is averaged over all of the plane waves generated by the illuminator.⁸ The relevant measured quantity, which is called the intensity, is proportional to the square magnitude of the electric field distribution, and this intensity then takes the form

$$I(\mathbf{x}) = \frac{\int_{\mathbb{R}^2} d^2\boldsymbol{\sigma} S(\boldsymbol{\sigma})|E(\mathbf{x}, \boldsymbol{\sigma})|^2}{\int_{\mathbb{R}^2} d^2\boldsymbol{\sigma} S(\boldsymbol{\sigma})}, \quad (3)$$

where S is the source distribution for the illumination system. That is, as mentioned above, the intensity is an average of the intensities emanating from the fields generated by the various point sources that comprise the illuminator. Substituting Eq. (2) into Eq. (3), the following expression results:

$$I(\mathbf{x}) = \int_{\mathbb{R}^2} d^2\mathbf{u}' \int_{\mathbb{R}^2} d^2\mathbf{u}'' M(\mathbf{u}')T(\mathbf{u}', \mathbf{u}'')M^*(\mathbf{u}'') \exp \left[+i2\pi \frac{NA}{\lambda}(\mathbf{u}' - \mathbf{u}'') \cdot \mathbf{x} \right], \quad (4)$$

where the transmission cross-coefficient T is given by

$$T(\mathbf{u}', \mathbf{u}'') = \int_{\mathbb{R}^2} d^2\boldsymbol{\sigma} S(\boldsymbol{\sigma})P(\mathbf{u}' + \boldsymbol{\sigma})P^*(\mathbf{u}'' + \boldsymbol{\sigma}). \quad (5)$$

Eqs. (4) and (5) are the fundamental relations on which the remainder of this section is based.

2.2. Periodic Unit Cells

A fundamental assumption in most lithography simulators is that each pattern to be simulated is part of a larger, periodic array. Without such an assumption about the global behavior of the mask, most aerial image computations would be impractical in both storage and speed. Consider a semitransparent, periodic object with x - and y -periods p_x and p_y , respectively. Such an object is represented mathematically as

$$M(\mathbf{u}) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} \delta \left(u_x - \frac{m}{p_x} \frac{\lambda}{NA} \right) \delta \left(u_y - \frac{n}{p_y} \frac{\lambda}{NA} \right), \quad (6)$$

where the c_{mn} are the mask Fourier coefficients. (Of course, the entire spectrum is not used in determining the image, as the objective lens filters out the higher-order coefficients.) When Eq. (6) is substituted into Eq. (4), a series for the intensity results:

$$I(\mathbf{x}) = \sum_{m'=-\infty}^{\infty} \sum_{m''=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} \sum_{n''=-\infty}^{\infty} c_{m'n'} c_{m''n''}^* T_{m'n'm''n''} \exp \left\{ +i2\pi \left[(m' - m'') \frac{x}{p_x} + (n' - n'') \frac{y}{p_y} \right] \right\}, \quad (7)$$

where

$$T_{m'n'm''n''} = T \left(\frac{m'}{q_x}, \frac{n'}{q_y}, \frac{m''}{q_x}, \frac{n''}{q_y} \right), \quad (8)$$

and

$$(q_x, q_y) = (p_x, p_y) \frac{NA}{\lambda}. \quad (9)$$

The form for the intensity in Eq. (7) does not seem to offer any clear advantage over the ray-based method. In principle, one could sum Eq. (7) directly, as the quantity $T_{m'n'm''n''}$ is bound to vanish when the indices are large enough, making the sum finite. To see how this works, we take a closer look at the definition of this coefficient in Eq. (7). The pupil function and source distribution function has finite support; that is, $|\mathbf{u}| > 1 \Rightarrow P(\mathbf{u}) = 0$ and $|\mathbf{u}| > \sigma_{\max} \Rightarrow S(\mathbf{u}) = 0$. This then reduces the infinite integration interval specified in Eqs. (4) and (5) into a finite interval, and this is illustrated in Figure 2 below.

From this point on, the source is assumed to be circular, with radius σ_0 . The summation limits in Eq. (7) are now deduced from the bounds of the transmission cross-coefficients. Two geometrical conditions must be met for nonzero values: first, the separation between the two offset pupils must be less than twice their radius, or 2 in this coordinate scale, and second, both pupils must touch the source. Letting $k = m' - m''$ and $l = n' - n''$, these conditions translate into the following:

$$|\mathbf{u}' - \mathbf{u}''| \leq 2 \Rightarrow \frac{k^2}{q_x^2} + \frac{l^2}{q_y^2} \leq 4, \quad (10)$$

$$\max\{|\mathbf{u}'|, |\mathbf{u}''|\} \leq 1 + \sigma_0 \Rightarrow \max \left\{ \frac{m'^2}{q_x^2} + \frac{n'^2}{q_y^2}, \frac{m''^2}{q_x^2} + \frac{n''^2}{q_y^2} \right\} \leq (1 + \sigma_0)^2. \quad (11)$$

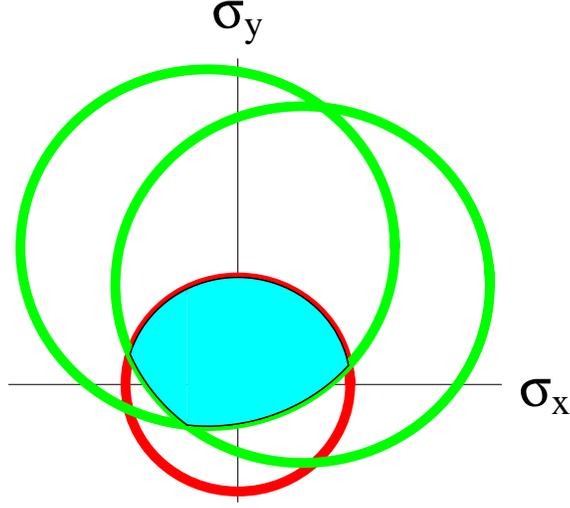


Figure 2. Geometry for computing the transmission cross-coefficient. The region of nonzero pupil values is represented by either of the larger circles, while that of nonzero source distribution is the smaller. The presence of two larger circles offset from both the origin and each other is a consequence of the convolution form of the integral in Eq. (5); the offsets are equal to the spatial frequencies. The transmission cross-coefficient is then equal to the integral over the intersection of these 3 regions and is represented by filled region.

The conditions in Eqs. (10) and (11), and the change of variable suggested above, lead directly to the following exact finite sum for the intensity. Denoting $\lfloor X \rfloor$ as the greatest integer less than X :

$$I(\mathbf{x}) = \sum_{k=-M_x}^{M_x} \sum_{l=-M_y(k)}^{M_y(k)} C_{kl} \exp \left[+i2\pi \left(k \frac{x}{p_x} + l \frac{y}{p_y} \right) \right], \quad (12)$$

where

$$M_x = \lfloor 2q_x \rfloor, \quad (13)$$

$$M_y(k) = \left\lfloor 2q_y \sqrt{1 - \frac{k^2}{4q_x^2}} \right\rfloor, \quad (14)$$

and

$$C_{kl} = \sum_{m=-N_x}^{N_x-k} \sum_{\substack{N_y(m,k)-l[l>0] \\ -l[l<0]-N_y(m,k)}} c_{m+k,n+l} c_{mn}^* T_{m+k,n+l,m,n}, \quad (15)$$

$$N_x = \lfloor (1 + \sigma_0)q_x \rfloor, \quad (16)$$

$$N_y(m, k) = \left\lfloor (1 + \sigma_0)q_y \sqrt{1 - \frac{(m+k)^2}{(1 + \sigma_0)^2 q_x^2}} \right\rfloor, \quad (17)$$

and where $\lfloor \mathcal{L} \rfloor$ is 1 if the logical condition \mathcal{L} is true, and 0 if \mathcal{L} is false.

Note that all of the information pertaining to the optical system is contained in the TCC's. Because the TCC's are computed via the integrations over the source, they can be stored in an array for a very efficient mask perturbation scheme. That is, as long as one only varies the mask features (without changing the pitch), then

the TCC's need only be computed once; the rest of the image computation is typically done in a small fraction of the time it took to compute the TCC's. Therefore, this frequency-space approach is ideal for optical proximity correction, in which layouts are optimized for optical diffraction and lithography process conditions.

3. ANALYTICAL EXPRESSION FOR THE IN-FOCUS AERIAL IMAGE

3.1. Derivation of the Geometrical Configurations

The computation of the in-focus TCC is reduced to simply computing the common area of 3 circles, 2 of which are offset from the origin and have unit radius (the pupils), while the other is centered at the origin and has radius of σ_0 (the source), as pictured in Figure 1. Kintner⁹ showed that, for the 1D case where the centers of the pupils are constrained to lie on the horizontal (or vertical) axis, and where $\sigma_0 < 1$, the TCC geometry takes the form of one of 4 distinct configurations. If we denote the intersection of the offset pupils by D and the source by S, then these configurations are as follows:

- S is completely contained within D.
- S protrudes out of one side of D.
- S protrudes out of both sides of D.
- D is completely contained within S.

Clearly, the situation for the general 2D mask is considerably more complicated. Up until now, no published work even enumerated the number of distinct geometrical configurations available; the typical approach to computing the 2D TCC has been to employ the minimum grid that contains the pupils and the source. It will be shown below, however, that, taking into account all possible symmetries, there are 18 distinct geometrical configurations to consider. We begin by observing that the coordinate system can always be defined such that the symmetry axis of the intersection region is vertical. By performing this rotation, the myriad of possible geometrical combinations is limited, and only a single vertical coordinate is necessary to specify the TCC. Organizing the different cases is also simplified, as the 2D TCC can now be viewed as a perturbation of the 1D TCC. The various geometries will then be organized by their configuration with the y-components of the frequencies set to zero.

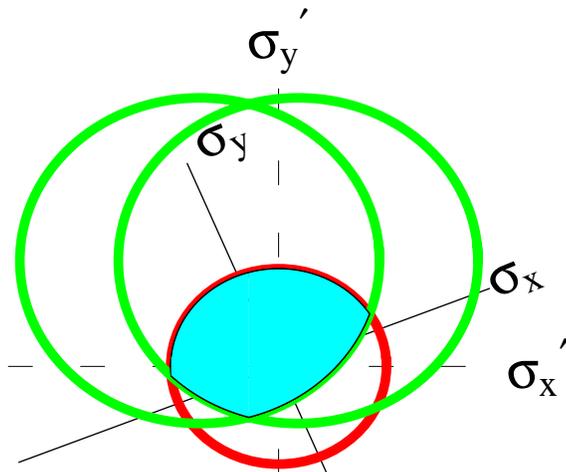


Figure 3. The TCC region in the rotated coordinate system.

Basically, one must track the vertical motion of the region D through the source, taking note of all critical points beyond which the nature of the integration region changes. The number of logical conditions easily

numbers in the hundreds; therefore, an exhaustive description of the algorithm is not possible here. To simplify the programming, it is essential that the unique regions be identified and mapped to the correct set of logical conditions. These unique regions and examples of the logical conditions that give rise to them have been enumerated in a previous work.¹⁰

3.2. Simplification Via Application of Stokes' Theorem

The main result of the previous subsection was to identify all of the possible geometrical forms for the region of integrations for the TCCs, and the conditions under which each particular form would occur. The in-focus TCCs are simply the areas of their respective regions. Given that each of the bounding contours of these regions is a circular arc, the determination of the areas exactly in terms of analytic functions is straightforward. All one needs are the locations of the intersection points in order to determine where a change in the integration limits is warranted.

On the other hand, the resulting functions that describe the TCC are, from a programming perspective, rather unwieldy. Even though orders of magnitude efficiency over straightforward numerical integration is still possible, one gets the feeling that further improvements are possible. Such an improvement would result if there were a simple, universal expression that could act as a building block for the computation of the area.

Such a building block is generated by integrating over the contour rather than the area. The mathematical tool for converting between contour and area integrals is Green's Theorem. A particularly simple consequence of Green's Theorem is an expression of the area A of a closed region D in terms of its bounding contour:

$$A(D) = \int_D d^2\boldsymbol{\sigma} = \frac{1}{2} \oint_{\partial D} \sigma_x d\sigma_y - \sigma_y d\sigma_x. \quad (18)$$

Because the closed contour consists of nothing more than a set of circular arcs, the parametrization of the above contour integral is trivial; the final result takes the form

$$A(D) = \frac{1}{2} \sum_{n=1}^N \left\{ \rho_k^2 (\phi_2^{(k)} - \phi_1^{(k)}) + \rho_k \left[u_0^{(k)} (s_2^{(k)} - s_1^{(k)}) - v_0^{(k)} (c_2^{(k)} - c_1^{(k)}) \right] \right\}, \quad (19)$$

where the following notation applies:

- ρ_k is the radius, and $(u_0^{(k)}, v_0^{(k)})$ is the center of the circle containing the k^{th} arc.
- $\phi_1^{(k)}$ and $\phi_2^{(k)}$ are the upper and lower angles, respectively, subtended by the k^{th} arc at its circle's center.
- $s_j^{(k)} = \sin \phi_j^{(k)}$, $c_j^{(k)} = \cos \phi_j^{(k)}$
- N is the number of arcs in the contour.

The angular quantities $\phi_1^{(k)}$ and $\phi_2^{(k)}$ are found by consideration of the intersections between one of the offset pupils and the source. These intersection pairs are found easily from the equations of their respective circles. The horizontal coordinates of these points are written as follows:

$$u_{\pm} = \frac{u [u^2 + v^2 - (1 - \sigma_0^2)] \pm v \sqrt{4(u^2 + v^2) - [u^2 + v^2 + (1 - \sigma_0^2)]^2}}{2(u^2 + v^2)}, \quad (20)$$

where the quantity u is $u_0^{(k)}$ and v is $v_0^{(k)}$. The following expression is then used to determine the correct sign for the vertical intersection coordinate:

$$v_{\pm} = \frac{u^2 + v^2 - (1 - \sigma_0^2)}{2v} - \frac{u}{v} u_{\pm}. \quad (21)$$

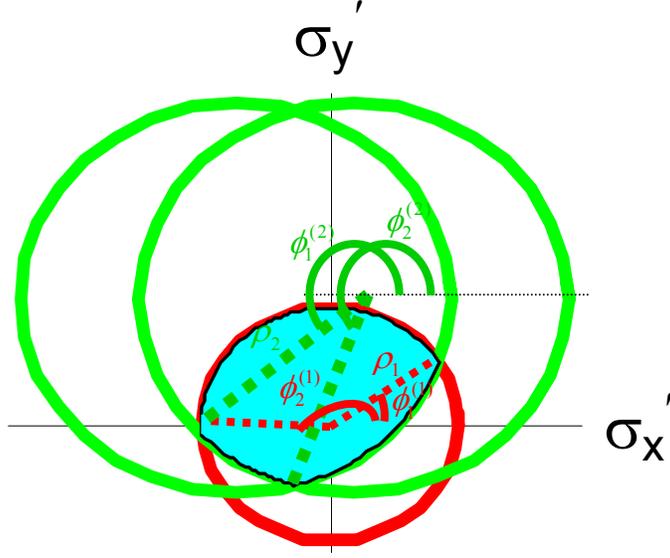


Figure 4. Illustration of the various parameters in Eq. (19).

The angles $\phi_1^{(k)}$ and $\phi_2^{(k)}$ are then computed from the following quantities:

$$c_{\pm} = \frac{u_{\pm} - u_0^{(k)}}{\rho_k}, \quad (22)$$

$$s_{\pm} = \frac{v_{\pm} - v_0^{(k)}}{\rho_k}, \quad (23)$$

$$\phi_{\pm} = \tan^{-1} \frac{s_{\pm}}{c_{\pm}}. \quad (24)$$

The \pm subscripts are mapped to 1 or 2 through consideration of the geometry of the TCC.

There are two immediate advantages to the contour integrations method:

- The number of unique geometrical configurations shrinks from 18 to 9, simplifying the computer code describing the algorithm.
- The final expression in Eq. (19) is evaluated with a single arctangent and a few square roots per arc, resulting in a more efficient algorithm.

4. EXTENSION TO DEFOCUSED IMAGES

4.1. Preliminary Development

The imaging algorithm outlined above is for the in-focus case only. Understanding the tolerance of a lithography process, however, requires the computations of out-of-focus images. These computations, without any mathematical simplifications, are far more computationally demanding than those for the in-focus case. We will show that, if the paraxial approximation is allowed, then vector calculus can be harnessed to produce a finite algorithm that will greatly speed up image computations.

Defocus is expressed as a pure phase pupil function. Here, for convenience, we express this phase minus a constant phase that contributes nothing to the image intensity:

$$P(\mathbf{u}) = \exp \left[-i2\pi \frac{z}{\lambda} \frac{NA^2 \mathbf{u}^2}{1 + \sqrt{1 - NA^2 \mathbf{u}^2}} \right]. \quad (25)$$

Eq. (25) represents an exact expression for the defocused pupil response for a telecentric system in air, valid for all $0 < NA < 1$, minus the constant phase term. Valid as it may be for high-NA lithographic systems, it leads

to unwieldy double integrals in the TCC expression that must be evaluated numerically and have no hope of the sort of simplification used in Eq. (19).

On the other hand, as we will see below, the paraxial approximation, which replaces the square root in the denominator of Eq. (25) with unity, will lead to an integral representation of the TCC that can be simplified using the Stokes Theorem. That is, the pupil response function takes the following form:

$$P(\mathbf{u}) = \exp \left[-i\pi \frac{z}{\lambda} NA^2 \mathbf{u}^2 \right]. \quad (26)$$

Generally speaking, this approximation is valid for $NA \lesssim 0.4$. In semiconductor lithography, however, we image in photoresist, which typically has an index of refraction in the 1.5 - 1.8 range, increasing the range of validity to $NA \lesssim 0.7$. Of course, more precise imaging models would take the variation of transmission and polarization with angle of incidence on the photoresist; nevertheless, at the level of approximation sought here, these effects can be neglected.

When a circular source is assumed, Eq. (26) is inserted into Eq. (5). Note that the terms quadratic in phase in the integration variable cancel, and the TCC is now expressed in terms of a Fourier-type integral:

$$T(\mathbf{u}', \mathbf{u}'') = \exp(-i\pi b) \int_D d^2\sigma \exp(-i2\pi \mathbf{a} \cdot \sigma), \quad (27)$$

where

$$\mathbf{a} = -\frac{NA^2}{\lambda} z (\mathbf{u}' - \mathbf{u}''), \quad (28)$$

$$b = \frac{NA^2}{\lambda} z (\mathbf{u}'^2 - \mathbf{u}''^2). \quad (29)$$

It is clear that the paraxial approximation allows for a much simpler mathematical form for the TCC. Left as is, however, Eq. (27) provides little real speed improvement over that which results from the full defocus expression in Eq. (25). Therefore, an expression similar in spirit to that in Eq. (19) is sought, where a reduction to a 1D integral will lead us to analytical expressions that can produce the desired results.

The complication here is that there is now a varying function inside the integral in Eq. (27). One way to deal with this is to find a function whose curl is the integrand of the integral in Eq. (27). Stoke's theorem states that, for a vector function \mathbf{F} of a vector variable \mathbf{x} that has at least continuous second derivatives, then the net flux of the curl of \mathbf{F} through a surface is equal to the line integral of \mathbf{F} over the (closed) boundary of that surface:

$$\int_{\mathcal{P}} d^2\mathbf{x} \hat{\mathbf{n}} \cdot \nabla \wedge \mathbf{F} = \oint_{\partial\mathcal{P}} d\mathbf{x} \cdot \mathbf{F}. \quad (30)$$

The immediate problem becomes the following: given $\nabla \wedge \mathbf{F}$, determine \mathbf{F} . In general, one also needs the divergence of \mathbf{F} in order to fully determine \mathbf{F} ; in this case, we choose \mathbf{F} such that $\nabla \cdot \mathbf{F} = 0$ so that we may proceed as follows. Although the integrand in Eq. (27) is scalar, it can be put in vector form by defining a coordinate system such that the normal to \mathcal{P} lies along the $\hat{\mathbf{z}}$ direction. That is, we write

$$\nabla \wedge \mathbf{F} = \exp(-i2\pi \mathbf{a} \cdot \sigma) \hat{\mathbf{z}}. \quad (31)$$

Clearly, \mathbf{F} will have the same exponential form as in Eq. (31), and, therefore, will be a solution of the vector wave equation:

$$\nabla \wedge \nabla \wedge \mathbf{F} = (2\pi \mathbf{a} \cdot \sigma)^2 \mathbf{F}. \quad (32)$$

Therefore, by taking the curl of both sides in Eq. (31), and using Eq. (32), the following expression for \mathbf{F} is deduced:

$$\mathbf{F} = -\frac{1}{-i2\pi} \frac{a_y \hat{\sigma}_x - a_x \hat{\sigma}_y}{a_x^2 + a_y^2} \exp(-i2\pi \mathbf{a} \cdot \sigma). \quad (33)$$

The integral over the intersection of the offset pupils and the source is now written as a single contour integral over the boundary of that region. Again, the integral over this closed contour is evaluated by parametrization over the circular arcs. Proceeding in exactly the same way as in the in-focus case above, we now express the defocused TCC in terms of a series of one-dimensional integrals:

$$T(\mathbf{u}', \mathbf{u}'') = \frac{\exp(-i\pi b)}{-i2\pi|\mathbf{a}|} \sum_{k=1}^N \rho_k \exp \left[-i2\pi \left(a_x u_0^{(k)} + a_y v_0^{(k)} \right) \right] \int_{\theta+\phi_1^{(k)}}^{\theta+\phi_2^{(k)}} d\phi \cos \phi \exp(-i2\pi|\mathbf{a}|\rho_k \cos \phi), \quad (34)$$

where $\theta = \tan^{-1} \frac{a_y}{a_x}$. It turns out that, in this rotated coordinate system, both a_y and θ are zero.

Eq. (34) expresses the defocused TCC explicitly in terms of a set of single integrals, each of which is expressible in terms of the following universal integral:

$$J(x, \phi_1, \phi_2) = \int_{\phi_1}^{\phi_2} d\phi \cos \phi \exp(-ix \cos \phi). \quad (35)$$

Eq. (34) represents the most simplified, exact form for the paraxially defocused TCC, as the integral in Eq. (35) has no exact, analytical representation in general. Similar integrals were found in studying the one-dimensional case, and these were evaluated numerically.¹¹ While this reduction from the usual double integral to this set of single integrals represents a dramatic reduction in computation, these integrals leave much to be desired. If left as is for numerical integration, we will be required to make many relatively expensive trigonometric function calls. Further, such a numerical integration will be very slow if done adaptively in order to achieve a given accuracy.

The first step toward speeding up this calculation, then, is eliminating the trig function calls altogether. This is done by a substitution and integration by parts:

$$J(x, \phi_1, \phi_2) = \sin \phi_2 \exp(-ix \cos \phi_2) - \sin \phi_1 \exp(-ix \cos \phi_1) + ix \int_{\cos \phi_1}^{\cos \phi_2} du \sqrt{1-u^2} \exp(-ixu). \quad (36)$$

Eq. (36) would, in fact, lead to a faster numerical integration scheme, although some care is necessary near the branch point at $u = 1$.

It turns out, however, that we can speed up the evaluation of the TCC further still by avoiding all numerical integration in Eq. (36). This is achieved by recognizing that, although there is no exact analytical representation of the integral in the right-hand-side of Eq. (36), we can find a function h of u that is very close to the $\sqrt{1-u^2}$ term inside the integral for all $u \in [0, 1]$; that is, $|\sqrt{1-u^2} - h(u)| < \epsilon$ for some small ϵ . Such a function would be chosen such that this integral can be expressed in terms of simple, analytical functions. The resulting integral would then satisfy

$$\left| \int_{\cos \phi_1}^{\cos \phi_2} du \left[\sqrt{1-u^2} - h(u) \right] \exp(-ixu) \right| < \epsilon |\cos \phi_2 - \cos \phi_1|. \quad (37)$$

4.2. Analytical Evaluation of an Integral

In this section, such a function h is found. Also, a more practical implementation of Eq. (36) is considered there for all possible integration limits. It turns out that the resulting integral is then expressed in terms of a complex polynomial in u and x , 4 trig function evaluations and 2 complex Fresnel integrals. The Fresnel integrals are in fact well-known, analytical functions which can be computed with few arithmetic operations.¹² Such an analytical representation of the TCC, then, is what we desire for subsequent computations.

We refer to the function

$$F(x, u_0) = \int_1^{u_0} du \sqrt{1-u^2} \exp(-ixu). \quad (38)$$

Many methods exist for approximate representation of arbitrary functions, e.g. Taylor/Laurent series, Padé approximant, etc. Nevertheless, the function in the integrand of Eq. (38), although simple, has a branch point at $u = \pm 1$ that does not make it amenable to polynomial approximation there. It turns out, however, that the branch point at $u = 1$ can be factored out if we consider only the interval $u \in [0, 1]$. That is, we can make the following representation:

$$\sqrt{1-u^2} = \sqrt{1-u} \left[1 + \sum_{q=1}^Q b_q u^q \right] + \epsilon. \quad (39)$$

Because of the finite interval of interest, the coefficients b_q may be found by a uniform sampling across this interval. It turns out that the value $Q = 4$ is sufficient to obtain $\epsilon < 10^{-5}$.

The expression on the right-hand side of Eq. (39) can be simplified further through the following equations for $u_0 \in [0, 1]$:

$$\int_1^{u_0} du \sqrt{1-u} \exp(-ixu) = -i \exp(-ix) x^{-3/2} \sqrt{\frac{\pi}{2}} \mathcal{F} \left[\sqrt{\frac{2}{\pi}} \sqrt{x(1-u_0)} \right] + i \exp(-ixu_0) x^{-1} \sqrt{1-u_0}, \quad (40)$$

$$\mathcal{F}(z) = \int_0^z dt \exp\left(i \frac{\pi}{2} t^2\right), \quad (41)$$

$$\int_1^{u_0} du u^q \sqrt{1-u} \exp(-ixu) = i^q \frac{\partial^q}{\partial x^q} \int_1^{u_0} du \sqrt{1-u} \exp(-ixu). \quad (42)$$

It turns out that this expression for the integral becomes invalid for small x , i.e. $x < 0.1$. (The reason for this can be attributed to the use of integration by parts earlier.) Such a breakdown is easily addressed by a Taylor series of the integrand about $x = 0$ out to, say, third order. That is,

$$F(x, u_0) = \sqrt{1-u_0^2} \left[\frac{1}{2} u_0 \left(1 - \frac{1}{8} (2u_0^2 - 1) x^2 \right) + i \left(\frac{1}{3} (1-u_0^2) x + \frac{1}{90} (3u_0^4 - u_0^2 - 2) x^3 \right) \right] - \frac{1}{2} \left(1 - \frac{1}{8} x^2 \right) \cos^{-1} u_0 + \epsilon, \quad (43)$$

and $|\epsilon| < 10^{-6}$ when $|x| < 0.1$.

With these expressions, we have the desired expression for the TCC. The error involved in this approximation is illustrated in Fig. (5). It is clear that this approximation results in at least 5 decimal places of accuracy for all possible values of x and u_0 and is a valid substitution for the numerical integral. Moreover, because there are far less arithmetic operations for each TCC in this analytical representation than that in the direct numerical approach, the time needed to compute paraxial images through defocus will be greatly reduced.

5. ILLUSTRATION OF THE THEORY

All the pieces are now in place for the numerical evaluation of the TCCs. This, along with the mask spectrum for an arbitrary set of polygons, which is derived in the appendix exactly in terms of analytical functions, is all that is needed to compute the aerial image. We now proceed with such a computation.

First of all, in order to get an idea of how much time savings this analytical technique provides, we create a table of computation time (in seconds) for the TCCs using both this method and a purely numerical method

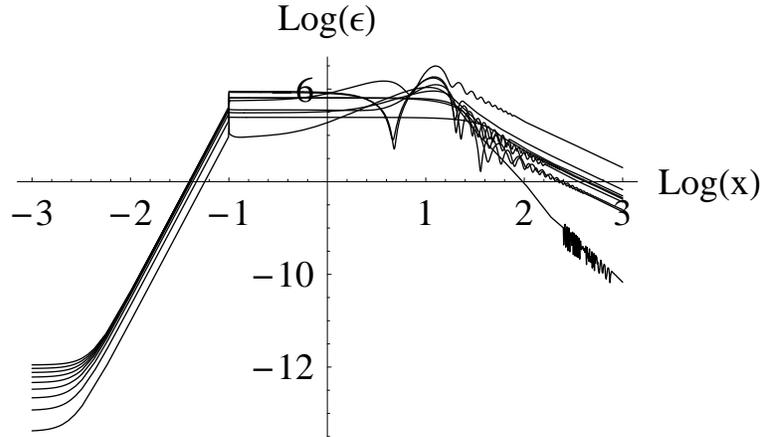


Figure 5. Log-log plot of the error in $F(x, u_0)$ for as a function of x for various values of u_0 .

over 3 focal planes on a 1.8 GHz processor with 512 MB of RAM. [This example uses $\lambda = 0.19337 \mu m$, $NA = 0.5$, $\sigma = 0.6$, and defocus = 0.0, 0.25, and $0.5 \mu m$.]

Area (μm^2)	T(analytical)	T(numerical)
0.5	0.06	49.4
1.0	0.121	113
2.0	0.440	436

It is clear that the analytical computation is taking on the order of $\frac{1}{1000}$ of the time it takes for the numerical calculation to complete. Further, it has been shown that the number of TCCs varies as the square of the area of the periodic unit cell,¹⁰ it is clear that for larger areas, the numerical method becomes unwieldy.

We now apply the theory to a large pattern that is seen in a semiconductor manufacturing environment. With the same optical parameters above, we consider a dark-field mask, with a background that is an attenuated phase-shift material. (That is, intensity transmittance of about 0.06, and a phase of 180 degrees.) The individual polygons then represent clear areas in the mask (see the Appendix). We compute the aerial image of such a pattern at a particular intensity threshold (or dose level).

This pattern is about $23.5 \mu m^2$ in area, or about $628\lambda^2$, and the TCCs took about 152 seconds to generate. From the previous estimate, the numerical computation would have taken over 42 hours.

6. CONCLUDING REMARKS

This work has shown that, not only in-focus imagery, but also paraxial defocused imagery can be computed for arbitrary circuit patterns very rapidly - on the order of 1000 times a purely numerical calculation. This has been done without being limited by any serious numerical error - the results here are only limited by the physical approximations imposed by the necessity for mathematical simplicity.

Of course, this work may seem, on the face of it, somewhat of an anachronism. We have been trumpeting the paraxial approximation, valid only for low angles, while the industry is rapidly moving toward the use of very high NA optics (> 0.8). The point of this work, then, is not to offer a possible simulator for exploration of lithographic technologies for such high NA systems. Rather, it provides a backup solution in the face of the necessity to simulate larger and larger patterns in order to preserve simulation accuracy. Here is a method that is

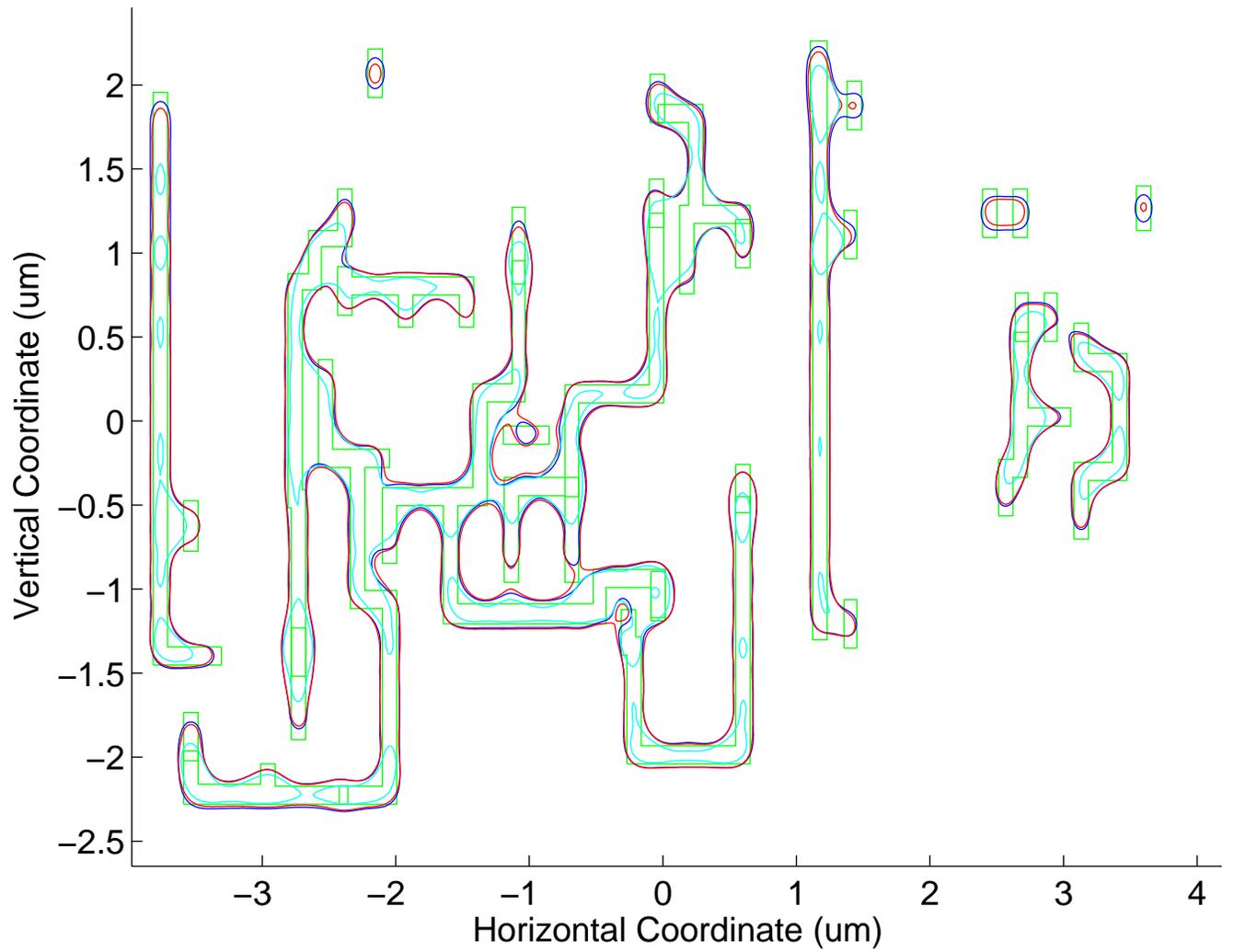


Figure 6. Aerial image of a large pattern seen in a semiconductor environment, both in-focus (blue) and for 0.25 (red) and 0.5 (cyan) μm defocus.

capable of computing these large patterns in a reasonable amount of time - can it be extended to more accurate physical models at larger value of NA?

This, then, is a work in progress. Non-uniform sources, vector diffraction, non-uniform transmittance into the photoresist layer, non-paraxial defocus, and aberrations are all important considerations as we attempt to image smaller linewidths across large fields. This work may be extensible if these effects can be expressed in terms of simple algebraic transformations of the paraxial results derived here. If this is possible, then larger dimensional scales may be explored than ever previously, leading to a more flexible MBOPC or more accurate design for manufacture.

APPENDIX A. ANALYTICAL EXPRESSION FOR THE MASK SPECTRUM

A typical level of a circuit design to be printed can be considered as a collection of a large number \mathcal{N} of arbitrarily complex polygons. Each polygon \mathcal{P}_k is taken to have a constant, complex transmittance t_k , and is immersed in a background of transmittance t_B . The mask transmittance function of such a setup is given by

$$m(\mathbf{x}) = t_B - \sum_{k=1}^{\mathcal{N}} (t_B - t_k) [\mathbf{x} \in \mathcal{P}_k], \quad (44)$$

Because of the linearity of the Fourier transform, only a single polygon need be treated in this analysis; Eq. (44) can then be invoked to get the total spectrum.

The mask spectrum is given by the Fourier transform relation

$$M(\mathbf{u}) = \int_{\mathbb{R}^2} d^2\mathbf{x} m(\mathbf{x}) \exp(-i2\pi \frac{NA}{\lambda} \mathbf{u} \cdot \mathbf{x}). \quad (45)$$

Consider a polygonal region \mathcal{P} of unit transmittance embedded in a perfectly opaque background (i.e. zero transmittance). The contribution to the spectrum from \mathcal{P} is then

$$M(\mathbf{u}) = \int_{\mathcal{P}} d^2\mathbf{x} \exp(-i2\pi \frac{NA}{\lambda} \mathbf{u} \cdot \mathbf{x}). \quad (46)$$

The polygon \mathcal{P} is completely specified by its P vertices $\{\mathbf{x}_1, \dots, \mathbf{x}_P\}$, and this representation is crucial in the transformation of the expression in Eq. (46) from an area integral to one over a boundary. Since a polygon simply connects adjacent vertices with straight lines, such a transformation of this integral could simplify the integration, even if the resulting integrand were more complicated.

Note that Eq. (46) is exactly of the same form as that in Eq. (27), and is therefore amenable to the same sort of analysis. Instead of a series of circular arcs around a closed contour, however, there is a series of line segments around the closed contour.

The spectrum now takes the following form:

$$M(\mathbf{u}) = -\frac{1}{i2\pi|\mathbf{u}|^2} \oint_{\partial\mathcal{P}} d\mathbf{x} \cdot (u_y \hat{\mathbf{x}} - u_x \hat{\mathbf{y}}) \exp(-i2\pi \frac{NA}{\lambda} \mathbf{u} \cdot \mathbf{x}). \quad (47)$$

The n th edge is a straight line connecting the point \mathbf{x}_n to the point \mathbf{x}_{n+1} ; it is parametrized by the variable $t \in \{0, 1\}$ and satisfies $\mathbf{x}(t) = (1-t)\mathbf{x}_n + t\mathbf{x}_{n+1}$ and $d\mathbf{x} = (\mathbf{x}_{n+1} - \mathbf{x}_n)dt$. The integration is then straightforward, and the final result is simply a sum over all the edges. If we define $\mathbf{x}_{N+1} = \mathbf{x}_1$, then an explicit expression for the spectrum is

$$M(\mathbf{u}) = \frac{\lambda}{NA} \frac{1}{i2\pi|\mathbf{u}|^2} \sum_{n=1}^{\mathcal{N}} [u_y(x_{n+1} - x_n) - u_x(y_{n+1} - y_n)] \exp[i\pi \frac{NA}{\lambda} (\mathbf{x}_{n+1} + \mathbf{x}_n) \cdot \mathbf{u}] \text{sinc}[\pi \frac{NA}{\lambda} (\mathbf{x}_{n+1} - \mathbf{x}_n) \cdot \mathbf{u}]. \quad (48)$$

Eq. (48) is the desired explicit expression for the spectrum of an arbitrary polygon.

Acknowledgments

The authors would like to thank Gregg Gallatin, Tim Brunner, Martin Burkhardt, Carlos Fonseca, and Nakgeoun Seong for useful discussions. They would also like to thank Mike Lercel and George Gomba for their continued managerial support.

REFERENCES

1. N. Cobb and A. Zakhor, "Fast low-complexity mask design," *Proc. SPIE* **3334**, p. 313, 1995.
2. F. Houle, W. Hinsburg, M. Morrison, M. Sanchez, G. Wallraff, C. Larson, and J. Hoffnagle, "Determination of coupled acid catalysis-diffusion processes in a positive-tone chemically amplified photoresist," *J. Vac. Sci. Tech. B* **18**, p. 1874, 2000.
3. M. Born and E. Wolf, *Principles of Optics*, Cambridge Univ. Press, Cambridge, 7th ed., 1999.
4. H. Hopkins, "On the diffraction theory of optical images," *Proc. Royal Soc. A* **217**, p. 408, 1953.
5. C. Sheppard and H. Matthews, "Imaging in high-aperture optical systems," *J. Opt. Soc. Am. A* **4**, p. 1354, 1987.
6. G. Forbes, "Validity of the Fresnel approximation in the diffraction of collimated beams," *J. Opt. Soc. Am. A* **13**, p. 1816, 1996.
7. G. Forbes, D. Butler, R. Gordon, and A. Asatryan, "Algebraic corrections of paraxial wave fields," *J. Opt. Soc. Am. A* **14**, p. 3300, 1997.
8. H. Hopkins, "The concept of partial coherence in optics," *Proc. Royal Soc. A* **208**, p. 263, 1951.
9. E. Kintner, "Method for the calculation of partially coherent imagery," *Appl. Opt.* **17**, p. 2747, 1978.
10. R. Gordon, "Exact computation of scalar, 2D aerial imagery," *Proc. SPIE* **4692**, p. 494, 2002.
11. S. Subramanian, "Rapid calculation of defocused partially coherent images," *Appl. Opt.* **20**, p. 1854, 1981.
12. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge, 2nd ed., 1992.