Optimal frequency coverages and parsings for imaging interferometric lithography

Thanis M. Tridhavee
University of New Mexico
Department of Electrical Engineering
and Computer Engineering
Albuquerque, New Mexico 87106
E-mail: tridhavee@ece.unm.edu

Balu Santhanam
University of New Mexico
Department of Electrical
and Computer Engineering
Albuquerque, New Mexico 87131-1356

Steven R. J. Brueck*
University of New Mexico
Center of High Technology Materials
1313 Goddard SE
Albuquerque, New Mexico 87106

Abstract. Imaging interferometric lithography (IIL) is an optical resolution enhancement technique, based on wavelength-division multiplexing, that combines off-axis illumination with multiple exposures and pupil filtering. In prior experiments, IIL was shown to be capable of improving the resolution of optics to subwavelength regimes but only in isolated parameter settings. The frequency parsing scheme plays a critical role in the resolution of aerial images from IIL and a platform for optimizing these frequency coverage parameters is currently lacking. We present three approaches to the optimization of IIL frequency coverage parameters. First, the exhaustive search approach and, then, the dynamic programming and greedy versions are presented. For unobstructed lenses and large numerical apertures, these approaches converge to frequency coverages with illumination points along the vertical and horizontal axes, while a lens with an obscured center results in a tilted frequency coverage. Two particular frequency parsing strategies are analyzed from the perspective of the quadratic term in the aerial intensity. Simulation results comparing the off-axis exposures with and without the on-axis exposure are presented. The effect of the quadratic terms on the exposure latitude is shown to have a positive effect in the dense edge areas and a negative effect on isolated edges. [DOI: 10.1117/1.2008969]

Subject terms: lithography; resolution enhancement; parsing strategies.

Paper 03079 received Dec. 18, 2003; revised manuscript received Dec. 13, 2004; accepted for publication Feb. 8, 2005; published online Aug. 10, 2005.

1 Introduction

The semiconductor industry pursues an effort to enable smaller feature sizes with each new generation of lithography manufacturing equipment motivated by the need for lesser power consumption, and higher clock speeds. A majority of the feature size reductions on an integrated circuit are accomplished through improvements in the optical lithography system. Several methods have been developed to decrease the printed feature sizes. The most simple methods used include (1) reducing the wavelength of the light source (λ) and (2) increasing the numerical aperture (NA). However, these simple methods are not without their own problems and their potential to improve resolution have nearly been exhausted. Hence, an area referred to as resolution enhancement techniques (RETs) or wavefront engineering has emerged.[1–3] RETs encompass optical proximity correction (OPC), subresolution assist features (SRAFs), off-axis illumination (OAI), and phase-shift masks (PSMs). Of course each of these RETs has its own advantages and disadvantages. See Refs. 4 and 5 for details of these RETs.

Imaging interferometric lithography[6–9] (IIL) is an optical RET analogous to holography, where a plane wave is inserted into the path of diffracted light passing through the lens at the image plane. The reference beam in OAI can in essence be viewed as the mechanism for obtaining higher resolution. With high NA, large off-axis illumination angles are not needed as in previous forms of IIL. (Note that earlier configurations of IIL dealt with smaller NA factors and required the reintroduction of a reference beam to retain the dc or zero-order beam.) Present configurations of IIL are similar to double-dipole or cross-quad RETs, however, IIL by itself uses coherent light, multiple off- and on-axis illumination exposures with matching pupil filters. TE polarization for horizontal and vertical spatial frequencies, and a single mask are used in combination to reach the theoretical limit of $\kappa_1 = 0.25$. In addition, IIL pupil filtering provides robust image fidelity to reduce[10,11] the need for extensive OPC that would be required with just double-dipole and cross-quad off-axis illumination techniques.

The IIL optical system can be viewed as an analog filter bank system where the spectrum is divided into small parts and each individual portion is transmitted through the band-limited system. The bandwidth of an optical system ($f_{\text{opt}} = n\ \text{NA}/\lambda$) is inversely proportional to the minimum critical dimension (CD), where n is the refractive index of the medium ($n = 1.44$ for water). IIL can approach the theoretical frequency bandwidth of optics to $2n/\lambda$ with little attenuation of high spatial frequencies and hence reduce the CD to sizes not achievable with single-exposure optical lithography. The celebrated Raleigh resolution formula,

*Also at Departments of Electrical and Computer Engineering and Physics and Astronomy, at The University of New Mexico

1537-1646/2005/$22.00

$$\mathrm{CD} = \kappa_1 \frac{\lambda}{n\mathrm{NA}} = \kappa_1 \frac{1}{f_{\mathrm{opt}}}, \tag{1}$$

is the primary predictor of how far feature sizes can be reduced to, while maintaining printability. With coherent illumination, the attainable CD is equal to the theoretical half pitch (h) when $\kappa_1 = 0.5$ for periodic structures. The κ_1 parameter for arbitrary patterns is interpreted as a process parameter and takes a lower value if RETs are used.

A coherent optical imaging system in air medium has a bandwidth of NA/λ, where $\mathrm{NA} \leqslant 1$. Immersion lithography will extend the effective $\mathrm{NA} \sim 1.4$, thereby increasing the ability to print denser features. Imaging at large NAs with immersion lithography will dictate the use of TE polarization rather than unpolarizied/circular polarization to retain the contrast needed to reach the 45-nm node.[10] The separate exposures for horizontal and vertical frequencies enable TE polarization for the high-spatial-frequency plane waves to provide strong contrast at the high-spatial-frequency features. Hence, lesser contrast due to TM polarization seen at the 45-nm node is eliminated with only TE polarizations. Without loss of generality, in this paper, only IIL scalar aerial images in air are considered since TE polarization is used.

The highest spatial frequency attainable is $2n/\lambda$ and the uniform lowpass filter characteristic of the optical system in coherent light has a maximum frequency of $n \sin(\theta)/\lambda$, where θ is bound by 90 deg. In this case, n/λ of frequency space is not utilized. IIL is a RET that can attain this spatial frequency limit. In other words, IIL in theory can make better use of this frequency space to achieve smaller feature sizes. Furthermore, higher spatial frequencies are not attenuated as in partially coherent/incoherent illumination. The nonuniform modulation transfer function across all frequencies for partially coherent illumination can lead to degradation of the image quality across patterns with various pitches.[5] Hence in IIL, there is less concern about sparse and dense pattern degradation than in partially coherent systems, since IIL breaks up the frequency space through the use of multiple exposures to reach the limits of optical resolution.

In addition, apodization of the various exposures in IIL is considered in Ref. 11 as a means to eliminate Gibbs ringing, which occurs because of the hard frequency stops between the on-axis and off-axis exposures. Binary pupil filters and apodization help control the impact of the nonlinear term on the aerial image by reducing low-frequency emphasis and isodense biases that reduce the resolution, thereby removing the need for the application of extensive OPC.

Several parsing strategies of frequency space have been discussed in past experiments,[8] but only for specific parameter settings. The goal of the optimization problem dealt with in this paper is the design of an illumination source for IIL that is analogous to the source design problem in partially coherent light and an OPC mask.[12] In this paper, we look at the problem of optimizing the frequency parsing parameters for a fixed number of exposures. We specifically look at an exhaustive search approach for solving the underlying combinatorial optimization problem and then consider the dynamic programming and greedy versions of the approach. For the discussions presented in the paper, we use binary masks without OPC and pupil filters to compensate for the loss in image quality. Finally, we analyze the nonlinear quadratic term in the IIL aerial image and its impact on pattern printability and quality from simulations.

2 IIL Imaging Model

In this section, we review the imaging model of IIL. The total scalar aerial image intensity of IIL is given by the sum of incoherent intensities:

$$\begin{aligned} I = \sum_{n>1} \Bigg(& \gamma_n^2 + \gamma_n \int_{\mathbf{f}\in P_n(\Omega_n)} \{\langle x, \exp(j2\pi\mathbf{f}\boldsymbol{\rho})\rangle \exp(-j2\pi\mathbf{f}\boldsymbol{\rho}) \\ & + [\langle x, \exp(j2\pi\mathbf{f}\boldsymbol{\rho})\rangle \exp(-j2\pi\mathbf{f}\boldsymbol{\rho})]^*\} \mathrm{d}\mathbf{f} \\ & + \left| \int_{\mathbf{f}\in P_n(\Omega_n)} \langle x, \exp(j2\pi\mathbf{f}\boldsymbol{\rho})\rangle \exp[-j2\pi(\mathbf{f} - f_n^{\mathrm{off}})\boldsymbol{\rho}] \mathrm{d}\mathbf{f} \right|^2 \Bigg) \\ & + \alpha \left| \int_{\mathbf{f}\in P_1(\Omega_1)} \langle x, \exp(j2\pi\mathbf{f}\boldsymbol{\rho})\rangle \exp(-j2\pi\mathbf{f}\boldsymbol{\rho}) \mathrm{d}\mathbf{f} \right|^2 . \end{aligned} \tag{2}$$

The symbol $\langle \cdot, \cdot \rangle$ denotes the inner product or the Fourier coefficients in $L^2(\mathbb{R}^2)$. Note that $\boldsymbol{\rho}, \mathbf{f} \in \mathbb{R}^2$ are the space and frequency vectors, respectively; and n is an index variable for the number of exposures. The parameter γ_n refers to the relative strength of the dc term to the linear and quadratic terms, and α is a parameter that relates the exposure dosage of the on-axis term to the off-axis terms. The optimization of γ and α are discussed in Refs. 11 and 13. The value of the γ is, in general, dependent on the mask, and optimization results on a dense mask are reported in Ref. 11 using the PROLITH vector model. In Ref. 13, the relative exposure energies γ and α are optimized using an iterative process between the desired image and a weighted error metric on the aerial image. We consider only the frequency coverage parameters here because joint optimization of the frequency coverage parameters and the relative exposure energies, albeit more optimal overall, would result in a computationally nontractable problem. In Ref. 11, γ and α are optimized in PROLITH for various combinations and it is shown that the biases can be reduced and that the contrast can be equalized such that the total aerial image has proportional contrast between the low- and high-frequency exposures. Since IIL requires several exposures, it becomes apparent that a choice of $\gamma \leqslant 1$ is required to reduce the dc background light for photoresist development.

Here, P_n represents the pupil filtering operation where certain frequencies on the set Ω are blocked, passed, or apodized. The frequency sets Ω are defined according to the circular shape of the pupil function of the optical system:

$$\Omega_n^{\mathrm{disk}} = \{\mathbf{f} : \|\mathbf{f} - \mathbf{f}_n^{\mathrm{off}}\|_2 \leqslant f_{\mathrm{opt}} \; \forall \; \mathbf{f}_n^{\mathrm{off}} \in \Psi\},$$

where $\|\cdot\|_2$ denotes the Euclidean distance, and $\mathbf{f}_n^{\mathrm{off}}$ denotes the shift in frequency and is controlled directly by the off-axis illumination point source.

Current IIL laboratory experiments use a high-NA lens. This type of lens has an obstruction in the center of the lens and this is typically modeled by a null of the Fourier coefficients in the center of the pupil plane. The lowest frequency before the obstruction in the Fourier plane blocks

any frequencies from being passed through is f_{obs}. For an optical system with an obscuration in the center of the Fourier plane, the frequency sets take the form

$$\Omega_n^{\text{donut}} = \{\mathbf{f}: f_{\text{obs}} \leqslant \|\mathbf{f} - \mathbf{f}_n^{\text{off}}\|_2 \leqslant f_{\text{opt}} \; \forall \; \mathbf{f}_n^{\text{off}} \in \Psi\},$$

where the offset frequency $\mathbf{f}_n^{\text{off}}$ has a maximum value of $1/\lambda$ in IIL and NA/λ in OAI.

The frequency offsets $\mathbf{f}_n^{\text{off}}$ are further related to the angle of illumination by the following:

$$\mathbf{f}_n^{\text{off}} = \frac{\sin(\theta_n^{\text{off}})}{\lambda}[\cos(\varphi_n)\sin(\varphi_n)]^T, \tag{3}$$

where φ_n is the angle in the frequency plane of the vector $\mathbf{f}_n^{\text{off}}$ relative to the horizontal line. The offset angle θ_n^{off} relates to the radius in the frequency plane of the off-axis illumination angle. If the offset frequency is along the vertical direction, then φ_n is 90 or 270 deg, and if the offset frequency is along the horizontal direction, then the angle is only 0 or 180 deg.

The offset frequency sets Ψ are defined as the sets of all possible off-axis illuminations:

$$\Psi_{\text{IIL}} = \{\mathbf{f}: \|\mathbf{f}\|_2 \leqslant n/\lambda\}. \tag{4}$$

The set just defined is bounded by the frequency of n/λ since the maximum offset angle is at 90 deg, i.e., $f_{\text{off}} \leqslant n \sin \theta^{\text{off}}/\lambda$ and the next set is bounded by the cutoff frequency of the optical system:

$$\Psi_{\text{OAI}} = \{\mathbf{f}: \|\mathbf{f}\|_2 \leqslant f_{\text{opt}}\}. \tag{5}$$

In either case, we refer to the set of offset frequencies Ψ_{IIL} or Ψ_{OAI} as simply Ψ. For simplification of notation, we refer to just Ω_n and not to the disk or donut type of frequency set shape. An index set of the offset frequencies is denoted as the index set Γ. In simulation, any Ω_n has a finite number of sample points, and we denote J as maximum size of this finite set, as discussed later in Sec. 3.4, $\Gamma \in \mathbb{N}^J$.

3 Optimization

Lithography imaging systems can be described by a simple homogeneous operator equation[14]

$$y = \mathcal{N}x, \tag{6}$$

where x is the object (mask), y is the image printed onto the wafer, and $\mathcal{N}$ is a nonlinear operator describing the nonlinear operations of the optical imaging and photoresist processes. The typical method used to invert such a system employs the contraction mapping theorem, i.e., the trial-and-error iteration algorithm, to find solutions. In our case, we know the input object x and we want to optimize the system operator $\mathcal{N}$ so that the output aerial image is as close as possible to x. We assume x is a binary mask without any OPC. OPC will enhance the image fidelity, however it is expected that the locations of the illumination source points should not differ significantly if an OPC mask is used.

3.1 *Nonlinear Bandlimited Approximation Problem*

Bandlimitedness of the optics limits the resolution capabilities of the entire optical lithography system.[15] In a broad view, an optical system implements a finite Fourier transform, and IIL in theory extends the bandwidth to high spatial frequencies with little attenuation to reconstruct the object (reticle) for smaller, denser features.

The illumination source points or frequency offsets determine the Fourier coefficients of the mask that are passed through the low-pass filter of the optical system. The set of frequency offsets in essence controls the set of basis functions onto which the signal is projected. Hence, basis selection can be performed appropriately for a given object (reticle). For Manhattan (rectangular geometry) patterns, the Fourier coefficients are distributed over the vertical and horizontal frequencies, hence a better choice would be to place the frequency offsets on the horizontal and vertical axes to recover those coefficients that are not passed through the conventional illumination based on axis exposure. Our optimization goal, therefore, is to find the best set of offset frequencies and to design the shapes of pupil filters to obtain the best aerial image quality. The locations of the circular frequency sets determine the extent to which the intersections of the coverages can be manipulated. The pupil filters used in the Fourier plane of the optical system are determined from the offset frequencies and the amount of overlap in frequency coverage (intersection of the circular frequency sets). The overlaps in frequency spectrum provide some freedom for manipulating the quadratic term while at the same time maintaining the same uniform frequency coverage. Therefore, the intersections of these circular frequency sets are where pupil plane filters can be placed to eliminate relative amplification of some frequencies.

The optimization problem can then be reformulated as a problem that seeks to minimize the distance between the binary mask without OPC (desired image x) and the aerial image (I) defined in Eq. (2). Only optimization of the frequency coverage or illumination source with a binary mask is addressed here, since IIL uses pupil filters (apodization) to remove the extensive OPC requirements on the mask. The filtering operation mitigates the degradations from the quadratic term in Eq. (2), thereby making the linear term the primary contributor. Comparisons with and without pupil filters are shown in Refs. 10 and 11.

Since we view the OPC problem as an inverse problem and the source illumination problem as an approximation problem, combining OPC into the mask should not change the solutions significantly since IIL assumes point sources and OPC does not change the frequency information tremendously since typically an OPC mask is highly correlated to its binary equivalent mask. The frequency coverage should not change much at a high NA, since most of the attain frequency space is covered by few exposures. At mid-NAs, the frequency coverage (point sources) locations will be more sensitive since frequency content changes depending on the OPC used, however the energy contribution from the non-OPC features should dominate the energy contributed from the OPC features. Thus, maximizing the energy criteria should give similar results in various cases.

Further work optimizing the OPC and source together to improve image contrast and fidelity can be done similar to that in Ref. 16.

The optimization problem can then be reformulated as a problem that seeks to minimize the distance between the binary mask without OPC (x) and the aerial image (I) defined in Eq. (2). Optimizing the error with respect to the aerial image makes the optimization more mathematically tractable. This optimization problem is similar to the PSM optimization problem of Ref. 14. The distance between the desired image and aerial image is

$$\epsilon = \|x - I\|, \tag{7}$$

where $\|\cdot\|$ can be some norm (metric), such as a weighted L^p norm, or possibly a complex cost function based on lithography metrics such as exposure latitude, degree of freedom (DOF), mask error enhancement factor (MEEF), corner roundness, line-width error, contrast, etc.[5,17,18] Typically, a metric is chosen for ease of optimization or determined by some prior information about the class of signals present. The better optimization procedure would be to optimize the whole system.[14] However, the formulation adopted in this paper makes assumptions to create a mathematically trackable solution.

From Eq. (7), we would first like to minimize ϵ with respect to γ_n, α, Ω_n, and P_n, where n can take at most N exposures in an index set Γ where $\Lambda \subseteq \Gamma$ and $\Lambda \in \mathbb{N}^N$. Simplifications are necessary to reduce the complexities of the optimization.

Heuristically, the most important terms of Eq. (2) are the linear terms of the first integral that contribute to the high-resolution capability of IIL. The quadratic term represents the interference of the non-dc spatial frequencies since the linear term is the interference of the dc plane wave (reference beam) with the non-dc spatial frequencies. For the discussions in this paper, the quadratic term is ignored in the optimization, since it contributes mainly to isodense bias, which can be mitigated with pupil filtering. Through simulation, its impact can be analyzed. The last square term is the on-axis image and its aerial image is nonlinear in nature due to the squaring. In addition, the on-axis exposure provides only low-frequency coverage and only the pupil size of the low-frequency coverage (on-axis pupil) can be varied to provide better results if the off-axis frequency coverage overlaps with the on-axis frequency coverage. The on-axis term is therefore ignored in the optimization, i.e., $\alpha=0$. Then one simple approach to optimize this system is to maximize the energy of the linear terms in Eq. (2).

Adopting the L_2 norm for mathematical tractability, we ignore the dc bias in the frequency coverage optimization because it does not contribute to the resolution. The γ parameter must be optimized at high NA using a vector model and photoresist process for accurate results. Coupling the optimization of γ and point source locations will create a computationally unattractive problem under a single metric. Figure 1 shows γ's for various error curves in the four and five exposure setup in PROLITH full vector simulator. The error curves represent the SE between the optimal binary mask and a threshold image at a certain threshold level, as shown on the horizontal axis of Fig. 1. The background

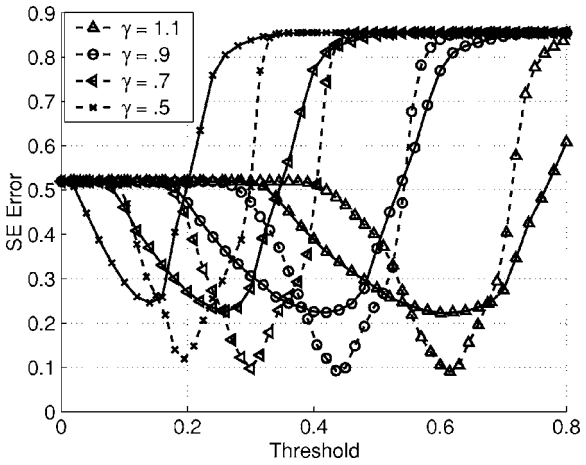

Fig. 1 Square error (SE) curves for various γ=0.5, 0.7, 0.9, 1.1, NA=0.9, and frequency coverage of Fig. 4. SE calculated between optimal binary image and various threshold images from intensity; (- -) curves are five exposures and (-) curves are four exposures; α=1 for a five-exposure system.

(dc) is no higher than in the four- or five-exposure case since the exposure dosage is adjusted so that the intensity falls into the threshold region of the photoresist. The parameter γ varies the dc bias and the gain of the linear terms, thus affecting the contrast of the intensity. Determining the frequency coverage is decoupled from the optimization of γ since any variation here does not affect the frequency coverage information. Varying γ will directly vary the energy of the linear term. Then optimizing γ only on an L_2 metric to be discussed later in Eq. (10) is not desirable because large values of γ would result in increased energy (L^2 norm) if not bounded. If $\gamma<1$, the contrast will decrease because there will be less variation in the linear term, however, the dc bias will decrease, thereby increasing the contrast until decreasing γ attenuates the linear term too strongly. If $\gamma>1$, the contrast will decrease since the dc bias will increase, but then the linear term will be amplified. A large γ is not acceptable since the intensity would then consist primarily of background light. A small γ is not acceptable since low-exposure latitude would not be robust for imaging. The contrast metric therefore is a natural metric to optimize the γ parameter, as seen Fig. 1 for various γ's since a maximum will exist to balance exposure latitude and bias.

When optimizing the linear terms of all the off-axis frequency coverages jointly, it is necessary to consider the union of all the frequency sets Ω_n. This is equivalent to the union of disjoint pupil filtered frequency sets, $P_n(\Omega_n)$. We define the union set as

$$\Phi = \bigcup_{n \in \Lambda} \Omega_n. \tag{8}$$

Then the optimization can be simplified to

$$\epsilon = \left\| x - \left[\int_{f \in \Phi} \langle x, \exp(j2\pi \mathbf{f}\boldsymbol{\rho}) \rangle \exp(-j2\pi \mathbf{f}\boldsymbol{\rho}) \mathrm{d}f \right] \right\|_2. \tag{9}$$

This optimization problem is similar to a nonlinear approximation problem[19] because the integration is computed over

the union of circular frequency sets. Then the correct frequency sets must be picked to give the best approximation by choosing the most significant coefficients. Using Bessel's inequality, an expansion into an orthonormal basis is minimized in the L_2 norm by maximizing the norm/magnitude of the coefficients. To achieve a good exposure latitude, the minimization then becomes equivalent to a maximization of the coefficient energy and results in the minimum achievable approximation error.[20] This error is minimal only if the linear terms are the only terms contributing to the aerial image. Later, we evaluate the exposure latitude of the total aerial image to observe the enhancement in the exposure latitude from the optimization. The effect of the quadratic term and bias can also be inferred from the figures.

The nonlinear BAP optimization problem can then be formulated as

$$\arg \sup_{f \in \Phi} \|\langle x, \exp(j2\pi \mathbf{f}\boldsymbol{\rho})\rangle\|_2^2. \tag{10}$$

Equation (10) maximizes the energy of the signal over all possible unions of the frequency sets for a given number of exposures. The transform coding area[21] deals with a similar optimization problem, where if a finite number of transform coefficients are to be retained, only the most significant coefficients are kept and the rest are set to zero. Typically in nonlinear approximation theory a threshold operation is placed on the coefficients, where all coefficients above a threshold value are kept and the rest are nulled. The circular frequency sets play a similar role to the threshold operation since all Fourier coefficients outside the bandwidth NA/λ are set to zero. Hence, it is necessary to find only the coefficients that are most significant.

The optimization associated with the BAP can be viewed as being equivalent to maximizing of energy in the projections and also equivalent to maximizing the correlation of the mask to a set of basis functions. Geometrically, this is viewed as two vectors, one in a large space and another in a subspace belonging to that space. The projection theorem geometrically says that a right triangle exists between vector representing the mask and the optimal vector belonging to the subspace of possible approximations. This projection onto the subspace is the optimal approximation in the least squares or the L_2 (Euclidean) sense.

3.2 *Algorithms to Solve Nonlinear BAP*

Several algorithms can be created to find the solutions to Eq. (10). Three algorithms are described here since there are three paradigms to algorithms.[22] First, we look at an exhaustive algorithm (EA), then a greedy algorithm (GA), and last a dynamic programming algorithm (DPA). The typical properties of a GA are that it will take the best immediate solution. However, this may not be the best overall solution. GAs find less than optimal solutions, but can be computationally faster and perhaps can attain solutions near the global solution with the application of suitable constraints. The EAs are globally optimal. However, they must search all possible combinations and this can be computationally difficult. Last, a DPA is discussed. Typically, they are based on solving smaller subproblems and saving the intermediate results for reuse later. The DPAs will converge to the same results as the EA with fewer computations or memory requirements at a given instance.

With all these searches, we assume pupil filters are implemented to eliminate the redundancy in the Fourier spectrum. In OAI, the reference beam is kept within the NA. As the illumination angles increase, the pupil ceases to pass the dc plane wave (reference beam) and we are in a IIL configuration for mid- and low-NA systems. Since future optical systems will have high NAs, bringing the reference beam outside the system and reintroducing the reference beam at the image plane is theoretically unnecessary.

3.3 *EA*

Since it may not be necessary to have a very fast algorithm for lithography related optimization, we explore the optimal algorithm to find the best coverages for any given mask pattern that is implementable and tractable. A combinatorial optimization is necessary to solve Eq. (10) optimally, i.e.,

$$\Lambda^* = \arg \sup_{f \in \Phi} \|\langle x, \exp(j2\pi \mathbf{f}\boldsymbol{\rho})\rangle\|_2^2. \tag{11}$$

The best linear approximation is obtained by choosing the best combination of basis functions (plane waves). Finding the index set Λ achieves the optimal approximation. Since in optics we are constrained to project the signal onto a circular frequency set (Airy disk) of basis functions, all combinations of these frequency sets must be searched in a joint fashion for N exposures. All the unions of the N possible frequency sets can be created by building all combinations of the offset frequencies ($\mathbf{f}_i^{\mathrm{off}}$) of Ω_i. Then it is just a simple search through all the possible projections of the mask onto the unions of N frequency sets for the best nonlinear approximation.

The computation complexity will increase with the number of exposures since $J \gg N$. The number of combinations to search is $\binom{J}{N}$, where N is the number of exposures (typically 2 for mid- to high-NA systems), and J is the number of possible frequency sets. Note that J is the maximum number of offset frequencies that are possible. Typically, for two exposures, the computations are not so large. For three exposures, the number of combinations is too numerous to track. The Hermitian symmetry in the pupil plane requires only a search over half the frequency plane since we assume a binary mask. The additional off-axis illumination points needed to preserve the telecentricity of the image are at the opposite end of the illumination pupil or similarly, the frequency coverage is located at the opposite side of the pupil plane (180 deg apart) with respect to the optical axis.

3.4 *GA*

The GA suffers from the possibility of converging to suboptimal solutions for the nonlinear bandlimited approximation problem. However, it is less computationally complex and with additional constraints, solutions near the globally optimal coverage can be found. The algorithm developed here is similar to matching pursuits algorithm in signal processing.[23] The algorithm tries to maximize the energy of the linear term by iterating through each exposure. Once it finds the exposure with the maximal energy, the coefficients

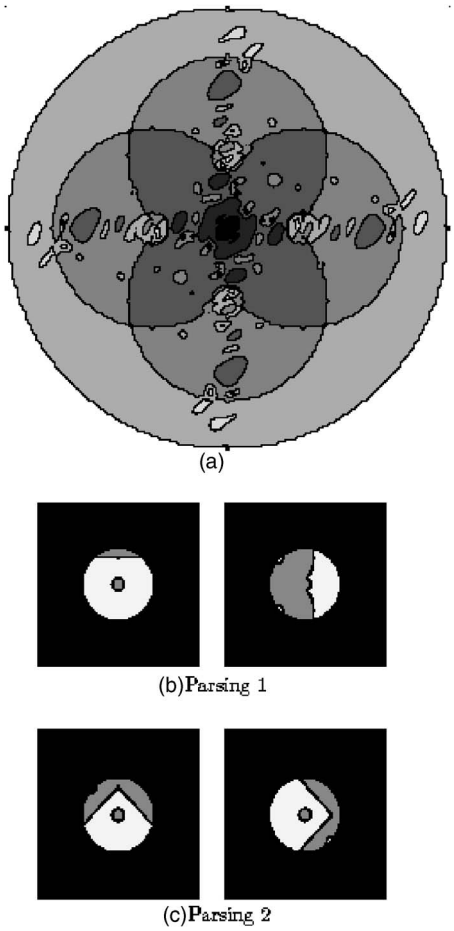

Fig. 2 Suboptimal frequency coverage at 71 nm from the GA plotted with Fourier coefficients with cross search: (a) and (b) pupil filters of parsing strategy 1 and 2 at 71 nm for same uniform frequency coverage. Frequency space extends to the highest optical frequency of $2/\lambda$. The small circle represents the central obscuration and the large circle represents linear frequency coverages. White represents clear areas in the Fourier plane, gray represents opaque areas in the Fourier plane, and black represents area outside Fourier plane. The left and right column of (b) and (c) are pupil filters for vertical and horizontal illumination points, respectively.

of that exposure are removed and then the next exposure with maximal energy is found until the number of exposures reaches N. This optimization algorithm for one exposure is then independent of prior knowledge of another exposure.

The mask is projected onto a finite number of possible frequency sets and the set with the best approximation is then removed. The coefficients are removed in the sense that the coefficients of the Fourier basis functions in a set are nulled. We can view the best nonlinear approximation problem geometrically as a right triangle where the hypotenuse is the mask and the best approximation is the closest vector in the subspace of vectors. Each vector making up the subspace is viewed as a projection of the signal onto a possible frequency set. Hence, the right triangle with the smallest angle between the object and its projection is the best approximation. In other words, the orthogonal projection onto these subspaces is the optimal approximation. Then the vectors making up the best approximation are removed and the signal is projected again onto the space of the remaining basis functions making up of a smaller subspace.

The computational requirements of the algorithm depend on size of the frequency domain grid J. The approximation with finer frequency sampling is better, but at the cost of more memory and CPU requirements. Maximizing the energy of the linear term requires projection of the signal onto all possible circular frequency sets Ω_n. The size of the possibilities are proportional to the size of the frequency grid. Half of the frequency grid is redundant due to the Hermitian symmetry, thereby reducing the search by one half. Since optics has a circular pupil transmission bandwidth up to a certain frequency, the most number of frequency sets that need to be searched for is $J=M/2 \times M \sim O(M^2)$, if we assume the frequency grid is $M \times M$. The frequencies near the corners of the square frequency grid are not required to be searched due to the circular pupil function. Hence, the size of the search is reduced to $J = \pi M^2/8 \sim O(M^2)$.

3.5 *GA Outline*

The steps in the GA are

1. Assume binary pupil filters are implemented to remove duplicate coverage.
2. First maximize the energy of coefficients in Ω_i with respect to all possible frequency sets, where l is initially equal to 1.

$$\Omega_l = \arg \sup_{\Omega_{i \in \Gamma}} \|\langle R^l x, \exp(j2\pi \mathbf{f}\boldsymbol{\rho})\rangle\|_2^2,$$

   where $R^1 x = x$ and Ω_l is optimal frequency set at I'th step.
3. The residual of the signal x is calculated as follows:

$$R^{l+1}x = R^l x - \int_{\mathbf{f} \in \Omega_l} \langle R^l x, \exp(j2\pi \mathbf{f}\boldsymbol{\rho})\rangle \exp(-j2\pi \mathbf{f}\boldsymbol{\rho}) \mathrm{d}\mathbf{f}.$$

4. Repeat step 2 until l equals the maximum number of exposures N.

Prior knowledge typically determines the number of exposures N. Typically, at high NAs four exposures are required to cover frequency space, as shown in Fig. 2(a). More exposures require more exposure time and eventually increases the lithography cost of ownership. The value of N is typically obvious from the parameters of the optical system's NA, wavelength, and knowledge of the feature sizes of the mask pattern.

The likelihood that the GA converges to a point near the global solution can be increased by constraining the search for the optical frequency sets to cover the vertical and horizontal axes of the frequency plane. Simple prior knowledge can be exploited to decrease the computational time and to find solutions near the global optimal solution without a need for an EA. This idea can be applied to any of the algorithms since it amounts to just reducing the search size. Later, we refer to this as the cross-hair search instead of a full search in the Fourier plane.

3.6 *DPA*

The EA discussed previously can be very computationally complex even though only a search through a combinatorial set is necessary. Building the combinations of frequency sets can be computationally difficult in itself since it requires building of simultaneously many frequency sets. A DPA may be more attractive from a computational viewpoint than searching through the all possible combinations. Then, a search through all combinations at the end may be faster by solving many smaller subproblems instead of searching through all the relevant combinations, as discussed in Sec. 3.3. Even though the complexity of the EA is smaller, it requires a large memory to achieve a fast search. The number of possible sets for a full combination is then $\mathcal{O}(J^N)$. However, N is computable since it is only two and since we assume high NAs and symmetry in the pupil plane.

This algorithm has a nature similar to the GA described in Sec. 3.4. The calculation of the energy (sum of squares) can be done separately over the Fourier coefficients in each exposure, instead of all together as is done in the EA over all combinations of frequency coverages. A DPA can then take advantage of this property to store the energy of the individual exposures until the correct corresponding exposures are found. Then the total sum of the energies of the individual frequency coverages will dictate the winning joint-frequency coverage.

The primary idea, is to first build all frequency sets in index set J only (all possible single-frequency coverages), as shown in Fig. 3. We then compute the energies of signal projected onto these frequency coverages into a table. This corresponds to energies at level 1 in a tree structure or the first exposure. For each element in level 1 (a possible frequency set), the remaining coefficients (residue) outside are used to find the best second exposure. The coefficients remaining are the residual of the first frequency coverage. For each possible second frequency coverage for a given first frequency coverage, the coverage with the most energy is tabulated into a table of energies that contains the highest energies. This value is then added to the energies of the given first exposure frequency coverage. Therefore, the optimal nonlinear approximation is the maximum energy of the sum of energy from the first exposure with its best second exposure frequency coverage location that has been stored away.

We can envision the DPA in a two-level tree structure, as just described and shown in Fig. 3. The numbers of levels of a tree structure correspond to the number of the exposure. The number of elements at the level 1 are indexed by the elements of set Γ. For each possible exposure in level 1, there are J possible other frequency sets as shown in level 2.

The steps to the DPA are

1. We assume pupil filters are used to eliminate all duplicate frequency coverage or in other words no Fourier coefficients are repeated.
2. The algorithm begins at exposure level $p=1$ and takes a bottom-up progression in the tree structure to the root, $p=0$.
3. In a transversal of a level, there are J^{p+1} nodes with each node containing J frequency sets. We define an index set $\Delta_p \in \mathbb{Z}^{Jp+1}$ that indexes the nodes of a level. The best child set is found according to the following criteria:

$$\Omega^p_{l^*_i} = \arg \sup_{\Omega_{l_{i\in\Gamma}}} \|\langle R^p_l x, F\rangle\|_2^2 \; \forall \; l \in \Delta_p,$$

where $R^0 x = x$ and $\Omega^p_{l^*_i}$ is optimal frequency set at the p'th level, l'th node, and i'th frequency set.

4. Then the best child frequency set for a parent node element is added to the parent, where the parent is $R^{l-1}x$. In other words, the basis functions of the best child space are added to the parents space. This is stated as

$$R^n_l x = R^{p-1}_l x + \int_{\mathbf{f}\in\Omega^p_{l^*_i}} \langle R^p_l x, F\rangle F \, d\mathbf{f} \; \forall \; l \in \Delta_p.$$

5. Then the next best set is found again by returning to step 3 until p reaches the root.

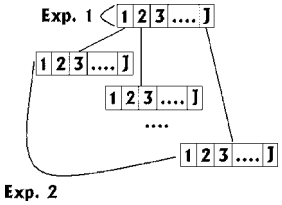

Fig. 3 DPA diagram. Two-level tree structure where level 1 is exposure 1 and level 2 is exposure 2. At level 1, each node represents the Fourier coefficients captured in the exposure, and at level 2, each node represents the remaining Fourier coefficients not captured at parent node.

4 Simulation Results

Listed here are the simulation parameters used for the figures in this paper:

1. Scalar model with polarization control (in Wu's IIL-TCAD vector model simulation tool[13] with polarization control, good agreement is show between cross sections of the aerial images of the vector model with scalar model)
2. NA=0.9 or 0.6
3. wavelength: $\lambda = 193$ nm
4. illumination scheme: OAI where reference beam is kept inside the numerical aperture with NA=0.9; IIL where reference beam is outside the numerical aperture with NA=0.6
5. number of exposures: $N=2$ or 3 (to retain telecentricity $N=4$ or 5)
6. $\gamma_n = 1 \; \forall \, n = 1 \ldots N$
7. $f_{\text{obs}} = 0.15$ (NA/λ)

4.1 *Definitions and Error Metrics*

The SE is defined as the L^2 norm difference between the expansion of the mask to $2/\lambda$, i.e., (x_{opt}) and the linear terms of the aerial image in Eq. (2) both of them excluding the dc coefficient in the expansion:

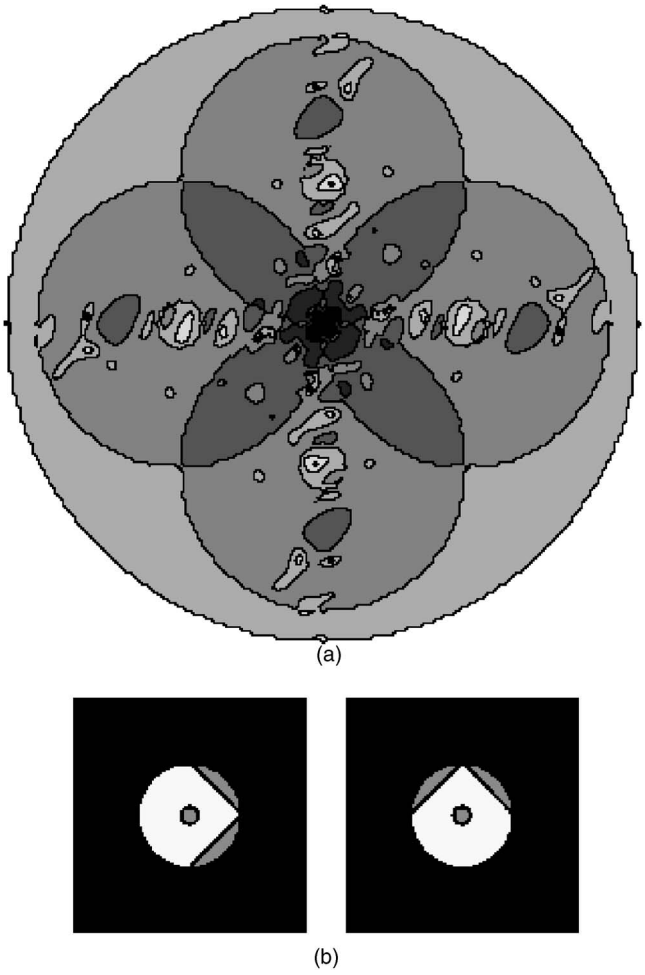

Fig. 4 (a) Optimal frequency space coverage from the DPA at all 71 nm with cross-hair search and (b) pupil filter for frequency coverage from parsing 2 strategy, where the left and right column pupil filters are for horizontal frequencies and vertical frequencies, respectively.

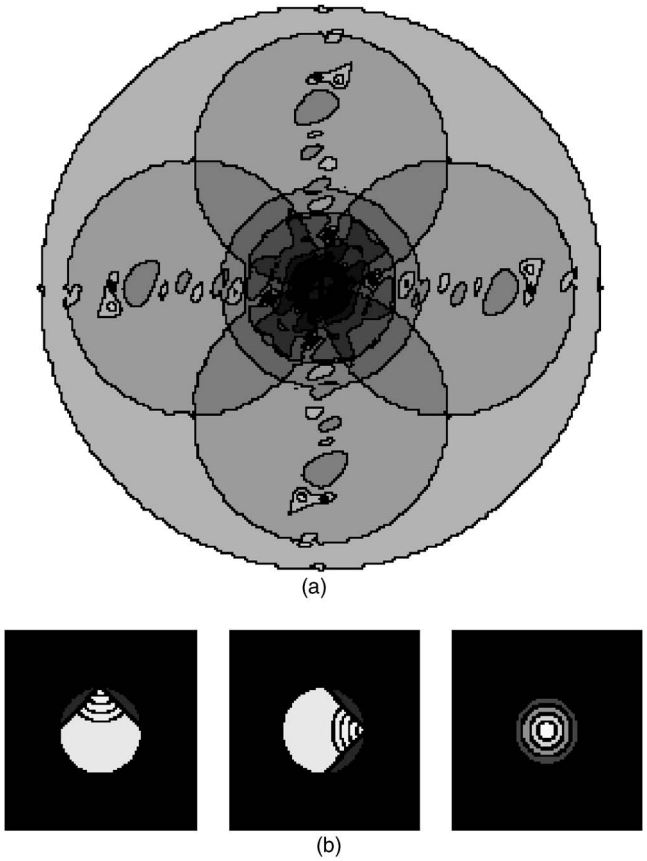

Fig. 5 (a) Optimal frequency space coverage from the DPA at all CDs. The small concentric circles are the on-axis exposures with their respective pupil sizes of NA=0.2, 0.4, 0.6, and 0.8. (b) Left and middle figures are vertical and horizontal frequency off-axis pupil filters with parsing 2 strategy, respectively, and the right figure is on-axis pupil filter. The on-axis filter has four various pupil sizes. In the off-axis filter, the arcs are the various on-axis pupil sizes.

$$\mathrm{SE}_1 = \|x_{\mathrm{opt}} - I_{\mathrm{in}}\|_2^2. \tag{12}$$

The only difference between the mean square error (MSE) and SE is the normalization by $1/N$, where N is the number of samples in the image. It is not done here since the numbers become very small for comparison, and since N is always fixed, it is not necessary.

We also define another metric to compare the binary mask to the aerial image

$$\mathrm{SE}_2 = \|x_{\text{opt-binary}} - I_{\text{aerial}}\|_2^2, \tag{13}$$

where J_{aerial} is the normalized aerial image and $x_{\text{opt-binary}}$ is the optimal binary mask.

The last metric is the same as the Hamming distance, which compares the binary mask and binary threshold of the aerial image:

$$\mathrm{SE}_3 = \|x_{\text{opt-binary}} - t_{\text{out}}\|_2^2, \tag{14}$$

where t_{out} is the threshold image of the aerial image. One type of search is the cross-hair search, where all possible frequency sets over the vertical and horizontal frequency axes are examined. Otherwise, the entire frequency plane is searched, which is referred to as the full search.

The first type of parsing of frequency space is referred to as parsing 1. The intersection of the linear terms is parsed such that the order of the exposures determines the divisions of the intersections of the circular frequency sets Ω_i. The frequency sets are then labeled in the order of the exposures. The first exposure takes the frequency sets that cover the entire intersection of the frequency sets and the linear terms and the next exposure then covers the coefficients that remain outside the previous exposure. This type of sequential parsing arises naturally from the GA, however it can also be applied to the DPA or EA algorithms if one labels the exposures in those algorithms. The GA optimizes the exposures independently and the EA/DPA optimizes the exposures jointly. This second type of parsing will be called as parsing 2. After all the linear terms are determined for each exposure, the intersection of the coverages is parsed with a line orthogonal to the line connecting the centers of the coverages. This is similar to a linear classifier where the orthogonal line between the means of two populations determine the linear discriminant function. This works only for the two-exposure case. In the three-exposure case, the

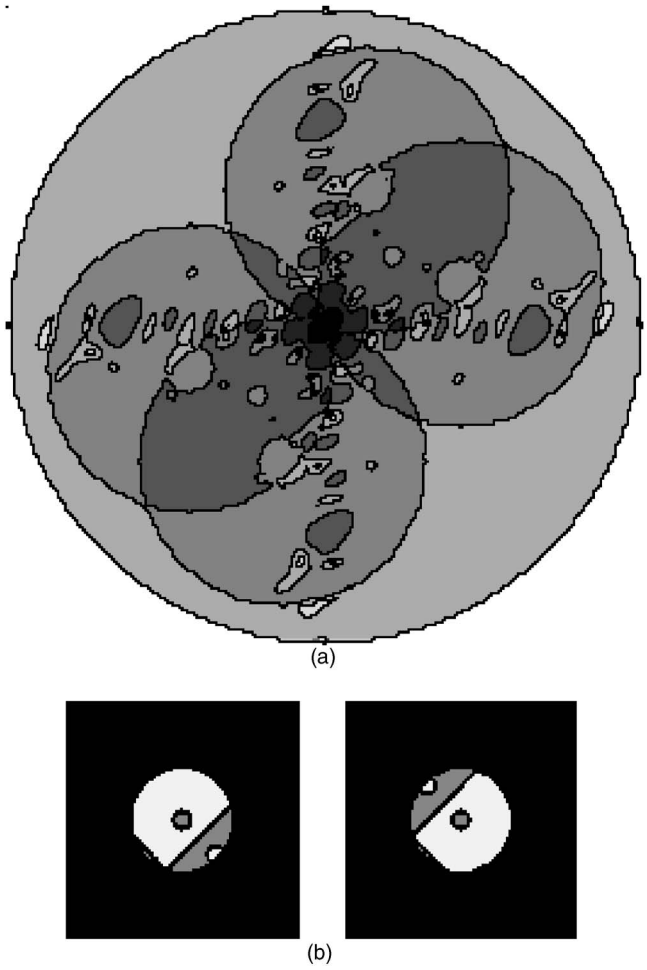

Fig. 6 Optimal frequency coverage at 71 nm with the full search plotted over Fourier coefficients. Note, the recovery of Fourier coefficients lost in the obscuration. Pupil filters at 71 nm using parsing 2 strategy. Left and right columns of (b) are vertical and horizontal frequency pupil filters.

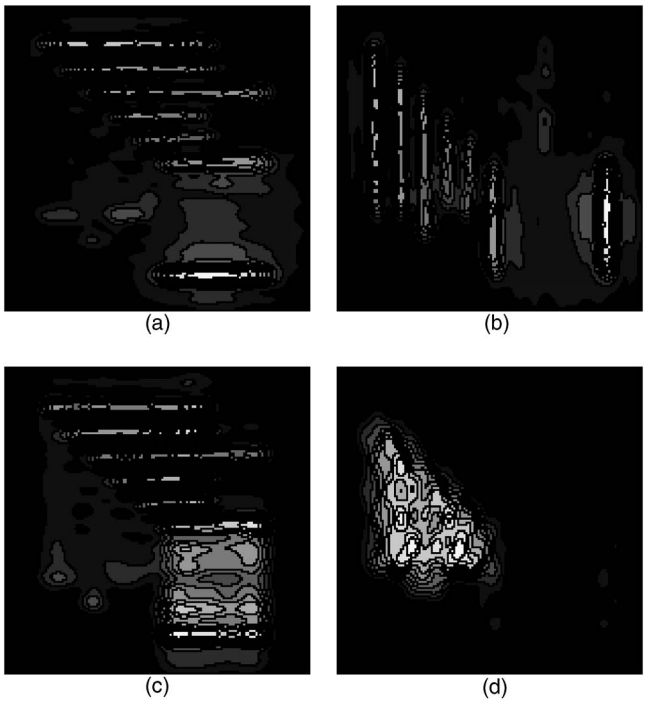

Fig. 7 Quadratic images at 71 nm for two exposures from frequency parsing from Fig. 2. (a) and (b) typical quadratic images of parsing strategy 2 at 71 nm and (c) and (d) from parsing strategy 1. Note, (c) shows a strong quadratic term, and (d) a negligible quadratic term.

parsings can be more complex and will require a more creative parsing algorithm for the two dimensional plane.

4.2 *Frequency Coverages—DPA*

The optimal nonlinear frequency coverage is a result of the distribution of the Fourier coefficients as discussed. Then the pupil filter used to provide uniform frequency coverage will be a by-product of the frequency coverage. Figure 4(a) shows the frequency coverage at 71 nm with $f^{\text{off}}=\text{NA}/\lambda$. The cross-hair search with a donut pupil is applied and the typical two exposure frequency coverage is found. The resulting pupil filters are shown in Fig. 4(b). The off-axis exposures of Fig. 5(a) show the frequency coverage for a disk pupil for a full search where the DPA algorithm converges to the typical vertical and horizontal illumination points. In addition, the on-axis exposure is included to simulate the resulting quadratic effect on the aerial image examined later in this paper.

Next, the case where the full Fourier plane is fully searched is considered and it is shown that the aerial image intensity will have a better approximation if Ω_n is not projected onto the vertical or horizontal frequency axes. The result is shown in Fig. 6(a) and their corresponding pupil filters in Fig. 6(b). This gain is due to the movement of the center obscuration away from the vertical and horizontal frequency axes. If the lens has no center obscuration (disk pupil shape) then the best linear coverage pattern at any target *CD* is the typical frequency coverage that is shown in Fig. 5(a).

Figures 7(a) and 7(b) illustrate typical images of quadratic terms. Cross-hair sections are shown in Figs. 8(a) and 8(b). Typically, the quadratic term has a strong amplitude near the isolated edges and will oscillate in the dense edge areas. In addition, it is observed that the two exposure option has more exposure latitude primarily due to the quadratic term. The degree of the effect of the quadratic term is controlled by the parsing of frequency space. Near the dc and low-frequency areas, the parsing becomes important primarily because these areas control the degree to which the strong low-frequency Fourier coefficients interfere with the high-frequency coefficients. In a three-exposure system or in a parsing strategy where the frequency space is parsed such that the high- and low-frequency Fourier coefficients do not interfere, the quadratic term will then interfere with fewer Fourier coefficients and will have a less pronounced effect. However, the nonlinear effect of the quadratic term can increase the exposure latitude as illustrated in Fig. 9 and Fig. 8(b) of the dense lines, but with an accompanied image quality degradation, as seen in PROLITH vector simulations.[11]

The three-exposures curves in Fig. 10(a) represents vari-

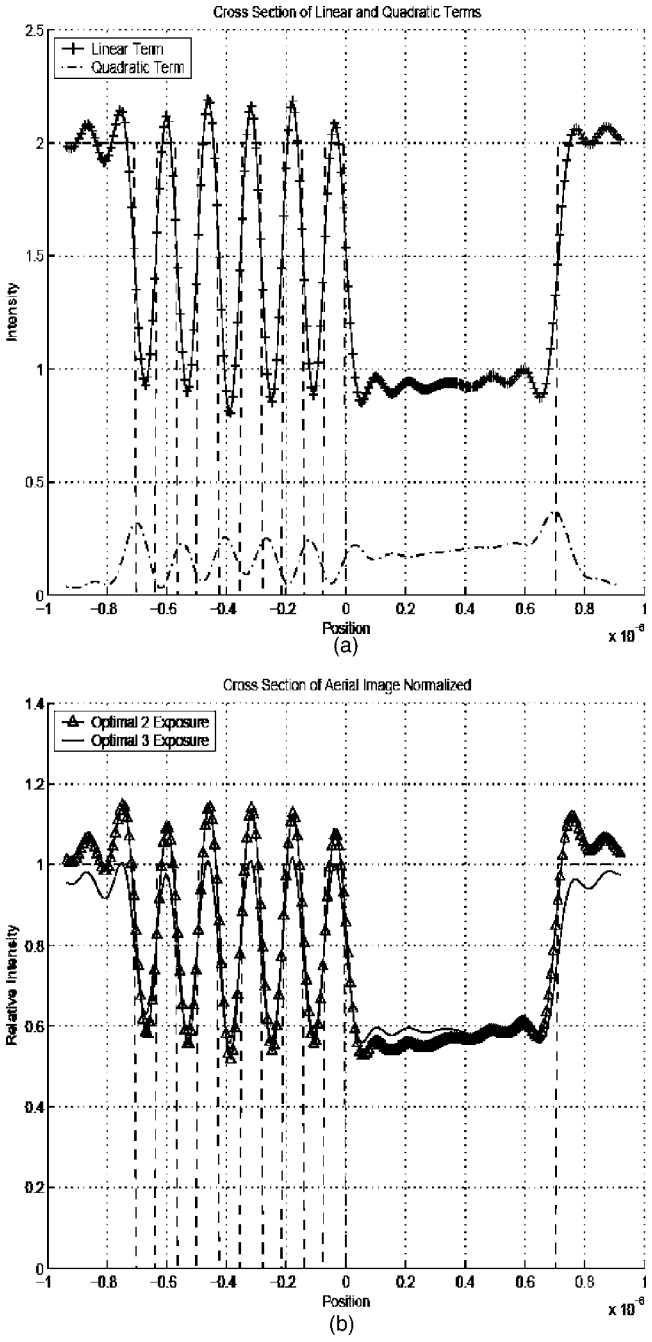

Fig. 8 (a) Cross section of optimal two-exposure comparison to optimal three-exposure aerial image. Note, the two-exposure option curve is closer to the optimal curve in the dense line region for high thresholds. Optimal curve is the dashed line. (b) Comparison of two- and three-exposure aerial images.

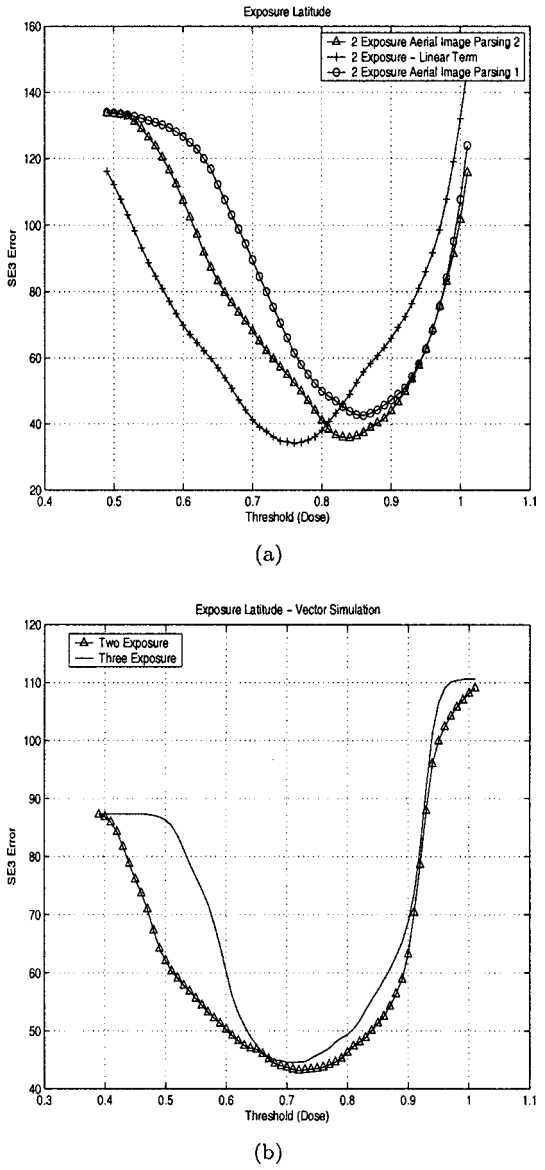

Fig. 9 (a) SE_3 curve versus threshold (dose) at 71 nm for the two-exposure versus the three-exposure options, where the third exposure is the on-axis exposure from aerial images from Fig. 4. (b) Similarly for PROLITH vector simulation.

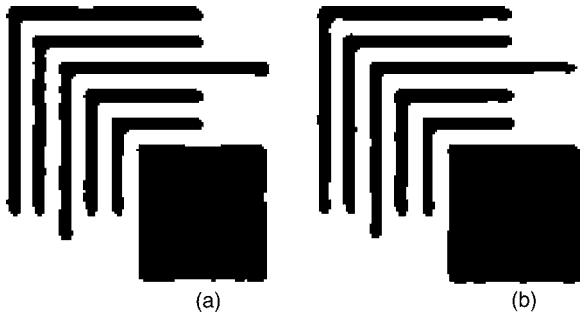

Fig. 10 (a) Optimal threshold image for two exposure at 71 nm with the optimal threshold picked at bottom of curve in Fig. 9(a) and 9(b) optimal threshold image for the three-exposure option at 71 nm on full search. Third exposure is on-axis exposure. Images are similar, but from the SE_3 versus threshold curves, the two-exposure option has a wider curve and more exposure latitude.

ous three exposure systems with different on-axis pupil sizes from NA=0.2, 0.4, 0.6, and 0.8. In this type of imaging, the expected difference is in the quadratic term. The plus curve (+) represents the linear terms in the two-exposure imaging system and the triangle line curve (Δ) is the total aerial image of the two-exposure frequency coverage. Then, the only difference between the + curve and Δ line curve is the quadratic term's effect. For a threshold above 0.8, the three exposures are approximately the same

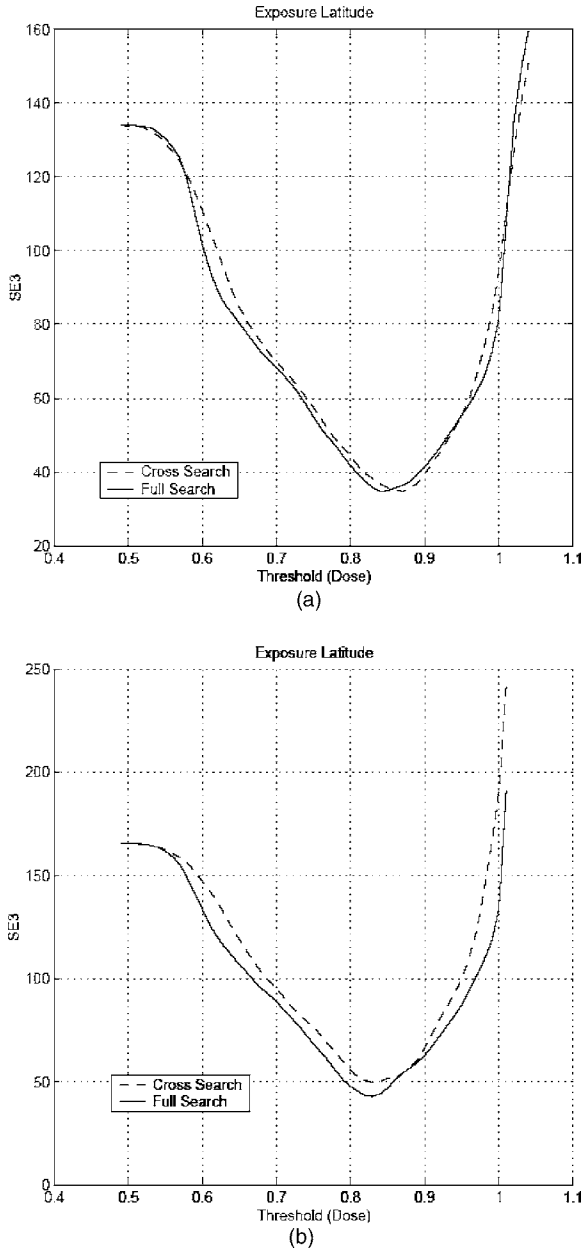

Fig. 11 (a) Comparison of SE_3 error versus threshold curves for two exposures with cross-hair search and full search for test mask 1, and (b) comparison for test mask 2 for full and cross-hair search where the frequency coverages. Note the small difference is due to the type of CD and spacing used for test mask 1 and more separation between curves for test mask 2.

as the + line curve (linear term of two exposure). The three-exposure option curve for thresholds below 0.8 falls to its minimum faster than the two-exposure curve, but for threshold above 0.8 the two-exposure option curve is lower than the three-exposure curve and has more exposure latitude over a range of thresholds (dosages).

In Fig. 11(a), the curves for the two-exposure aerial images are compared for the optimized cross-hair search and full search for frequency coverages of Figs. 4 and 6, respectively. The SE_1 of the image decreases by a half from a $SE_1=354.5$ to $SE_1=174.3$, but only a small gain is seen in

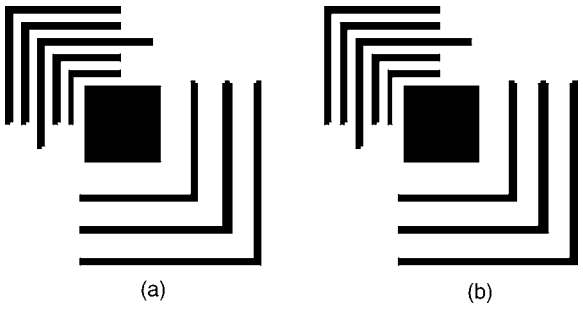

Fig. 12 (a) Test mask 1 and (b) test mask 2.

the second metric, where $SE_2=89.73$ compares to $SE_2=88.67$. When comparing the errors induced into the exposure latitude curves there appears only a small change, as reflected in Fig. 11(a). A second test mask [shown in Fig. 12(b)] is generated to induce more significant Fourier coefficients near the obscuration in the Fourier plane. The exposure latitude curves for this test mask 2 in Fig. 11(b) for full and cross-hair search show more difference over a large set of threshold values since mask 2 contains features with larger pitches.

In Fig. 9(b), the error versus threshold is plotted, but it uses the aerial images from PROLITH's vector model with polarization control.[10] This is shown for comparison with the scalar model figures shown previously to verify the gain in exposure latitude from the quadratic term. One three-exposure setting with an inner pupil at 0.8 (NA/λ) is shown and it has a curvature with generally less exposure latitude than the two exposure. The two-exposure curve is shifted left relative to the other curve in the figure due to the unknown normalization constant and biasing used in PROLITH. It appears there is good agreement between PROLITH and the IIL scalar model simulations with regard to the wideness of the error versus threshold curve and general shape. More comparisons in PROLITH are shown in Ref. 11.

4.3 *Frequency Coverage—GA*

In this section, we discuss the frequency coverages resulting from the GA. The coverages obtained are suboptimal, but they provide insights into a simulation of how an obscure frequency coverage and pupil filtering can change the quadratic term and how the aerial image quality can vary with extreme pupil filtering. These results show a search over the subset of horizontal and vertical frequency axes. Figure 2(a) illustrates the suboptimal frequency coverage. The parsing 1 and 2 strategies result in pupil filters that are shown in Figs. 2(b) and 2(c), respectively. Since the pupil filters manipulate the quadratic term, the resulting quadratic images are shown in Fig. 7. The quadratic term of Fig. 7(c) is very strong with a maximum value of 0.64 and this is also seen in Fig. 13. In the large-line area, a significant bias is introduced. This is likely to be the cause of any artifacts appearing in the large-line region since the bias introduced adds to the ringing that typically occurs in the linear term. As the bias in the flat region nears the threshold, the chances of an artifact arising there increase. However, in the dense region, the quadratic term enhances the exposure

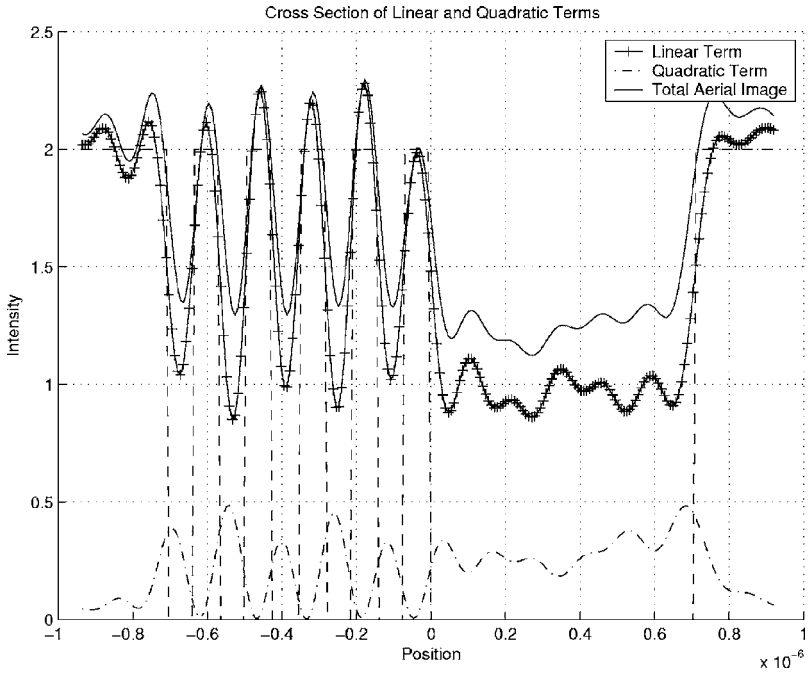

Fig. 13 Cross-hair section of linear and quadratic terms for parsing 1 from frequency coverages in Fig. 2. Quadratic term has strong oscillation in the dense region and strong bias in flat region from the parsing 1 strategy.

latitude making the curve then closer to the optimal mask curve as closely seen in Fig. 14. Figure 14 shows the cross-sectional slice of the aerial image when parsing 2 is used and it illustrates the lower quadratic term bias in the large-line region, and therefore it has a smaller chance to introduce an artifact, although less exposure latitude is given to the dense lines. The bias introduced in the large-line regions is lowered since it is spread throughout the region, instead of being absorbed through a single exposure as done in parsing 1. The corresponding exposure latitude curves are shown in Fig. 15. The symmetric pupil filters induce less error as the threshold is varied because the quadratic term bias is more distributed spatially throughout the large-line regions.

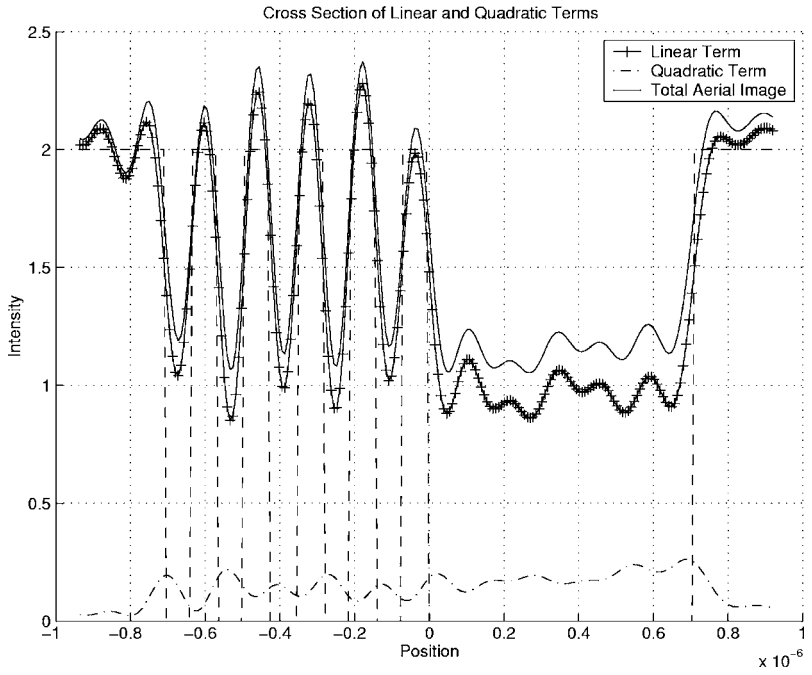

Fig. 14 Cross section of linear and quadratic terms for parsing 2 from frequency coverages in Fig. 2. The quadratic term has less oscillation and less bias in the flat region from parsing 2 strategy and less chance to cause artifacts by pushing the ringing near the threshold.

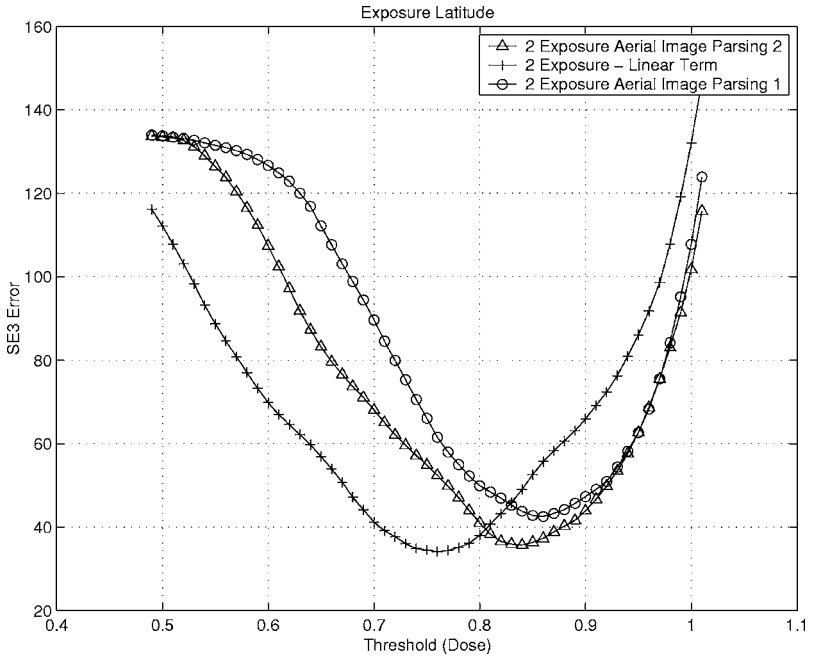

Fig. 15 Exposure latitude comparison for parsings 1 and 2 for coverages from the GA at 71 nm. Two exposures with parsing 2 pupil filters have a wider curve than with parsing 1 pupil filters.

For a three-exposure system, the GA is preferred since it is computationally feasible. A low-NA system of 0.6 is simulated for a three exposure system since a high-NA of 0.9 is sufficient to cover frequency space as is seen in previous figures. In this case, the reference beam is outside the optical system and is inserted at the image plane. Figure 16 shows the first exposure tilted at 135 deg to recover the frequency coefficients at the corners of the pattern and the second and third exposures are placed near the edge of frequency space to cover the vertical and horizontal frequencies.

5 Conclusions

In this paper, we have presented a combinatorial approach to finding the optimal frequency coverage parameters, discussed the associated parsing schemes, and analyzed their impact on the resolution of the aerial image. The optimal exhaustive search approach, dynamic programming, and GAs presented for solving the nonlinear approximation problem can be applied to any type of pattern. The GA, with a lower complexity, optimized the frequency sets for each exposure independently while the exhaustive and dynamic programming approaches optimized the frequency sets for the exposures jointly. We saw that the application of suitable constraints, such as the cross-hair search, can reduce the tendency of the greedy approach to converge to suboptimal solutions.

We have shown that under tilted frequency coverages, the errors in the threshold image are reduced due to the better placement of the obscuration in frequency space. We presented simulation results, at high NAs for a pupil with an obscuration, which illustrate the fact that the lost information in the obscuration can be recovered in the pupil filters without increasing the number of exposures. The optimization algorithms for a disk-shaped pupil return the typical coverage and parsings that have been discussed in prior papers (i.e., Fig. 10 of Ref. 7), but the algorithms presented here also have potential use in low-NA optical systems for disk-shaped pupils, where a majority of the frequency space is not covered and the utility of these algorithms for optimal coverages will be more important. We

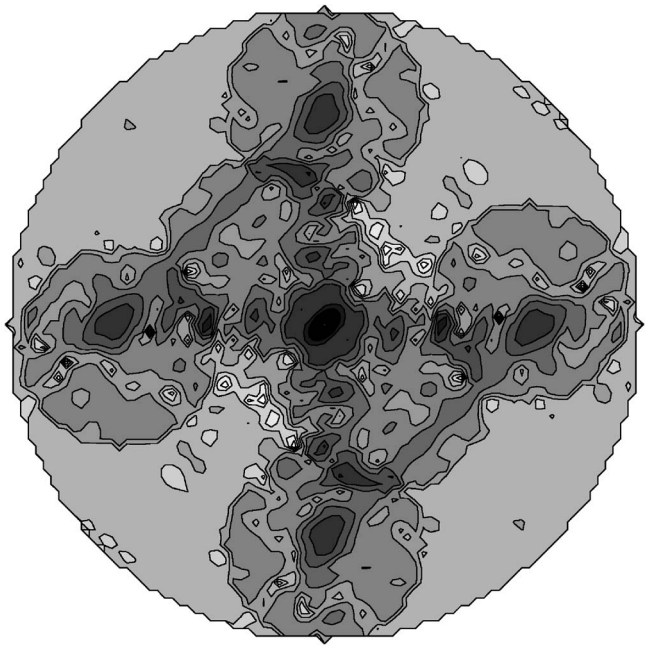

Fig. 16 Three-exposure frequency coverage for NA=0.6 from greedy algorithm for mask 1 with CD=71 nm.

also looked at how Hermitian symmetries in the pupil plane and the geometry of Manhattan pattern masks could be exploited to reduce the size of the search grid in these approaches.

Lastly, we discussed parsing strategies associated with the different frequency coverage algorithms. We saw that the pupil filters control the degree of interference and they affect the quadratic terms of the isolation/dense bias. We showed that under the two-exposure option, the quadratic term has the most interference effect, and the dense edges gain printability, but this comes at the cost of linewidth variation of the isolated edges. This, in turn, dictates the necessity for precise optical alignment. A possible strategy to mitigate this is to introduce SRAF/OPC features to induce a proximity effect near isolated edges to decrease the dense/isolated bias and keep the number of exposures to a minimum. Under the three-exposure system, the quadratic term has less interference between the high and low frequencies since the parsing occurs in the mid-spatial frequency area; thus, removing the resolution degradation, but at the cost of an additional on-axis exposure and narrower exposure latitude curve. As κ_1 factors lower, the nonlinear quadratic term may also degrade and the use of pupil filters (apodization) may be required to make the optical imaging behave linearly such that the dense features can also be resolved well.

Acknowledgments

This work was supported by ARO/MURI under Grant No. DAAD19-99-1-0196.

References

1. M. Levenson, "Wavefront engineering from 500–to 100–nm CD," *Proc. SPIE* **3049**, 2–13 (1997).
2. M. Levenson, "Extending the lifetime of optical lithography technologies with wavefront engineering," *Jpn. J. Appl. Phys., Part 1* **33**, 6765–6773 (1994).
3. M. Levenson, "Wavefront engineering for photolithography," *Phys. Today* **46**, 28–36 (1993).
4. H. Levinson, *Principles of Lithography*, SPIE Press, Bellingham, WA (2001).
5. A. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Vol. TT47, Bellingham, WA (2001).
6. X. Chen and S. Brueck, "Imaging interferometric lithography: a wavelength division multiplex approach to extending optical lithography," *J. Vac. Sci. Technol. B* **16**, 3392–3397 (1998).
7. S. Brueck and X. Chen, "Spatial frequency analysis of optical lithography resolution enhancement techniques," *J. Vac. Sci. Technol. B* **17**, 908–920 (1999).
8. X. Chen and S. Brueck, "Experimental comparison of off-axis illumination and imaging interferometric lithography," *J. Vac. Sci. Technol. B* **17**, 921–929 (1999).
9. X. Chen and S. Brueck, "Imaging interferometric lithography: approaching the resolution limits of optics," *Opt. Lett.* **24**, 124–126 (1999).
10. A. Biswas and S. Brueck, "Simulation of the 45-nm half-pitch node with 193-nm immersion lithography and dipole illumination," *J. Microlithogr., Microfabr., Microsyst.* **3**, 35–43 (2004).
11. T. Tridhavee, B. Santhanam, and S. Brueck, "Optimization and apodization of aerial images at high NA in imaging interferometric lithography," in *Optical Microlithography XVII*, *Proc. SPIE* **5377**, 1544–1554 (2004).
12. A. Rosenbluth, S. Bukofsky, M. Hibbs, K. Lai, A. Molless, R. Singh, and A. Wong, "Optimum mask and source patterns to print a given shape," *J. Microlithogr., Microfabr., Microsyst.* **1**, 13–30 (2002).
13. E. Wu, B. Santhanam, and S. Brueck, "General framework for parameter optimization in imaging interferometric lithography," *J. Microlithogr., Microfabr., Microsyst.* **4**, 023009 (2005).
14. Y. Pati and T. Kailath, "Phase shifting masks for microlithography: automated design and mask requirements," *J. Opt. Soc. Am. A* **11**, 2438–2452 (1994).
15. Y.-T. Wang, Y. Pati, J.-W. Liang, and T. Kailath, "Systematic design of phase-shifting masks," *Proc. SPIE* **2197**, 377–347 (1994).
16. Y. Granik, "Source optimization for image fidelity and throughput,"*J. Microlithogr., Microfabr., Microsyst.* **3**, 509–522 (2004).
17. K. Tounai and N. Aizaki, "Resolution improvement of isolated line pattern in quarter-micrometer level by layout-optimized assistant pattern method," *Proc. SPIE* **2726**, 82–87 (1996).
18. K. Tounai, S. Hashimoto, S. Shiraki, and K. Kasama, "Optimization of modified illumination for 0.25-μm resist patterning," *Proc. SPIE* **2197**, 31–41 (1994).
19. R. DeVore, "Nonlinear approximation," *Acta Numerica* **7**, 51–150 (1998).
20. D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley & Sons, New York (1969).
21. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA (1993).
22. H. Lewis, *Data Structures & Their Algorithms*, Addison-Wesley, Reading, MA (1991).
23. S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press (1999).

Thanis M. Tridhavee received his BS degree in electrical engineering from the University of Illinois, Urbana-Champaign in 1999 and his MS degree in electrical engineering at the University of New Mexico in 2003. He has worked at Lucent Technologies as a member of the technical staff in Naperville, IL, working on and to the Brazil telecommunication network. In addition, he has been a co-op student at Advanced Micro Devices in Austin, Texas working on product development of a codex IC chip. He also has been a research assistant at the U.S. Army Construction Engineering Research Lab in Champaign, Illinois, working on military gun range noise data. His research interests include resolution enhancement technology, imaging interference lithography, and optical signal processing.

Balu Santhanam received his BS degree in electrical engineering from St. Louis University, St. Louis, Missouri, in 1992. He obtained his MS and PhD degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1994 and 1998 respectively. From 1998 to 1999, he was a lecturer and postdoctoral researcher with the Department of Electrical and Computer Engineering at the University of California, Davis. In 1999, he joined the faculty of the Department of Electrical and Computer Engineering at the University of New Mexico, Albuquerque, where he is presently an associate professor of electrical and computer engineering. He currently serves as the associate editor for the International Journal on Computer and Electrical Engineering, Elsevier Science. His current research interests include multiple access interference reduction in CDMA systems, multi-component AM-FM signal separation, modulation/demodulation, and time-frequency representations for nonstationary signals, imaging interferometric lithography and IIL-related parameter optimization for sub-wavelength optical lithography. He is a member of the SPIE, the signal processing and communication societies of the IEEE, and a member of the ASEE.

Steven R. J. Brueck received his BSc from Columbia University in 1965 and his MSc/PhD from MIT in 1967/1971. In 1971, he joined the Quantum Electronics Group at MIT Lincoln Laboratory. In 1985, he became the director of the Center for High Technology Materials (CHTM) and a professor of electrical and computer engineering and physics at the University of New Mexico. He has made extensive experimental and theoretical research contributions in many aspects of optics and laser spectroscopy. Material systems investigated include semiconductors, simple molecular liquids, and plasmas used for semiconductor processing. Under his direction, CHTM has become an internationally recognized center for optoelectronics and microelectronics research. He has published over 200 research articles, has edited 7 books, been awarded 21 patents, and serves as an editor in many journals. He is a fellow of IEEE and OSA societies and recipient of the IEEE Third Millennium Award.