Optimum Mask and Source Patterns to Print a Given Shape

Alan E. Rosenbluth[a], Scott Bukofsky[b], Michael Hibbs[c], Kafai Lai[b], Antoinette Molless[b], Rama N. Singh[a], and Alfred Wong[b]

[a]IBM T.J. Watson Research Ctr., Yorktown Heights, NY, [b]IBM Semiconductor Research and Development Center, Hopewell Jct., NY, [c]IBM Microelectronics Div., Essex Junction, VT

ABSTRACT

New degrees of freedom can be optimized in mask shapes when the source is also adjustable, because required image symmetries can be provided by the source rather than the collected wavefront. The optimized mask will often consist of novel sets of shapes that are quite different in layout from the target IC patterns. This implies that the optimization algorithm should have good global convergence properties, since the target patterns may not be a suitable starting solution. We have developed an algorithm that can optimize mask and source without using a starting design. Examples are shown where the process window obtained is between 2 and 6 times larger than that achieved with standard RET methods. The optimized masks require phase shift, but no trim mask is used. Thus far we have only optimized 2D patterns over small fields (periodicities of ~1µm or less). We also discuss mask optimization with fixed source, source optimization with fixed mask, and the re-targeting of designs in different mask regions to provide a common exposure level.

Keywords: Off-axis illumination, source optimization, RET, OPC, global optimization.

INTRODUCTION

An important synergy can be exploited in jointly optimizing mask and source to print a given shape. Often the resulting mask and source patterns fall well outside the realm of known design forms. For this reason it is desirable that the optimization algorithm provide good global performance, e.g. the algorithm should not be constrained to use a known starting design. Our work suggests that standard OPC approaches may have difficulty converging on the mask solution that is globally optimal.

Previous work on optimization of the source alone has described general algorithms[1] and specific implementations[2-4] for customizing illumination to print particular shapes. Enhancement techniques to customize masks (e.g. RET methods like assist features, serifs, phase tiling, etc.) are usually applied as adjustments or modifications to the nominal circuit patterns. In formal terms, one can say that the nominal patterns (or some simple extension of them) effectively serve as the starting solution when masks are optimized.

In this respect RET technologies are linked to classical lithography, wherein axially illuminated mask shapes that reproduce the target patterns are used to project a wavefront with all attendant symmetries into the lens. The wavefront section collected by the lens (whose finite NA acts as a cutoff filter) is likewise symmetrical under axial illumination, and as a result the input

Figure 1 – Degrees of freedom in collected wavefront using different illumination directions. Reticle phases other than 0° or 180° are ruled out to prevent distortions through focus. a) Only 2 independent orders are collected under axial illumination, since +1 and −1 orders must be complex conjugates when reticle transmittance is real-valued. b) 3 independent orders can be collected from (sufficiently oblique) illumination directions, aiding optimization. c) Stability through focus is restored by illuminating reticle from mirrored directions.
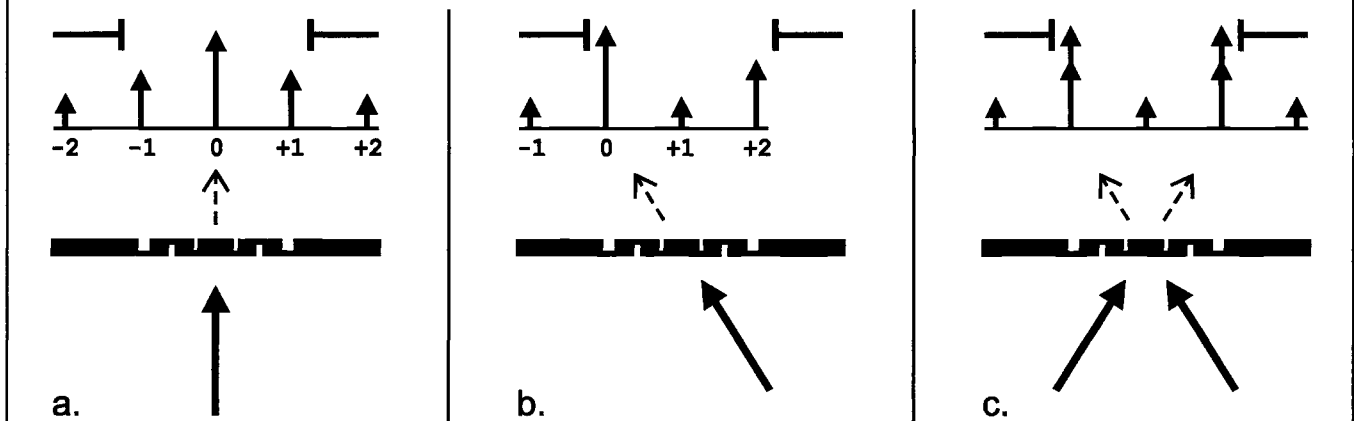
Figure 2 – Capacitor pattern. Horizontal period is 260nm, vertical period 390nm. Rectangles (130nm×247nm) are bright. Dashed boundary shows plot area for images in later figures.
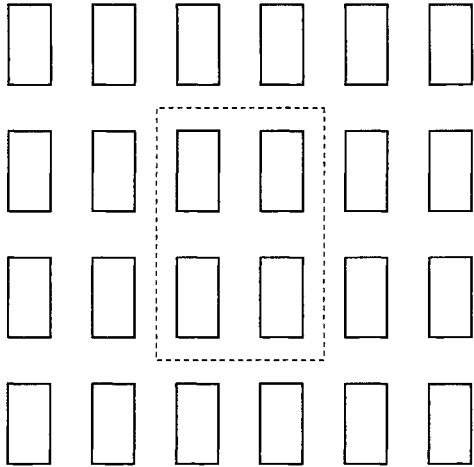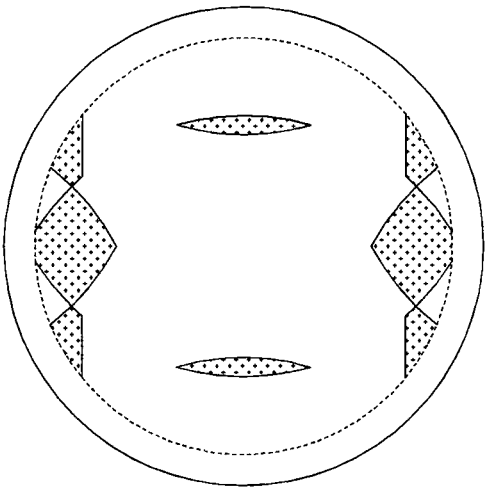
Figure 3 – Optimized source for Figure 2 capacitor pattern. λ=248nm, NA=0.68 (solid circle). Process window through focus is optimized. Hatched areas are bright. Dashed circle is 0.85 σ limit. Figure 4 shows mask.

symmetry is transferred to the image. Wavefront symmetry constraints include Hermitian radial symmetry (if the reticle phase is restricted to 0° or 180° to avoid distortions through focus), as well as any bilateral symmetries that the target pattern may have.

These constraints substantially reduce the number of truly independent orders that can be collected under axial illumination. Once a particular positive order is determined, the corresponding negative order is also fixed (to within an unimportant translational phase). From an optimization viewpoint, the quasi-symmetry of typical wavefronts implies that the number of degrees of freedom in the lithographic image will be little larger than that corresponding to one quadrant of the NA, or half the NA if the mask shapes are highly non-symmetric (but still restricted to 0° or 180° phase). Figure 1 illustrates this idea in schematic form.

However, when we illuminate the mask obliquely it is not necessary to impose a symmetry constraint on the decentered section of the wavefront that is collected. In practice the illumination is limited to e.g. $\sigma \leq 0.85$; in this case the number of truly independent diffraction orders that can be addressed from an oblique illuminating direction will typically be ~2X larger than can be addressed with axial illumination. In many cases the availability of these extra degrees of freedom significantly enhances the quality of the optimized solution, and we can restore the required symmetries and focal insensitivity to the printed pattern by using a suitably symmetric source. The optimized diffraction pattern will therefore tend to be dominated by the way in which diffraction orders combine coherently from illumination directions that are strongly non-axial, thereby forming the dominant image components of the incoherent sum.

The collected set of oblique orders usually has more structure after optimization than would be present with e.g. the typical diffraction falloff from coarse mask rectangles. This means that if the optimized mask were to be illuminated axially rather than obliquely, a completely different interference pattern would usually be produced on the wafer (since the centered collection of orders would combine some subset of the optimized oblique orders and matching negative orders in an undesirable way). The axial image is therefore unlikely to resemble the optimized wafer image (which resembles the target pattern). It also follows that the optimized reticle pattern, which can be thought of as comprising a very large number of axially centered orders, will likewise tend not to resemble the optimized image (or the target pattern).

This means that enhancement techniques which use the target patterns as a starting solution may not provide fully optimized reticles when the source shape can be freely adjusted. Note that most algorithms for nonlinear optimization are essentially local minimizers, and so are strongly dependent on the quality of the starting solution. Of course, lithographers face no explicit requirement to begin the design process using any particular trial layout; indeed, global algorithms are of interest as conceptual tools for bringing forward new design forms.

Casual experimentation with a local optimization routine suggests that changing the magnitude of individual orders by ~0.3 can move a trial solution into the vicinity of a new local minimum (in a test case where the average order intensity was set to about 1). This sensitivity reflects the oscillatory nature of the plane wave components that define the image. If we suppose that the orders typically span a range from about −3 to +3, and that the minimum field size needed to adequately bound the tails of the lens resolution (e.g. ~2λ/NA) can be characterized by 7 collected diffraction orders (allowing non-axial illumination, but counting only truly independent orders), then if we wish to find globally optimal values for these amplitudes via the simple expedient of trying a large number of starting solutions, we would be required to run the optimizer from roughly 3.2×10^7 different starting points. Inclusion of the source variables entails a further combinatorial explosion.

This estimate is crude, but it demonstrates that even the most robust local convergence is insufficient for RET optimization. To address this disadvantage we have devised global algorithms that can optimize mask and source to print a given shape without using a starting design. The wavefront from any individual off-axis direction is allowed to have arbitrary decentration (above

Figure 4 – a) Optimized mask patterns (chromeless) for Figure 2 capacitor pattern. Black represents 0° phase-shift, white 180°. Area shown corresponds to dashed region of Figure 2. b) Aerial image (screen capture from Prolith[5] simulation). c) Superposition of mask and image. The "battery-shaped" mask features create dark horizontal separations in the image, and are positioned in between the bright image rectangles. Pattern layouts on mask and wafer are quite different.

0.889
0.808
0.728
0.648
0.568
0.488
0.407
0.327
0.247
0.167
0.087
0.007

a.   b.   c.
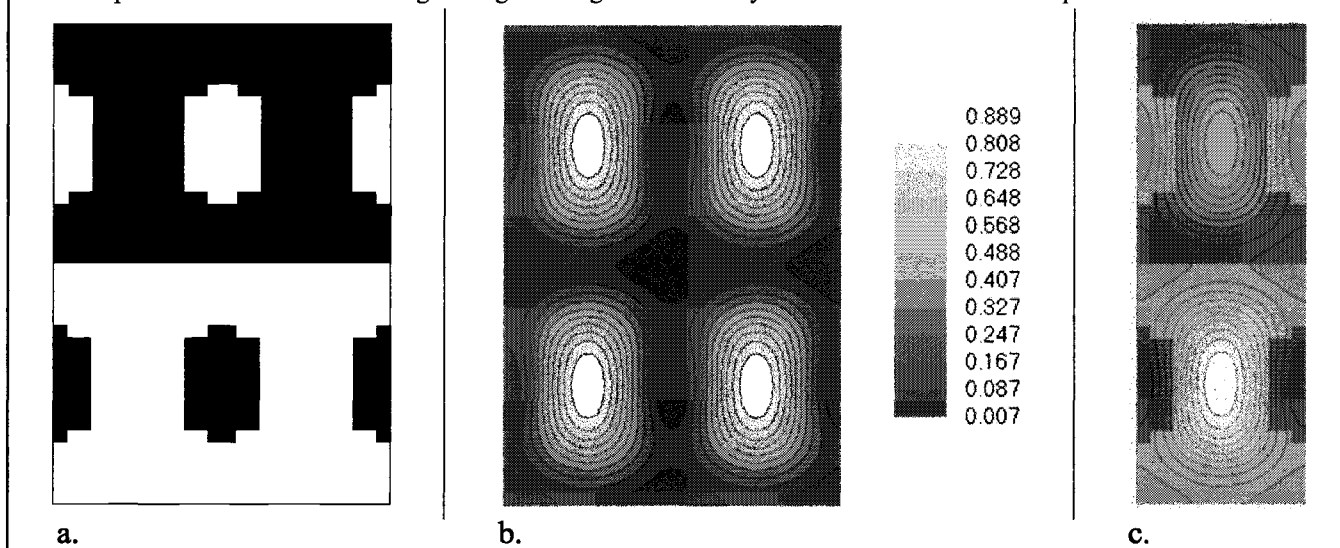
and beyond that produced by the tilted illumination), and arbitrary lateral asymmetry. Focal tilt and bilateral asymmetries in the final image are removed by using symmetric illumination distributions. Several simplifying approximations are adopted, but many of these approximations can be avoided in the sub-problems of calculating the optimal mask for a given source, the optimal source for a given mask, and the most efficient mask to produce a given set of collected orders (yielding fully global solutions to these sub-problems under the formulations given below). An optimized wavefront generally requires 180° phase shift in the mask, which can be provided by either attenuating chrome, chromeless shifters, or phase-reversed openings in opaque chrome. No trim mask is used.

In the present paper we will describe a global optimization algorithm that uses exposure latitude as the merit function. However, we have made considerable progress toward developing an algorithm that optimizes against full process window through focus, and we will show results from this latter algorithm as developed thus far.

Consider, for example, the DRAM capacitor level shown in Figure 2. One critical dimension in this pattern is the width of the printed rectangles (bright for positive resist), which in this example we take to be 130nm. Though difficult, it is desired that the rectangles print with an aspect ratio of at least 1.9:1. At low k-factor this elongated aspect ratio poses considerable difficulty for conventional RET methods. The DRAM cell uses a 2F×3F layout[6], and the pitch ratio is only 1.5:1. Contrast in the dark

Figure 5 – Successively defocused image slices from Figure 3,4a solution, taken through centerlines of bright rectangles. Dashed are vertical slices, solid horizontal. Images are normalized against peak intensity of 89%.
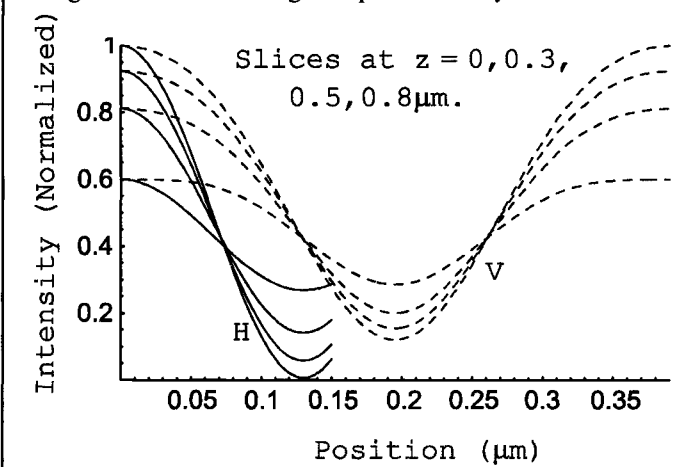

Slices at z = 0,0.3, 0.5,0.8µm.

Figure 6 – Process window obtained with the solution of Figures 3 and 4a. An aberration-free lens is assumed. CD tolerances are ±15nm on width, ±30nm on length. Curve is calculating from thresholded aerial images. Horizontal axis is single-side defocus, equal to half DOF. Integrated window (two-sided) is 45%-µm.
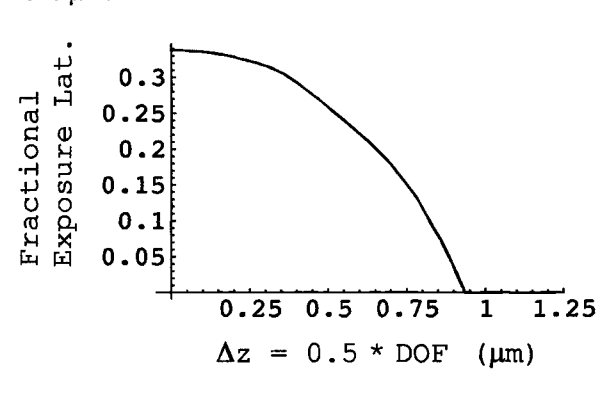

$\Delta z = 0.5 * DOF$ (µm)

Figure 7 – Solution for Figure 2 pattern in attenuating phase-shift chrome. The area shown corresponds to Figure 4a,b, and to dashed region of Figure 2. Mask openings are shown white. Chrome transmission is 6.5%, phase-shifted 180° (black shaded).
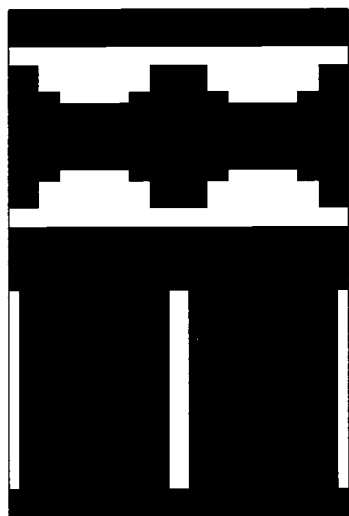

Figure 8 – SEM image of chromeless mask that implements Figure 4a solution.

gaps that separate the rectangle tips is poor, and the rectangles tend to print with considerable shortening. When shortening is compensated by narrowing the gaps, contrast degrades further. For example, at λ=248nm and NA=0.68, even an ideal thresholded aerial image model predicts that we will only be able to print the array using an attenuated phase-shift mask (T=6.5%) and annular illumination if we allow fairly relaxed CD tolerances, and accept poor contrast in the dark separations between the tips of the rectangles. If we impose a requirement that the intensity at the center of the focused rectangle be at least 3 times larger than that midway between the tips (i.e. if we allow the feature to be biased to the point that max-to-min contrast in a vertical slice across the tips drops as low as 3:1), then the ED window achieves a DOF of ±0.56µm when tolerances of ±30nm and ±15nm are applied to the length and width, respectively. The process window is 7%-µm (using integrated area under the two-sided ED curve as the process window metric[7]). If we remove all constraints on contrast, biasing can increase theoretical process window to 16%-µm, but contrast drops to 2.3:1. Experimentally, such low contrasts prove unusable, and printed resist images show zero common process window for length and width using conventional enhancement methods, unless separate exposures are used to print alternate rows of the array[6].

Figures 3 and 4a show the result of optimizing mask and source to print the Figure 2 pattern (at λ=248nm, NA=0.68), using process window through focus as the merit function. Image slices are shown in Figure 5. A chromeless mask technology is used, though the same underlying solution can be realized in essentially any mask technology that provides 180° phase shift. Note that the bright rectangular features in the image actually fall in between the vaguely brick-like openings in the reticle, i.e. the direct resemblance of these reticle shapes to the image patterns is coincidental. Indeed, the reticle shapes in Figure 4 that are optimized for off-axis illumination have a distinctly different "topology" from the image shapes, i.e. their basic layout has a different internal connectedness. It would have been quite difficult for a conventional optimizer to have devised a path of smooth and continu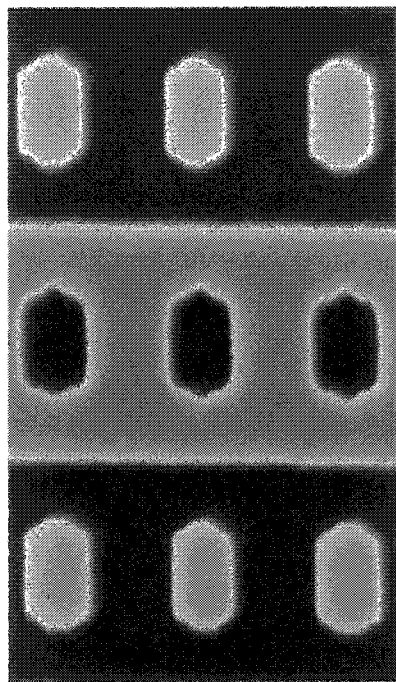ous adjustments that reached the Figure 3,4a solution from starting shapes that matched Figure 2; moreover, even if such a path could be defined, a local algorithm would not follow it unless process window increased monotonically at every point.

Figure 6 plots the ED window obtained with the optimized source and

Figure 9 – Approximate realization of the Figure 3 source. a) Simplified aperture pattern, designed to ease fabrication of stencil illumination stop in model shop. b) Pupil-gram (highly defocused image through mask pinhole) showing the illumination pattern actually realized in the exposure tool. Discretization from the light tunnel homogenizer is apparent.
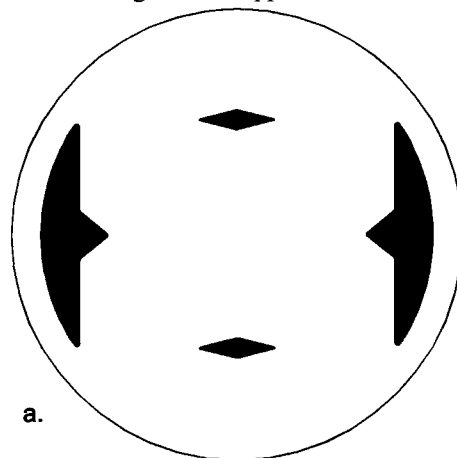

a.
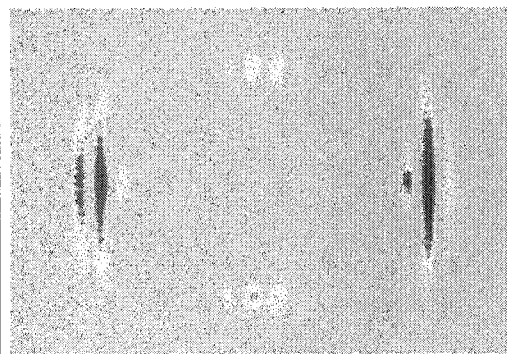
b.

Figure 10 – Idealized model of source discretization by homogenizer. a) Source pattern. The input σ=0.85 disk is sparsely filled, simulating the effect of homogenizing optics in a slot-field exposure tool. Plot shows source pattern after truncation by ideal Figure 3 aperture. b) Difference between image with discretized Figure 10a source, and ideal image (continuous Figure 3 source).
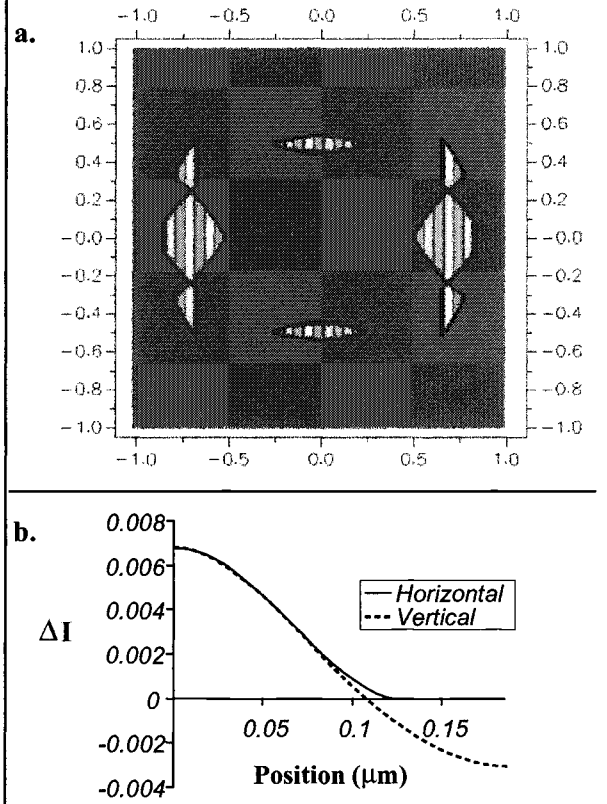
reticle, using the same ±30nm and ±15nm tolerances on length and width considered above. Integrated process window is 45%-μm under a thresholded aerial image model (assuming no aberrations). This is between 3X and 6X better than the calculated performance of standard enhancement methods (see above). Max-to-min contrast across the rectangle tips is 8.2:1, also much improved over the conventional result. The solution in Figures 3,4a was obtained by optimizing against process window; however a similar solution with quite good process window (37.6%-μm) is obtained by optimizing against exposure latitude in focus (algorithm P below). The optimized solutions can also be realized in attenuating phase-shift masks. The attenuating phase-shift solution in Figure 7 achieves the same large process window as the Figure 4a chromeless solution; however, overall intensity is quite low because the optimizer has realized the solution by printing through the chrome as if it were a "hard" phase shifter.

Our global optimization approach provides novel design forms with high theoretical performance. Of course, in practice lithography cannot really be reduced to a purely formal optimization. After describing our method in more detail we will comment briefly on some issues of practical implementation. We will also discuss the prospect for extending our methods to optimize multiple patterns simultaneously. Global methods show promise for increasing the common process window of a suite of patterns. Indeed, in principle the common process window for a globally optimized set of patterns cannot be lower than that provided by conventional optimization methods. However, as with conventional methods, the common process window cannot be larger than is achieved for a single pattern when optimized individually. Pattern diversity is limited within the field sizes that can be optimized at present (~1μm), and source solutions for such small fields tend to be fairly specialized. Source directions at large-σ along the 45° azimuths tend to maximize the number of collected degrees of freedom, providing an advantage in optimizing a diverse set of patterns.

EXPERIMENTAL TEST

Though the treatment in this paper is primarily theoretical, we felt it important to include an experimental demonstration of the theory. Figure 8 shows our implementation of the Figure 4a chromeless mask. To obtain results within deadline for the present proceedings, we implemented the source of Figure 3 in the form of a simple illumination stop (located in a plane conjugate with the entrance pupil), and adopted the simplified hole pattern shown in Figure 9a for ease of fabrication. Figure 9b shows a measurement of the illumination pattern as realized in the exposure tool. The source apertures are sparsely filled because the input σ=0.85 disk is realized by discrete multiple foldings within a homogenizing rod. The exposure tool uses a scanning slot field, so the input source appears striped under the limited resolution provided by the pupil-gram. This coarse discretization would not be present if the custom source were defined by diffractive elements[8]; moreover, the impact of this discretization need not in principle be severe, as shown in Figure 10. However, considerable source distortion was incurred in the present experiment (compare Figure 3 with Figure 9b).

Nonetheless, we achieved reasonable wafer images with this compromise source, as may be seen in Figure 11a. Figure 12 shows focus/exposure data from the experiment (top-surface SEMs). Measured exposure latitude is about 14%, DOF approximately 0.7μm, and process window roughly 7%-μm. This is quite a respectable result (though well below the ideal performance of the Figure 6 simulation), considering that in practice the pattern proves impossible to print within tolerance using conventional enhancement methods[6]. The investigations reported in ref.[6] show that capacitor aspect ratio for 130nm trenches is limited to about 1.4:1 when annular illumination and phase-shift chrome are employed, even if the pitch is relaxed slightly to permit increased mask bias. Figure 11b shows the approximate limit of what can be achieved (same NA and λ as Figure 11a, but different exposure tool). Because of the narrow separation between adjacent capacitors, it is impossible to introduce a bias sufficient to meet tolerance unless every other row in the array is removed from the mask to free up more real estate; the array must then be printed in two separate exposures (see Figure 6 in ref.[6]).

Figure 11 – Images of Figure 2 pattern in 5300Å of UV82 resist, exposed at λ=248nm, NA=0.68. a) Exposures using the Figure 8 mask and Figure 9b source. Pitch is 260nm horiz., 390nm vert., per Figure 2. b) Attempt to print elongated capacitors of 130nm width using conventional enhancement methods (annular illumination, phase-shift chrome, mask bias), and expanded pitch (relaxed to 300nm horiz., 405nm vert.). Figure 11b image was scaled to same magnification as Figure 11a using graphics software.
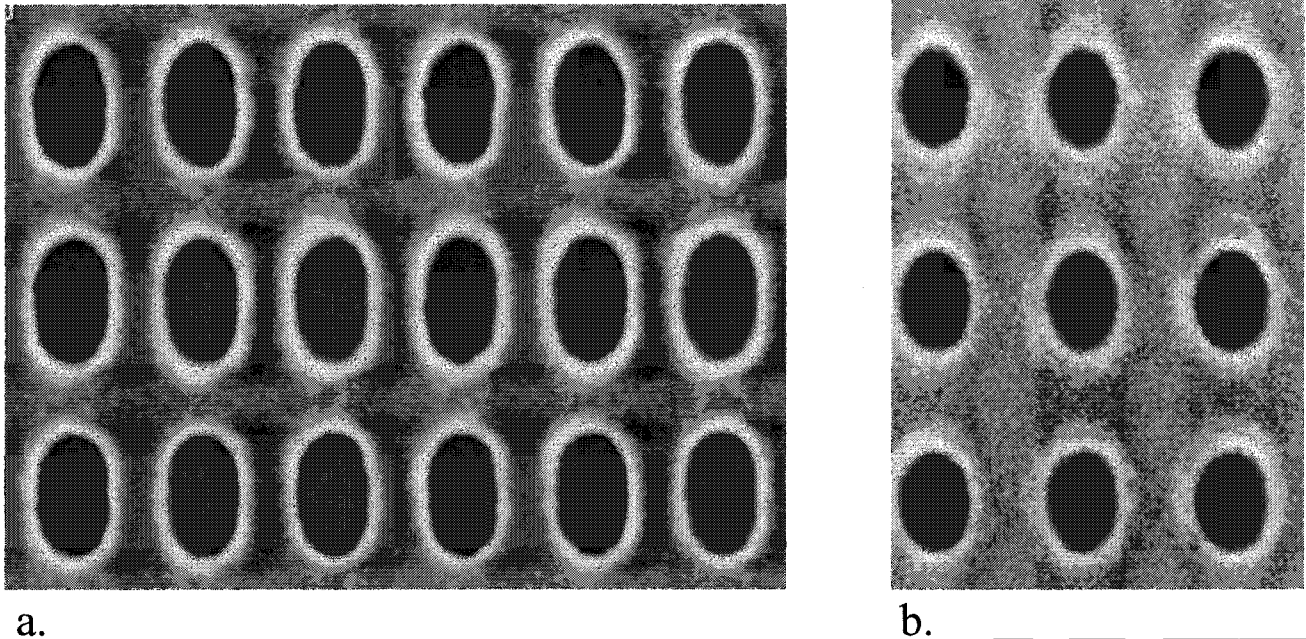

a.   b.

Figure 12 – Focus-exposure measurements using the Figure 9b source and Figure 8 mask. Each point represents the maximum CD error found in an adjacent pair of measurements. Errors are normalized, so that 1.0 represents the tolerance limit (±15nm horiz., ±30nm vert.). Gray triangles are width errors, black rectangles are length errors. Solid lines are nominal dose; dashed and dotted lines show the effect of increasing or decreasing dose by 4%.
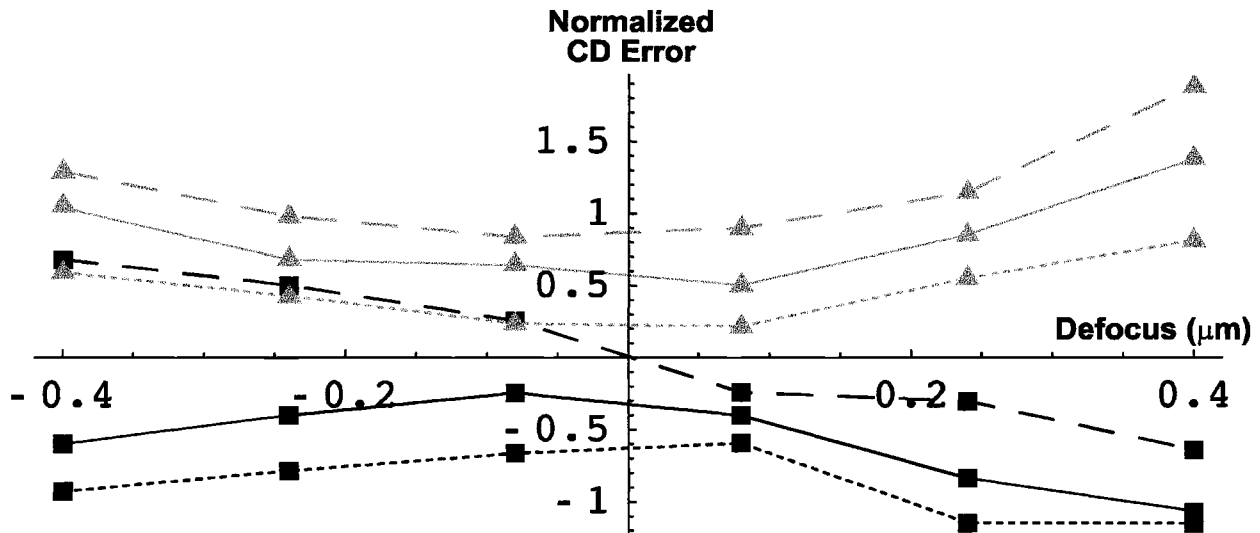
ALGORITHM TO OPTIMIZE EXPOSURE LATITUDE

We now describe an algorithm for global optimization of mask and source against exposure latitude in focus. First, we note that highly efficient algorithms have been developed for local optimization[9]; these are available, for example, in packages like MATLAB[10], Mathematica[11], and IMSL[12]. Such algorithms can converge to local maxima in the merit function within polynomial time, even when the merit function is nonlinear. If one can model the system in the "forward direction", and if one

Figure 13 – Pupil diagram for array with staggered pitch. X pitch is 1120nm, y pitch is 560nm, and one basis vector is diagonal. Lens pupil radius (NA) is 0.68 (heavy circle). Dashed circle indicates $\sigma_{Max} = 0.85$. Diffraction orders (under axial illumination) are plotted as gray points. Circles of radius NA are erected about each order. Numbered overlap regions (53 in all) are source variables.
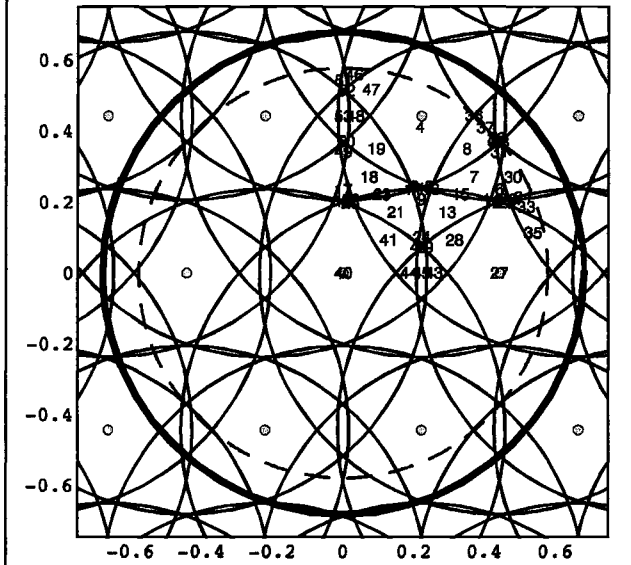
can devise a merit function to quantify the suitability of a given solution, then a nonlinear optimizer will efficiently refine a given starting design so that it converges to the nearest local maximum of the merit function.

In the case of global optimization, it has been proven that for a general merit function no global algorithm can be guaranteed to perform better than simple exhaustive grid search of the parameter space [Nemirovsky and Yudin, as cited in ref.13]. However, by exploiting the particular structure of the lithographic problem we can find solutions on a far more rapid basis. Knowledge of this special structure provides a very strong advantage. For example, our tests of two general-purpose global optimization programs found them unable to solve even limited sub-problems (e.g. source held fixed) of joint source/mask optimization problems that our specialized algorithms can handle.

The difficulty in lithographic problems is that the merit functions are usually not convex; indeed, the plane-wave orders that comprise the image are intrinsically oscillatory, giving rise to a great many local maxima. To achieve efficient global performance we adopt the following two-part strategy:

1) Seek the global solution to a simplified version of the problem.

2) Use a local optimizer to refine the step 1 solution against a more complete model.

The robustness of widely available local optimization routines allows us to divert many detailed optimality criteria to step 2. Step 1 is solved under a scalar aerial image model.
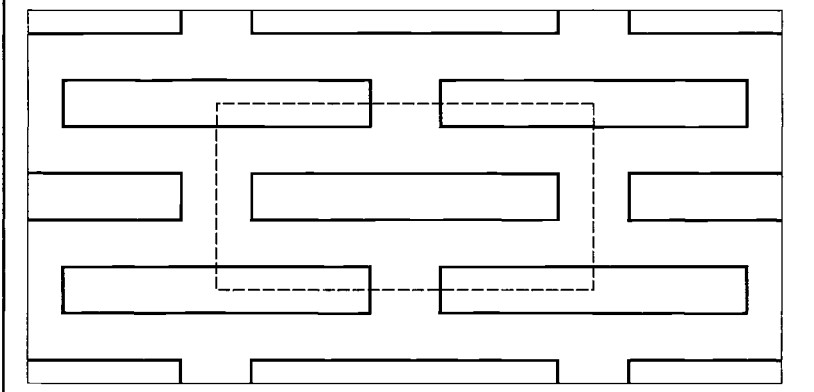
The imaging solution determined in steps 1 and 2 is defined in the pupil plane (as a set of illumination and diffraction amplitudes), so to complete the solution we add a third step:

3) Calculate a reticle pattern that provides the optimized wavefront determined in step 2.

We describe below a simple approach to step 3 which exploits the linearity of the diffraction Fourier transform. Step 2 can be handled by standard routines, as we have seen. For the more difficult step 1 global optimization we simplify the problem by considering only an aberration-free image (aberrations can be deferred to step 2). Further, the algorithm described in this section optimizes only the focused image during step 1, i.e. defocus aberration is also zero. Of course, the step 2 local refinement need not be restricted to optimization of exposure latitude.

With target patterns that are periodic (or to which we apply periodic boundary conditions), optimization of a focused image allows us to partition the continuous space of possible illumination directions into a fairly small number of distinct regions, since two illumination directions are equivalent (when aberrations including defocus are zero) if they direct the same set of diffraction orders into the collection pupil. This is illustrated in the k-space diagram of Figure 13. The entrance pupil (centered on the origin) has radius NA=0.68 in this example. σ_{Max}=0.85 is assumed as the illumination limit imposed by the stepper (shown as a dashed circle). The optimization program next divides the entrance pupil into independent source regions whose boundaries are formed by circles of radius NA centered on each diffraction order. The diffraction orders plotted in Figure 13 assume λ=248nm, and a staggered array with 1120nm horizontal pitch and 560nm vertical pitch.

Figure 14 – Isolation pattern with periodicity matching Figure 13. Width of dark rectangles is 140nm; separation between tips is 210nm. Later figures plot optimized images over the region shown dashed.

Figure 14 shows a DRAM isolation pattern laid out on such an array. The rectangles (dark for positive resist) have width F equal to 140nm. The vertical spacing of the rectangles is also F, and their length 6.5F. The desired horizontal separation between the rectangles is 1.5F.

The diffraction pattern shown in Figure 13 is produced by illumination on axis. The orders shift as the illumination is tilted, but the associated array of pupil-sized circles should be considered fixed in the lens aperture. Each circle then represents the range of illumination directions for which a given order can be collected, and each overlap region represents a range of illumination directions that provides the same set of collected orders. We can without loss of generality represent the fourfold symmetric source which optimizes any focused image (laid out on the Figure 13 pitch) using 53 distinct variables, with each variable representing the illuminating intensity from one of the different pupil regions identified in the Figure 13 construction. We will denote these unknowns as a vector variable $\bar{s}$ (of length 53 in this example). Note that each element of $\bar{s}$ represents a set of 1, 2, or 4 equally intense illuminating directions that impinge on the mask from mirrored directions. Usually the illuminator will fill all open illumination directions with a fixed power per unit solid angle; in this case the variables must be constrained according to

$$0 \leq s_j \leq S_{Max,j} , \qquad\qquad [1]$$

where $S_{Max,j}$ is the area of the jth illumination region in the pupil. If the source distribution is defined by diffractive elements it is more appropriate to constrain the summed intensity.

The m,nth diffraction order would ordinarily be defined as the amplitude $a_{m,n}$ that (under axial illumination) diffracts from the reticle in a direction $\bar{k} = k_0(m\lambda/p_x, n\lambda/p_y)$, with p_x and p_y the unit cell periodicities. However, for our problem it is desirable that the unknown amplitude variables represent independently adjustable components of the wavefront, and individual collected orders as conventionally defined are not all independent. For bilaterally symmetric patterns we adopt a notation in which m and n are non-negative; $a_{m,n}$ then represents a single non-redundant unknown. Thus, in the Figure 13 example, three independent orders ($a_{0,0}$, $a_{1,1}$, $a_{2,0}$) are collected with axial illumination [source region 40], whereas seven are collected under illumination from off-axis region 8 ($a_{0,0}$, $a_{1,1}$, $a_{2,0}$, $a_{3,1}$, $a_{2,2}$, $a_{0,2}$, $a_{4,0}$).

For a given source direction j, the normalized wafer-plane amplitude $b_{m,n,j}$ that is produced by an unknown amplitude $a_{m,n}$ may then include the result of interference between superimposed waves from the $\pm m,\pm n$ directions. In other words, $b_{m,n,j}$ may be given by

$$b_{m,n,j} =$$

$$e^{2\pi i\left(\frac{m\,x}{p_x}+\frac{n\,y}{p_y}\right)}, \text{ or } 2e^{2\pi i\frac{m\,x}{p_x}}\cos\left(\frac{n\,y}{p_y}\right), \text{ or } 2e^{2\pi i\frac{n\,y}{p_y}}\cos\left(\frac{m\,x}{p_x}\right), \text{ or } 4\cos\left(\frac{m\,x}{p_x}\right)\cos\left(\frac{n\,y}{p_y}\right), \qquad [2]$$

depending on whether or not particular negative orders in the x,y mirror directions are collected simultaneously. It is convenient to write the $a_{m,n}$ and $b_{m,n,j}$ quantities as vectors; $\bar{a}$ for the unknown order amplitudes (including all orders that can be captured from at least one feasible illumination direction), and $\bar{c}_1$ and $\bar{c}_2$ for the real and imaginary parts respectively of $\bar{b}$. To provide proper symmetry in the image we illuminate the reticle symmetrically from mirrored directions, which we distinguish with an index q. Using an index h to separate real and imaginary parts, we have for the image intensity

$$I(x,y) = \sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j\left(\bar{c}_{q,j,h} \cdot \bar{a}\right)^2 . \qquad\qquad [3]$$

To optimize exposure latitude we now seek the global solution to the generalized fractional programming problem:

$$\text{Maximize} \quad \Psi(\bar{s},\bar{a}) \equiv \underset{r}{\text{Min}} \left(\frac{\displaystyle\sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j\left(\bar{c}_{q,j,h,r} \cdot \bar{a}\right)\left(\nabla_\perp \bar{c}_{q,j,h,r} \cdot \bar{a}\right)}{\displaystyle\sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j\left(\bar{c}_{q,j,h,r} \cdot \bar{a}\right)^2} \right)$$

subject to:

$$\sum_{j=1}^{J_{Max}} s_j \geq S_{Min} , \qquad\qquad [4]$$

$$0 \leq s_j \leq S_{Max,j} \qquad\qquad (\forall j \,|\, 1 \leq j \leq J_{Max}),$$

$$\sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j\left(\bar{c}_{q,j,h,r} \cdot \bar{a}\right)^2 = \text{non-preset constant Q, independent of r} \qquad (\forall r \,|\, 1 \leq r \leq r_{Max}),$$

$$\sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j \left(\bar{c}_{q,j,h,u} \cdot \bar{a} \right)^2 \geq I_{Bright} \, Q \qquad\qquad (\forall u \mid 1 \leq u \leq u_{Max}),$$

[4 cont.]

$$\sum_{q=1}^{4} \sum_{j=1}^{J_{Max}} \sum_{h=1}^{2} s_j \left(\bar{c}_{q,j,h,v} \cdot \bar{a} \right)^2 \leq I_{Dark} \, Q \qquad\qquad (\forall v \mid 1 \leq v \leq v_{Max}).$$

Here the index r refers to sample points (x_r, y_r) along the edges of the target patterns. $\nabla_\perp \bar{c}$ represents the derivative of c in a direction normal to the feature edge. Maximization of Ψ ensures that the shallowest log-slope among feature edges is as steep as possible. If desired a weighting factor can be applied to the log-slope at each position. The indices u and v run over sample points that must be bright and dark, respectively. Constraints are imposed to i) require achievement of minimum acceptable pupil fill, ii) enforce geometric restrictions on the size of the s_j source regions, iii) prevent line shortening and other CD errors in the printed pattern, iv) require adequate exposure in bright areas, and v) prevent excessive exposure in dark areas.

Techniques are reported in the literature for solving fractional optimization problems like eq.[4], often reducing them to a parametric problem in the difference between numerator and denominator[14]. Eq.[4] can also be approximated as a cubic polynomial optimization; a global optimum is then guaranteed in principal if a homotopy algorithm is used to solve the Lagrangian. However, we have found that problems of the size considered here pose considerable difficulty for the homotopy algorithms reported in the literature[15].

Our solution scheme for eq.[4] exploits global solutions we have found for two simplified sub-problems in the equation. This decomposition method constitutes step 1 of our overall algorithm to optimize exposure latitude (denoted P), which is outlined in the following table:

0) Preliminary:

    a. Problem definition; input of image sample points.

    b. Identify the J_{Max} source variables via Figure 13 construction.

1) Global Optimum:

    a. Considering each source variable one at a time, calculate a solution for $\bar{a}_j$ that is globally optimum under simplified criteria.

    b. Initialize amplitude variables $\bar{a}$ to the best value obtained in previous step. Initialize S_{Min} to 0.

    c. Calculate the globally optimum source distribution $\bar{s}$ for the current values of $\bar{a}$ and S_{Min}.

    d. Use a local algorithm to optimize $\bar{s}$ and $\bar{a}$ together (per eq.[4]).

    e. Increase S_{Min} by small increment and return to step c, until stopping criteria are met.

2) Fix S_{Min} at desired final level and choose corresponding solution from step 1, then refine using local optimizer against more complex criteria.

3) Calculate the optimum reticle pattern that provides the wavefront $\bar{a}$ obtained in step 2:

    a. Find global solution that produces wavefront with maximum intensity.

    b. Refine step 3.a solution using local optimizer to e.g. satisfy mask CD tolerances, reduce shapes to Manhattan geometries, etc.

(P)

Let us now consider these steps in more detail. In calculating the step 1.a amplitude sets $\bar{a}_j$, we defer constraints on equal feature bias and minimum pupil fill to step 1.c. As a further simplification, we optimize in step 1.a against the finite difference between adjacent dark and bright points across feature edges. Moreover, the overall intensity scaling of the amplitudes $\bar{a}$ is arbitrary until the step 3 mask calculation. This allows us to set the intensity at active bright point constraints to 1 (the other bright points then being above 1), eliminating the need to optimize against log-slope per se until step 1.c (since slope and log-slope are equalized at unit intensity).

The step 1.a optimization problem for the jth source direction is then to minimize intensity in dark points under these constraints, and we can write this problem in matrix form as:

$$\text{Minimize} \quad \Phi_j(\bar{a}) = \bar{a}^T A_0 \bar{a}$$

subject to:

[5]

$$\bar{a}^T A_u \bar{a} \geq 1 \qquad\qquad (\forall u \mid 1 \leq u \leq u_{Max}).$$

The symmetric A_0, A_u matrices (obtained from eq.[3]) take into account any orders that may be collected from negative directions, as well as the effect of mirroring the illumination. The $\bar{a}^T A_u \bar{a}$ quadratic forms in the constraints of eq.[5] represent the intensity at bright sample points, while the $\bar{a}^T A_0 \bar{a}$ term in the demerit function provides the average intensity within dark

Figure 15 – Joint mean-intensity eigenvectors for bright and dark regions of Figure 14 isolation pattern, with illumination incident from source region 8 (see Figure 13; illumination is fourfold mirrored). As in Figure 13, the imaging conditions are λ=248nm, NA=0.68. 00) Perspective view of target pattern (central region of Figure 14). 0) Magnified view of target pattern (the dashed upper right quadrant of previous view). 1-7) The 7 eigenvectors, plotted as images over upper right quadrant. Sorted in decreasing order of bright region intensity. All eigenvectors provide unit average intensity at dark sample points.
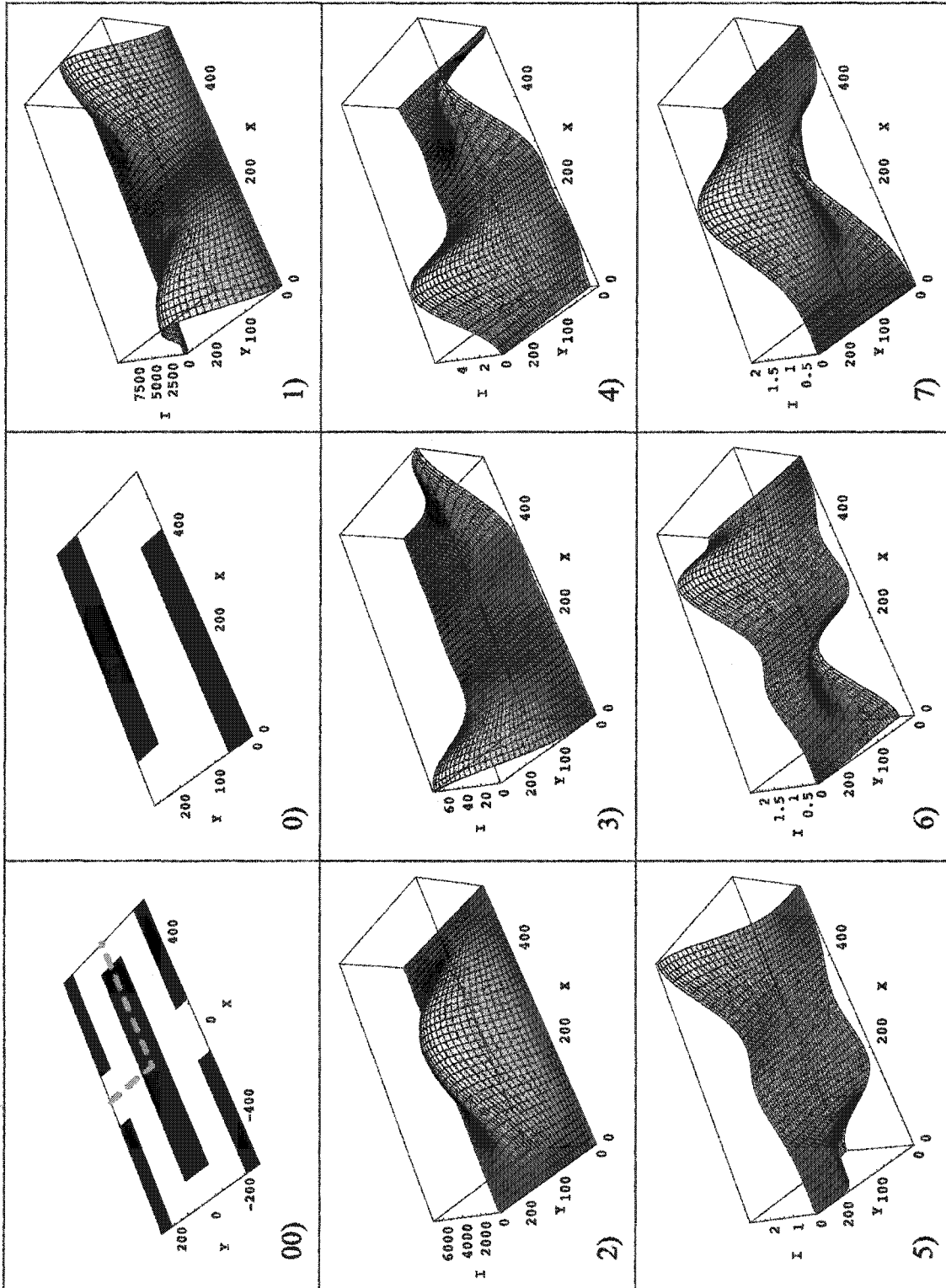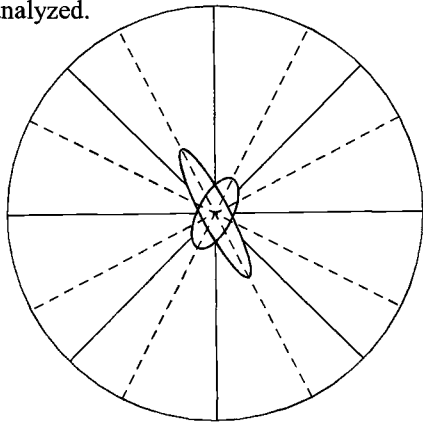
Figure 16 – Schematic of search space decomposition, for a pattern having 2 sample points in bright region. Example in text yields 3 significant eigenvectors, but for ease of drawing this figure assumes 2 (yielding a 2D subspace). In 2D the "spherical triangles" become arcs (bounded by dashed lines) whose midpoints are shown solid. Note that by symmetry only half the triangles need be analyzed.

areas of the image. This average is typically a very small quantity since we are optimizing exposure latitude in focus. Proper polarity in all dark points is thus ensured, since conversion of even a single dark point to bright would drastically raise the average, i.e. Φ could not be minimal in such a case. (Note that we are free to suppose that only a limited number of dark points participate in this average, since points are not mutually constraining if their separation greatly exceeds the lens resolution.) On the other hand, it is necessary that each bright point be entered as a separate constraint, since it is sometimes possible to make an invalid improvement in the average bright-to-dark contrast by switching a few critical bright points to dark.

Though the matrices in the eq.[5] quadratic forms (ellipsoids) can be made positive definite, the problem is nonconvex because the inequality constraints are lower bounds. However, two aspects of the eq.[5] structure allow the multiple local minima to be fully mapped in a very efficient way. First, the eq.[5] ellipsoids share a common center, and second, their principal axes (whose lengths are the reciprocal square-roots of the matrix eigenvalues) must range between very small and very large amplitudes (since for feasible values of λ/NA it must be possible to print a wide range of intensities on at least a subset of the sample points).

To exploit these properties we first calculate the eigenvectors and eigenvalues of the black region matrix A_0. (It may be necessary to shift the sample points in x,y,z by e.g. 0.001λ to make the matrix positive definite.) We then scale the eigenvectors by the square root of the reciprocal of the eigenvalues, thereby effectively scaling the diagonalized black region matrix to the identity matrix. The eigenvector basis can now be rotated into alignment with the eigenvectors of the matrix for mean intensity at bright points (average of the A_u, denoted $A_{\overline{U}}$). If we use the symbol E to denote eigenvector column matrices (i.e. the eigenvectors of a matrix A_Q are denoted E_Q) then the transformation W from the new basis to the old is given by:

$$W = E_0 D_0^{-\frac{1}{2}} E_B, \quad \text{where} \quad A_B \equiv D_0^{-\frac{1}{2}} E_0^T A_{\overline{U}} E_0 D_0^{-\frac{1}{2}}, \qquad [6]$$

with the reciprocal square root of D_0 denoting a diagonal matrix formed from the reciprocal square roots of the eigenvalues of A_0. In basis W the summed squared amplitudes give the mean black intensity, and also the mean bright intensity when weighted by the eigenvalues of A_B.

It is only possible to simultaneously diagonalize two matrices in this way (see treatment in ref.[16]), and no single eigenvector for the mean bright and dark region intensities is likely to provide high brightness at all bright sample points. Since the eigenvectors are only common to the mean dark and bright intensities, we cannot immediately calculate the relative eigenvector weightings that are required to provide an optimum image from the fixed source (e.g. region j, four-fold mirrored). However,

Figure 17 – Mask and source solution for Figure 14 isolation pattern using algorithm P (with step 2 omitted). [Results from a more sophisticated algorithm than P are shown in Figure 21.] a) Chromeless [non-alternating] mask ($T_{Min} = -1$ [shown black], $T_{Max} = +1$ [shown white]). Plotted region matches Figure 14. The mask features have a very different shape from the target patterns. b) Binary source. Circle represents 0.68 NA. Illumination directions are shown dark.
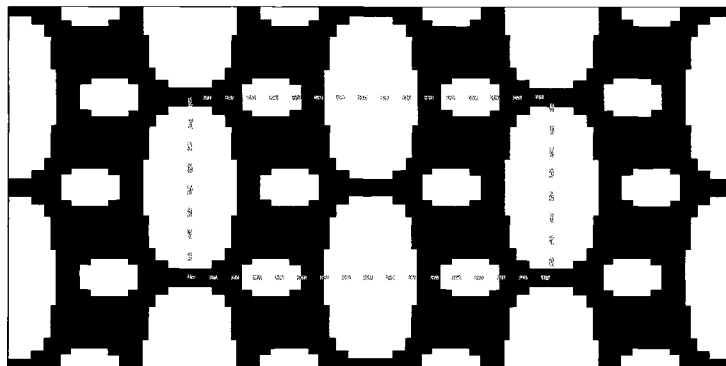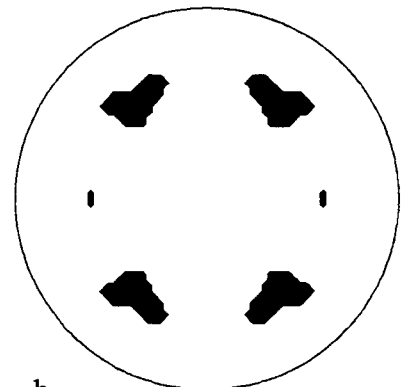

a.   b.

Figure 18 – a) Focused aerial image from the Figure 17 solution (same perspective as Figure 15.0). Thick curve shows contour slice at nominal threshold. (Only the contour for the front rectangle of Figure 15.0 is visible.) b) Horizontal (dashed) and vertical (solid) centerline slices through rectangle image. The vertical slice is shifted by the difference between the nominal length and width to show that the aerial image contour prints without line shortening. c) Process window (thresholded aerial image model, assuming no aberrations). Exposure latitude is 55%, but DOF is small (less than ±0.4μm), reducing process window to 24.7%-μm (compare with Figure 23).
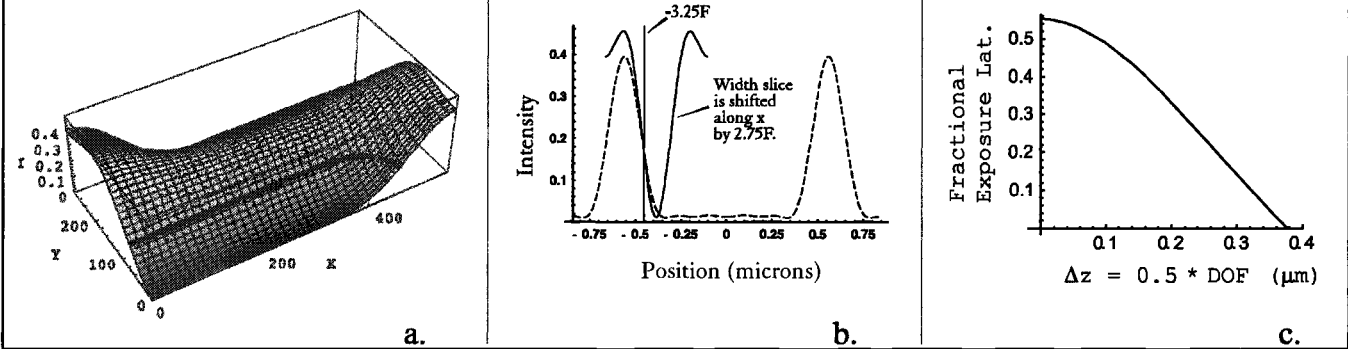
the solution vector must lie approximately within a subspace spanned by a limited number of these eigenvectors, namely the minimal set of eigenvectors such that for each of the bright sample points, at least one eigenvector in the set provides intensity above 1.

Consider, for example, the eigenvectors shown in Figure 15, which correspond to illumination from region 8 in Figure 13 (four-fold mirrored). The first 2 eigenvectors provide very high contrast, but do not allow the horizontal separations between the rectangles to be printed bright. Eigenvector 3 must also be employed in order to provide high intensity at all bright sample points, indicating that black region contrast is significantly impacted by the need to achieve high intensity between the rectangle tips. (Printing the isolation rectangles is thus more difficult than printing non-terminating lines and spaces.) Eigenvectors 4 through 7 degrade contrast in the image, and so can only contribute to the solution in small amounts.

To solve eq.[5] we now need to find the point in basis **W** which is closest to the origin while remaining outside each of the individual ellipsoids representing unit intensity at particular bright points. We can consider the search to take place within the subspace spanned by the dominant eigenvectors for mean intensity (e.g. in the Figure 14 example, the 3D subspace spanned by eigenvectors 1,2,3 of Figure 15). In order to fully probe the "nooks and crannies" of the intersecting ellipsoids in an efficient way, we organize the search space by erecting spherical triangles on the "celestial sphere" (i.e. a sphere where the intensities at all bright points are much higher than unity). The first set of vertex nodes for these bounding spherical triangles is defined by projecting the eigenvectors for individual bright points to the celestial sphere, i.e. by projecting vectors outward along the principal axes of the ellipsoids. (Of course, the algorithm must in general handle problems of arbitrary dimensionality. The number of vertices in each "triangle" is equal to the dimensionality of the subspace, and the "sphere" is a surface of dimensionality one less.) The other half of the node set is then generated by splitting the triangles through the addition of a new vertex at the central coordinate of each. We then proceed from each node by decreasing all amplitudes in a common proportion until we reach the outermost ellipsoid intersecting the ray. A local optimizer then settles into the nearest local minimum in the solution space (the innermost pocket of the intersecting ellipsoids in that region). Our local optimizer uses the augmented Lagrangian algorithm in Bertsekas' textbook[9]. To exactly solve eq.[5] during step 1.a, the local optimization should take place in the full vector space **W**. This decomposition is illustrated in Figure 16.

We should note that the method of eqs. 5 and 6 allows the globally optimum mask to be determined for arbitrary fixed source under the simplified constraints of eq.[5].

Once the step 1.a sub-problem is solved, algorithm P uses the solution to initialize $\vec{a}$, and proceeds to the source optimization loop in step 1.c. Step 1.c requires that we solve eq.[4] for $\vec{s}$, with $\vec{a}$ given. Even when $\vec{a}$ is fixed, eq.[4] is nonlinear, since the merit function involves log-slope. However, we can transform eq.[4] to the linear program:

Minimize z_0,

subject to:

$$z_0 + \vec{z} \cdot \sum_{q=1}^{4} \sum_{h=1}^{2} \left(\vec{c}_{q,j,h,r} \cdot \vec{a} \right) \left(\nabla_\perp \vec{c}_{q,j,h,r} \cdot \vec{a} \right) \geq 0 \qquad (\forall r \mid 1 \leq r \leq r_{Max}), \qquad [7]$$

$$0 \leq S_{Min} z_j \leq S_{Max,j} \sum_{k=1}^{J_{Max}} z_k \qquad (\forall j \mid 1 \leq j \leq J_{Max}),$$

$$\vec{z} \cdot \sum_{q=1}^{4} \sum_{h=1}^{2} \left(\vec{c}_{q,j,h,r} \cdot \vec{a}\right)^2 \ = \ 1 \qquad\qquad (\forall r \,|\, 1 \le r \le r_{Max}),$$

$$\vec{z} \cdot \sum_{q=1}^{4} \sum_{h=1}^{2} \left(\vec{c}_{q,j,h,u} \cdot \vec{a}\right)^2 \ \ge \ I_{Bright} \qquad\qquad (\forall u \,|\, 1 \le u \le u_{Max}), \qquad\qquad \text{[7 cont.]}$$

$$\vec{z} \cdot \sum_{q=1}^{4} \sum_{h=1}^{2} \left(\vec{c}_{q,j,h,v} \cdot \vec{a}\right)^2 \ \le \ I_{Dark} \qquad\qquad (\forall v \,|\, 1 \le v \le v_{Max})$$

Eq.[7] is linear in the transformed set of $1+J_{Max}$ variables $z_0, z_1, z_2, z_3, \cdots \equiv z_0, \vec{z}$, and so can be solved globally using standard linear programming algorithms. After eq.[7] is solved, the step 1.c source intensities are given by:

$$\vec{s} \ = \ \frac{S_{Min}\vec{z}}{\displaystyle\sum_{k=1}^{J_{Max}} z_k}. \qquad\qquad [8]$$

In general, eqs. 7 and 8 provide the global solution to the problem of optimizing the source for a given mask, under the criteria of eq.[4].

To complete our discussion of algorithm P we now describe the step 3 reticle calculation. (As noted above, it is straightforward to carry out the various local optimization steps in P using standard routines.) To begin with, we calculate the set of reticle patterns that provide the brightest possible image consistent with the step 2 solution for $\vec{a}$. This initial layout must be then refined using standard criteria; for example, the optimized patterns must be rendered on the mask as polygons, preferably as a set of rectangles. The rectangles can be fairly coarse, e.g. of dimension only moderately smaller than the lens resolution. We use a local optimizer to do this refinement.

For the basic reticle calculation we approximate the Fourier diffraction integral as a summation over discrete sample points. The mask transmission function T(x,y) is sampled on a 2D grid, and then unraveled into a 1D vector of unknowns $\vec{T}$ indexed by g:

$$\int_{-P_x/2}^{P_x/2} \int_{-P_y/2}^{P_y/2} dx\,dy\, T(x,y)\, e^{2\pi i\left(\frac{mx}{P_x}+\frac{ny}{P_y}\right)} \ \cong \ \sum_{k=1}^{K} \sum_{\ell=1}^{L} T(x_k, y_\ell)\, e^{2\pi i\left(\frac{mx_k}{P_x}+\frac{ny_\ell}{P_y}\right)} \ \equiv \ \sum_{g=1}^{KL} T_g\, b'_{g,m,n} \ \equiv \ \sum_{g=1}^{KL} T_g\, b'_{g,w}. \qquad [9]$$

The symbol b' has been introduced in eq.[9] as shorthand for the exponential, and an unraveled index w is introduced to represent the m,n indices of the wth captured amplitude in $\vec{a}$.

Step 3 now becomes a linear programming problem:

$$\text{Maximize } \Omega\!\left(\vec{T}\right) \ \equiv \ \text{Sign}\!\left[\sum_{w=1}^{W_{Max}} a_w\right] \sum_{g=1}^{KL} \sum_{w=1}^{W_{Max}} T_g\, b'_{g,w},$$

subject to:

$$\sum_{g=1}^{KL} T_g \left[\left(a_{w'} \sum_{w=1}^{W_{Max}} b'_{g,w} \right) - \left(b'_{g,w'} \sum_{w=1}^{W_{Max}} a_w \right) \right] \ = \ 0 \qquad\qquad (\forall w' \,|\, 1 \le w' \le W_{Max}), \qquad [10]$$

$$T_{Min} \ \le \ T_g \ \le \ T_{Max}.$$

Eq.[10] forces the mask Fourier orders to be in the same ratio as the elements of the optimized diffraction order list $\vec{a}$ obtained in step 2. T_{Min} and T_{Max} are determined by the mask technology. T_{Max} would generally be +1, while T_{Min} would be e.g. −1 for a chromeless mask, $-\sqrt{0.065}$ for an attenuating phase-shift mask with 6.5% chrome transmission, etc. In general we must set $T_{Min} < 0$ for eq.[10] to provide a solution.

Eq.[10] can be modified to adjust the exposure threshold of the printed pattern (e.g. to match its intensity with that provided by some other set of mask patterns) by adding the constraint:

$$\text{Sign}\!\left[\sum_{w=1}^{W_{Max}} a_w\right] \sum_{g=1}^{KL} \sum_{w=1}^{W_{Max}} T_g\, b'_{g,w} \ = \ \Omega_{Match}. \qquad\qquad [11]$$

This adjusts the intensity of the aerial image without changing its shape. Ω_{Match} must of course be smaller than the unmodified
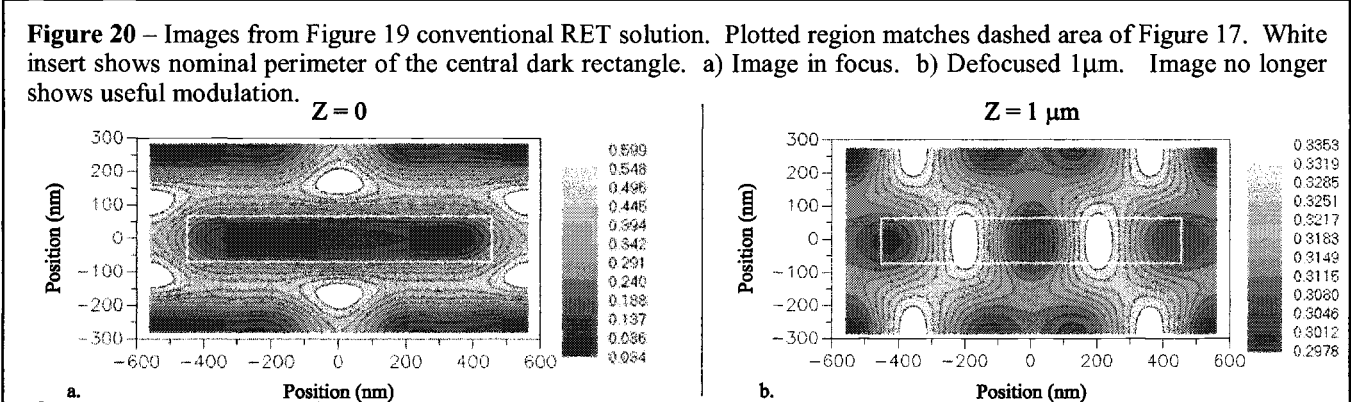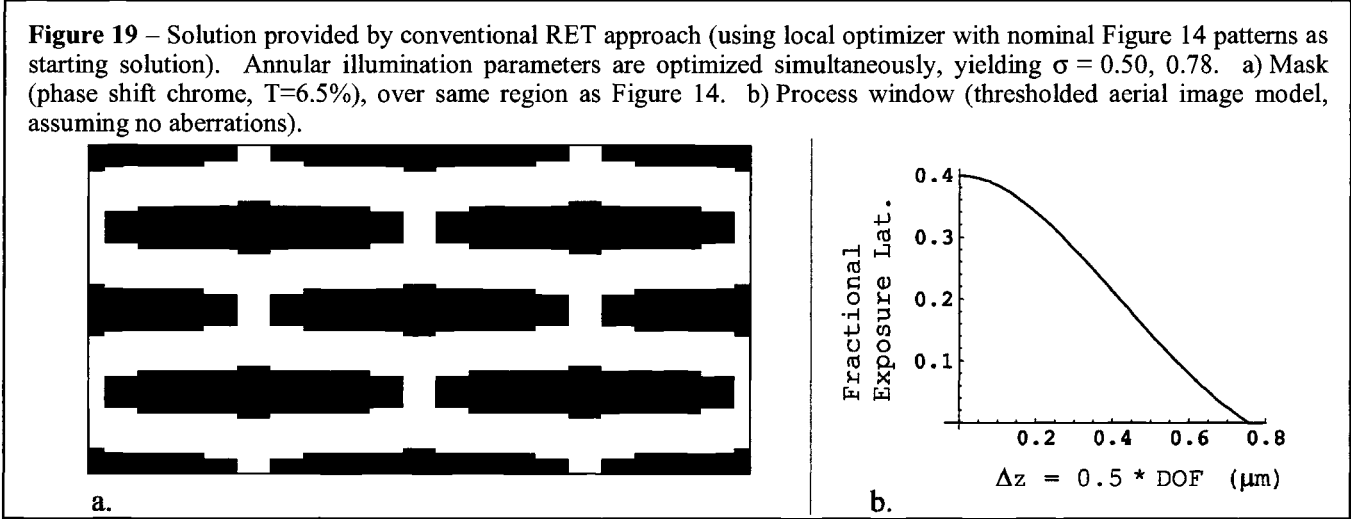
eq.[10] maximum. To prevent excessively fine features in the retuned solution, one can introduce a spatially smoothed version of the unmodified solution as a new objective vector. This gives preference to pixel adjustments near the edges of features, where the absolute value of the smoothed pattern is less (so that correlation with the new objective vector is maximized when adjustments are made at the edges of existing features, rather than in newly introduced features.) Alternatively such criteria can be enforced in the step 3.b local optimization.

In the limit of an arbitrarily fine grid, the solution provided by eqs.[10] and [11] will be "two-tone", in that all pixels will be driven to either T_{Min} or T_{Max}. To design a Levenson-type mask (i.e. a mask with $0°$ and $180°$ apertures opened in opaque chrome), we modify eq.[10] with a change of variables and added constraints:

$$T_g \rightarrow T_g^+ - T_g^- ,$$

$$T_g^+ \geq 0, \quad T_g^- \geq 0,$$

$$\sum_{g=1}^{KL} \left(T_g^+ + T_g^- \right) \leq (1-G) \cdot KL$$

[12]

If parameter G were allowed to float, the change of variables in eq.[12] would not revise the solution of eq.[10] (assuming $T_{Min} = -1$, $T_{Max} = +1$), since the first two lines of eq.[12] permit a transmission of ± 1 to be realized whenever the third line is not binding. This latter constraint is activated by setting G to a positive value; a fraction G of the reticle area is then driven to opaque chrome (i.e. $T_g^+ = T_g^- = 0$).

Figure 17 shows the solution provided by algorithm P for the isolation pattern of Figure 14, in the simple case where the step 2 local optimization is omitted. Log-slope across the narrow width of the rectangles is given a 1.5X higher weighting than log-slope at the tips of the rectangles, corresponding to a tighter CD on the width than the length (tighter in absolute terms; relative tolerances are the same). Figure 18 shows the aerial image in focus. The intensity along the centerline of the dark rectangles is roughly 1/30th that at peak. When spacewidth tolerances of $\pm 20\%$ are applied to the bright horizontal and vertical separations between the rectangles, the exposure latitude is 55%. This is about a 1.4X improvement over the 40% exposure latitude achieved by a more conventional approach (described below).

Figure 19 – Solution provided by conventional RET approach (using local optimizer with nominal Figure 14 patterns as starting solution). Annular illumination parameters are optimized simultaneously, yielding $\sigma = 0.50$, 0.78. a) Mask (phase shift chrome, T=6.5%), over same region as Figure 14. b) Process window (thresholded aerial image model, assuming no aberrations).

$\Delta z = 0.5 * DOF$ (μm)

a.

b.

Figure 20 – Images from Figure 19 conventional RET solution. Plotted region matches dashed area of Figure 17. White insert shows nominal perimeter of the central dark rectangle. a) Image in focus. b) Defocused 1μm. Image no longer shows useful modulation.

a.

b.

OPTIMIZATION OF PROCESS WINDOW VS EXPOSURE LATITUDE

Unfortunately, the depth of focus provided by the Figure 17 solution is not very large ($\pm 0.38 \mu m$ under the above $\pm 20\%$ CD tolerance), leading to an integrated process window of only 24.7%-μm (using a thresholded aerial image model), despite the large exposure latitude in focus. This process window is considerably better than can be achieved with a simple opaque chrome mask incorporating the nominal patterns without bias. However, standard OPC methods can do appreciably better. Figure 19 shows the result of using a local optimizer to adjust the shapes of mask openings in phase-shift chrome, with the nominal Figure 14 pattern serving as a starting solution. The inner and outer radii of annular illumination were adjusted simultaneously. Depth of focus is $\pm 0.75 \mu m$, substantially exceeding that of the Figure 17 solution, and a better process window overall is achieved (33.3%-μm). Figure 20 shows plots of the aerial image.

We should emphasize that this decoupling of process window and exposure latitude does not always arise. For example, in optimizing mask and source to print the Figure 2 pattern, though the Figures 3,4a solution was obtained using an algorithm that optimizes against full process window, a very similar solution is provided by algorithm P. (Process window with algorithm P is 37.6%-μm, versus 45%-μm for the solution of Figures 3 and 4a.) Indeed, the Figure 3,4a solution can probably be recovered exactly from algorithm P if process window is used as the merit function in the step 2 local optimization.

The Figure 14 problem can be attacked in a similar way; i.e. by refining the step 1 solution (Figure 17) against process window using a local optimizer (step 2 of algorithm P). The solution found in this way yields a process window of 36.2%-μm, slightly exceeding that of the more conventional Figure 19 approach. The step 2 refinement is found to improve depth of focus by 50% while decreasing exposure latitude by only 2%-μm, demonstrating again that process window and exposure latitude are not always strongly coupled. Clearly, it is preferable to have a global algorithm that can directly optimize the solution for maximum process window.

We have made considerable progress toward development of such an algorithm. Integrated area under the ED window is maximized (under a thresholded aerial image model). Figure 21 shows the solution obtained by this method for the Figure 14 isolation pattern; Figures 22 and 23 show the resulting image and process window. (The solution of Figures 3 and 4a was also obtained with this algorithm.) Integrated process window is 67%-μm (see Figure 23), about double that obtained with the more conventional RET optimization of Figure 19 (and also about double that obtained by optimizing for process window in step 2 of algorithm P). The improvement in depth of focus may be seen by comparing Figures 20 and 22. Figure 24 emphasizes the dramatic difference between the optimized mask shapes of Figure 21 and the printed pattern.

Figure 21 – Globally optimized solution to maximize process window for Figure 14 pattern. (See also Figure 17 solution, which only optimizes exposure latitude.) a) Chromeless mask (non-alternating). Black represents $0°$ phase-shift, white $180°$. Plotted region matches that in Figs. 14, 17, and 25. b) Jointly optimized gray-scale source.
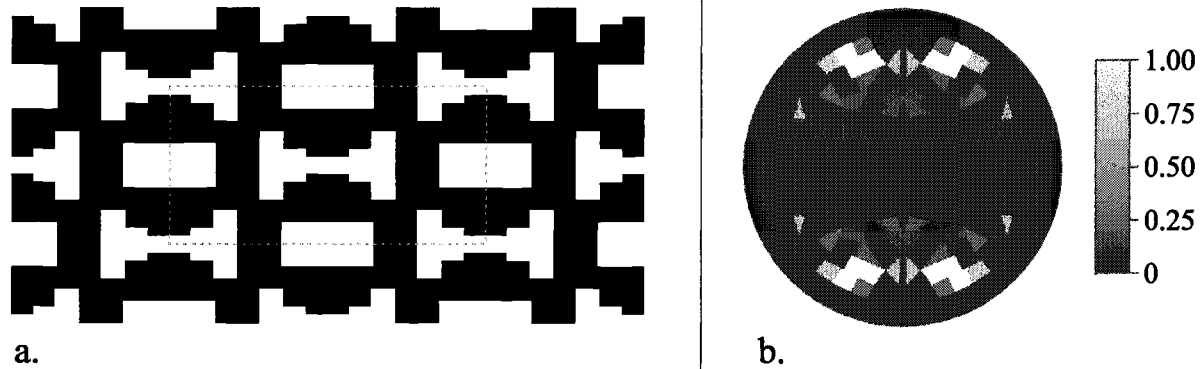

Figure 22 – Aerial images for the Figure 21 solution (screen captures from Prolith[5] simulations). Plotted region matches dashed area of Figure 21 (also matches Figure 20). White insert shows nominal perimeter of the central dark rectangle. a) Image in focus. b) Defocused 1μm. DOF is considerably larger than with conventional enhancement approach.
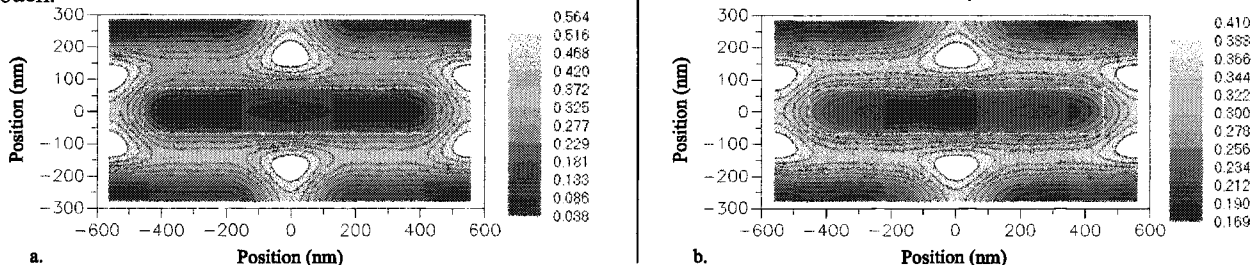
Figure 23 – Process window plot for Figure 21 solution, with ±20% CD tolerances on the bright horizontal and vertical separations between rectangles. A thresholded aerial image model is used, and an aberration-free lens assumed. The integrated window (two-sided) is 67%-μm.

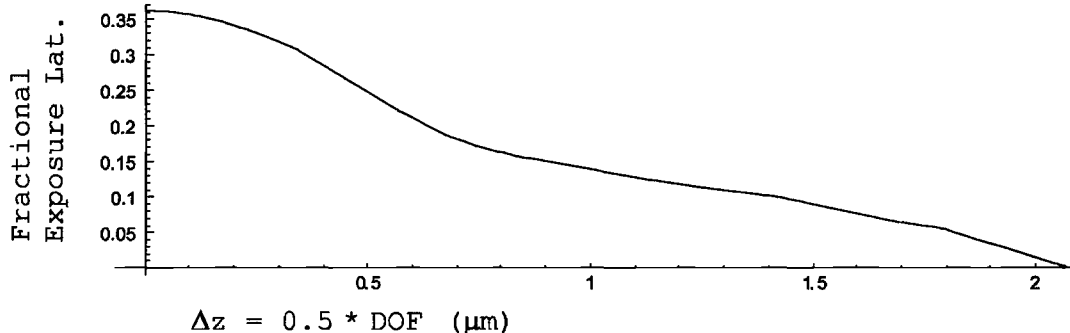$$\Delta z = 0.5 * DOF \quad (\mu m)$$

Figure 25 shows an implementation in opaque chrome (i.e. a Levenson mask where features have unit transmittance and 0° or 180° phase shift). In general, eq.[10] and related methods provide highest efficiency in chromeless technology, and Figures 5 and 22 demonstrate that reasonably high intensities can be achieved. We have found these methods to be quite successful in compensating the greater difficulty in maximizing intensity when a decentered wavefront slice is optimized. Of course, exposure time will be significantly degraded if the optimized source is provided by an attenuating aperture rather than diffractive elements (as in exposure tools that provide software-selectable source distributions via a library of pre-loaded diffractive elements[17]; see also ref.[8]).

CONCLUSIONS AND FUTURE DIRECTIONS

Figure 24 – Superposition of Figures 21a and 22a. The dark image rectangles are centered on the "bow-tie" shapes. The centers of the rectangular mask features print bright. Plotted area corresponds to dashed regions of Figures 14, 17, 21, and 25.
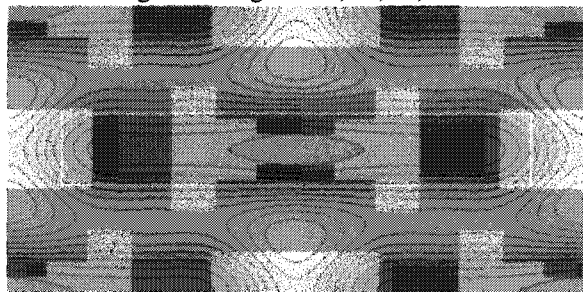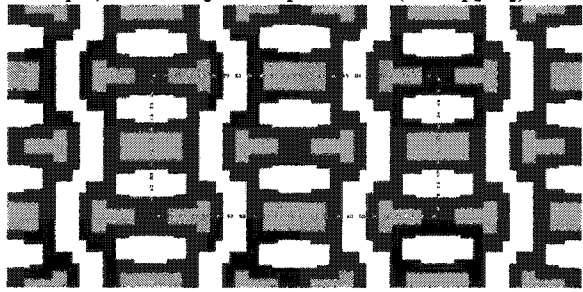

Figure 25 – Implementation of Figure 21 solution as Levenson mask. Opaque chrome is shown black; white and gray represent openings of 0° and 180° phase shift. (Mask is not alternating.) Plotted region is the same as Figure 14. Chrome coverage (low in this example) can be adjusted up or down (see eq.[12]).

To maximize process window it is necessary that the reticle shapes not be constrained to follow the inherent "topology" of an initial design form. By considering the implications of off-axis illumination in a detailed way, we have devised a design algorithm that is not encumbered by such restrictions. The theoretical improvement in performance from this global approach can be substantial.

However, many practical issues remain to be considered. The present paper focuses on development of the basic algorithm, but it is important that the solutions be compatible at a detailed level with practical constraints imposed by the illuminator and the mask-making process. For example, it is possible that the illumination will need to satisfy tighter requirements on directional uniformity when pattern symmetry is provided by the source rather than the collected wavefront.

Global optimization must also be integrated into an overall strategy to print a given IC level. The field sizes considered above are sufficient for e.g. separate exposure of the array region of a DRAM level, but for general purposes this is not adequate. Several approaches are available to accommodate larger sets of patterns. While globally optimized designs are often somewhat novel and unexpected, one can generally understand them "after the fact" in an intuitive way that is more compatible with a lithographer's "bag of tricks" than is possible for a purely mathematical result. Our discussion of global algorithms has been couched in terms of optimizing mask and source together; however, once the source has been optimized for critical patterns, it is possible to globally optimize less critical mask patterns with the source distribution held fixed (e.g., see eqs. 5, 6). The source can also be "softened" to improve compatibility with a wider range of shapes[18].

Though the algorithm can be extended by such techniques, computational limitations make it necessary to interface the globally

optimized solutions with neighboring patterns that are derived by other means. Periodic boundary conditions entail additional computational burden when target patterns are non-periodic, e.g. to feather overlapping solutions across redundant buffer regions. Eq.[11] allows the exposure threshold in a given aerial image to be adjusted up or down to maximize the common window with other patterns.

Though computational requirements make these hybrid approaches inevitable over full IC levels, it is interesting to speculate on how the benefits from fully global optimization might scale if no compromises were made, i.e. to assess the potential advantages of global optimization as the dimensional scale and pattern diversity of the simultaneously optimized feature set is increased.

Off-axis illumination continues to provide access to more degrees of freedom when a pattern is optimized as a member of a group rather than individually, and in principle these degrees of freedom are best optimized with a global algorithm. With conventional enhancement methods the common process window for a group of features is generally less than that of the features considered individually. Global optimization may prove a useful tool to bring to bear on this problem. On the other hand, the relative advantage of global optimization over conventional methods might decrease when a suite of patterns is optimized, since conventional methods already employ broader and more symmetric sources than are required for individual patterns. The Figure 13 construction implies that large-σ illumination directions along the 45° azimuths provide the largest number of independent collected orders when patterns are highly symmetric, potentially improving the prospects for optimizing a broad set of patterns. Global optimization can theoretically allow the less critical patterns to be printed with a narrower and more discrete source than usual (i.e. a source optimized for critical patterns), but this may entail optimization of a great many shapes. While a fully global algorithm cannot in principle do worse than local optimization, it imposes a distinctly greater computational burden, which may force significant compromises. It remains to be established how these factors will trade-off when optimizing the pattern content of different IC levels.

ACKNOWLEDGEMENTS

REFERENCES

[1] J. Gortych and A.E. Rosenbluth, "Method and system for optimizing illumination in an optical photolithography projection imaging system," US Patent 5,680,588 (1997).

[2] A.E. Rosenbluth, T. Brunner, and D.G. Flagello, "Optical lithography as the NA=1 limit approaches," 1992 Optical Society of America Annual Meeting (Albuquerque, NM, 1992), paper MVV-3.

[3] M. Burkhardt et al., "Illuminator design for the printing of regular contact patterns," Microelectronic Engineering 41 (1998): p. 91.

[4] T.-S. Gau et al., "The Customized Illumination Aperture Filter for Low k1 Photolithography Process," in SPIE v.4000 - Optical Microlithography XIII, ed. C.P. Progler (Santa Clara, CA: SPIE, 2000), p. 271.

[5] See product description for Prolith at Finle website, http://www.finle.com.

[6] S. Bukofsky et al., "A Lithographically-Friendly $6F^2$ DRAM Cell," Proc. VLSI-TSA 2001, to be published.

[7] A. Wong et al., "Level-specific lithography optimization for 1-Gb DRAM," IEEE Transactions on Semiconductor Manufacturing 13, no.1 (Feb. 2000): p. 76.

[8] M.D. Himel, R. Hutchins, and M. Poutous, "Design and fabrication of customized illumination patterns for low k1 lithography: a diffractive approach," SPIE v.4346 - Optical Microlithography XIV, ed. C.P. Progler (Santa Clara, CA: 2001), to be published.

[9] D.P. Bertsekas, Nonlinear Programming (Belmont, Mass.: Athena Scientific, 1995).

[10] Matlab Optimization Toolbox, http://www.mathworks.com/products/optimization/8513v03_optim.pdf.

[11] S. Wolfram, The Mathematica Book, 4th ed. (Cambridge University Press, 1999).

[12] IMSL Online Documentation (Visual Numerics Inc.), http://www.vni.com/products/imsl/index.html.

[13] S.A. Vavasis, "Complexity Issues," in Handbook of Global Optimization, ed. Reiner Horst and Panos M. Pardalos (Dordrecht, Netherlands: Kluwer Academic Publishers, 1995), p. 27.

[14] A.I. Barros et al., "A new algorithm for generalized fractional programs," Mathematical Programming 72 (1996): p. 147.

[15] A. Morgan, Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems (Englewood Cliffs, NJ: Prentice-Hall, 1987).

[16] G. Hadley, Linear Algebra (Addison-Wesley, 1961).

[17] D. Williamson et al., "Micrascan III, a 0.25 micron step and scan system," SPIE v.2726 - Optical Microlithography IX, ed. G.E. Fuller (1996): p. 780 .

[18] J. Fung Chen, private communication.