Lithography and Level-Sets:
Inverse Lithography Technology (ILT) made possible

ABSTRACT

Long considered the holy-grail of photomask design, ILT has only recently become a practical commercial possibility. Not a moment too soon, as traditional approaches, such as model based OPC, are running into severe difficulties at advanced technology nodes (65nm and beyond). We discuss how one can use level-set methods to find the optimal photomask by rigorously solving the lithography inverse problem. The design of the optimal mask takes into consideration not only pattern fidelity under nominal conditions, but also the size of the process window and the constraints and costs of mask manufacturing. By formulating the problem in a rigorous mathematical framework, we find highly optimal solutions which do not arise from traditional ad hoc approaches. The resulting masks often provide substantially improved depth-of-focus and exposure latitude, enabling geometries that may be otherwise unattainable.

INTRODUCTION

Increasingly, for semiconductor manufacturers moving to advanced nodes –65nm, 45, 32, and below – the greatest challenge is lithography. This is because lithography is fundamentally constrained by basic principles of optical physics. At 65 nm, a line is less than a third of the effective wavelength; as the industry moves forward, optical diffraction and interference are becoming fundamental obstacles, not just second order effects.

It has long been known that the best lithography can be achieved in theory by considering the design of photomasks as an inverse problem - and then solving the inverse problem to find the optimal photomask, using a rigorous mathematical approach. Inverse Lithography Technology (ILT) has been explored for many years, starting with the pioneering work of B. E. A. Saleh and his students in the 1980s [1,2]. Although these earlier approaches to ILT often resulted in the design of superior photomasks, they were generally impractical in a production environment: run-times were many orders of magnitude too slow, and the resulting masks were often too complex to manufacture. Luminescent has developed the first commercial ILT solution, suitable for use in a production environment. It can rapidly solve for the optimal manufacturable photomask design, with turn-around times comparable or superior to other mask synthesis solutions such as OPC.

Stated simply, the goal of inverse lithography is to find the optimal mask pattern to print a given target pattern, given the known transformation from mask to wafer. It is a rigorous mathematical approach to the design of photomasks for optimal lithography. By its nature, ILT is a methodology that is not limited to simple heuristic modifications of the target pattern; in other words, it explores regions of solution space that are very different from the original pattern. Perhaps most importantly, it is inherently pattern independent.

There are also some common myths associated with ILT. One of these is that ILT always results in a unique global optimum. Although this may be true for some algorithms, many inverse lithography approaches are based on local search heuristics which find a good -- but not necessarily global -- optimum. Moreover, because the inverse problem is mathematically ill-posed, there are many situations in which the global optimum is not unique. It is then up to the algorithm (or algorithm designer) to determine which of many equally good solutions is desired.

Another ILT myth is that such methods must be image or pixel based. The algorithm developed at Luminescent and discussed below is pixel based, but not every approach to inverse lithography necessarily is.

Finally, there is a common perception that inverse lithography algorithms must represent a solution in closed form, or are in some sense direct solutions to the problem. In point of fact, however, most ILT methods that have been developed are iterative at some level, and it is quite true in general that engineering inverse problems are often solved with iterative methods. This should not be surprising, because many (if not most) modern mathematical solvers involve iterations, and, done properly, such methods can be very efficient.

The key distinctive feature of ILT is the lack of pattern-dependent heuristics, and the ability to broadly explore wide areas of solution space. ILT algorithms routinely lead to mask patterns which are unanticipated. Consider the problem of placement of subresolution assist features (SRAFs). These are mask features that do not print on the wafer, are detached from the edges of the semiconductor pattern, and yet manipulate the light reaching the wafer so as to accentuate the wafer image. In the past, these were placed empirically, with great care, and frozen in place during the computation of the rest of the mask. ILT, on the other hand, can produce SRAFs "out of thin air", and determine optimal SRAFs simultaneously with the rest of the mask.

The benefits of such an approach are many. The absence of segmentation scripts is a significant advantage. In addition, ILT is generally not susceptible to errors resulting from unanticipated patterns (because it is inherently pattern independent). By finding optimal mask patterns, superior pattern fidelity, larger process windows, and improved yields can all be achieved.

ILT HISTORICAL BACKGROUND

ILT was first developed by B. E. A. Saleh and others at the University of Wisconsin-Madison. For example, in 1981 Saleh and Sayegh [1] found optimized photomasks by a variation on simulated annealing, or "pixel flipping"— that is, they started with an initial guess, randomly flipped individual pixels, accepted changes that improved the quality of the solution (and rejected changes that degraded the solution), and repeated this process until the system converged on an optimal photomask. A few years later, Saleh and Nashold [2] described an algorithm using a sequence of projection operators in order to find a band limited function (corresponding to a continuous-tone or gray scale mask) which would optimally result in the desired image. Later, the same authors used a similar approach to find complex valued functions that corresponded to continuous tone phase masks.

In the early '90s, Yong Liu and Avideh Zakhor (at Berkeley) wrote a series of papers [3,4] describing various approaches to ILT. In one case, they used branch and bound and the simplex method. In another, they used what they called a "bacteria" algorithm in order to satisfy mask constraints.

In 2001, Rosenbluth et. al. (at IBM) described an ILT algorithm that analyzed diffraction orders in order to jointly optimize the photomask and the stepper illumination [5]. This approach solved first for an optimal wavefront and then in a second step tried to find the optimal photomask to generate the same diffraction pattern.

Although the researchers described above made significant contributions to the development of ILT, there are many others who have also made important contributions: for example, the work done by Wang. et. al. [6] (at Stanford, and later Numerical Technologies), and the OPERA program, by Oh et. al. (at Wonkwang University in South Korea) [7]. Most recently, Fuhner and Erdmann of the Fraunhofer Institute developed ILT using genetic algorithms [8]. The above summary is merely intended as a survey, and is certainly not one hundred percent inclusive.

As mentioned previously, these early approaches to ILT usually resulted in superb lithography. The patterns found often resulted in superior accuracy, improved process windows, and better pattern fidelity. However, they were generally impractical in a production environment, due to intractable run-times and/or unmanufacturable masks.

MATHEMATICAL METHODS

In order to formalize the problem, we require some definitions:

Mask function: ψ
Target pattern: Φ
Forward operator: f
Wafer pattern: ω

The forward operator may take into account all of the elements of the transformation from mask to wafer: for example, the electromagnetics of the 3D mask, the optics of illumination and the lens, the behavior of the photoresist, the dose and focus conditions, aberrations, etc. Thus

$$\omega = f(\psi) \tag{1}$$

and we seek to find

$$\psi^{*} = f^{-1}(\Phi) \tag{2}$$

where ψ^{*} is the optimal mask function.

However, the problem thus stated is ill-posed; because the forward operator *f* is many-to-one (that is, many different masks will yield identical on-wafer results), the function has no well-defined inverse. Moreover, for typical target patterns Φ (e.g., a drawn layout with Manhattan geometry and sharp corners), there does not exist any mask function ψ for which $\Phi = f(\psi)$.

These issues are addressed by recasting the inverse problem as an optimization problem. We define a merit function, also called a cost function, energy function, or Hamiltonian (by analogy to quantum mechanics), and label it $H(\psi)$. This function is indicative of the quality of the solution, or the "goodness" of the mask. A simple example would be

$$H = \iint \left| f(\psi) - \Phi \right| \tag{3}$$

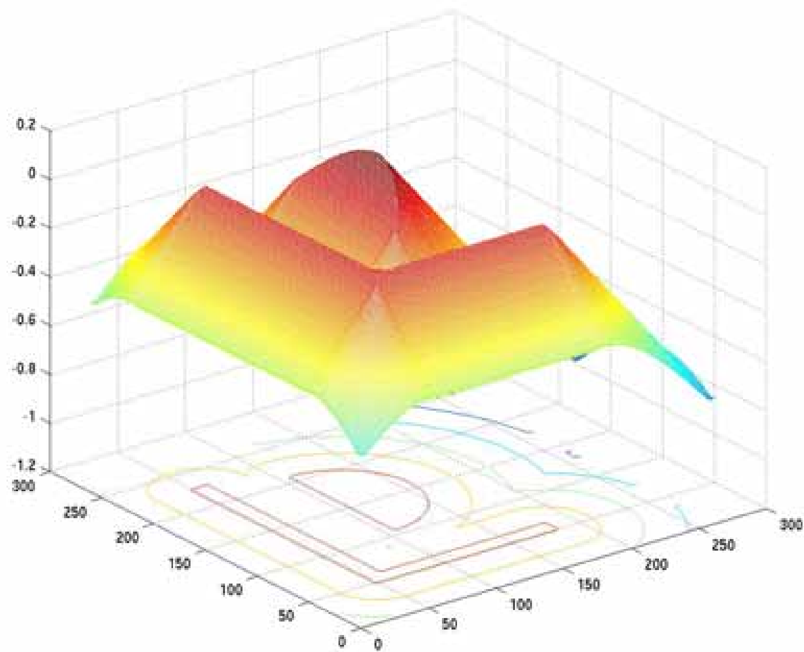
In other words, this Hamiltonian is the absolute value of the difference between the wafer image and the target pattern, integrated over the area of the mask. In practice, a number of additional elements may be included in the Hamiltonian. For example, the wafer pattern at various conditions throughout the process window (i.e., over or under exposed and/or plus/minus focus), the NILS of the image, the robustness against MEEF, or other factors as deemed appropriate. The actual functional form may be different from the form as described above as well. Elements that are not directly related to lithography may be included; for example, simple masks may be preferred over complex masks, and terms to this effect may be included in the Hamiltonian as well. What is essential is that the Hamiltonian is a functional of the mask function, and that minimizing said Hamiltonian allows us to find the optimal mask, according to the criteria we have chosen.

More specifically, because a binary photomask is defined by its contours, we define the mask function according to the following principle:

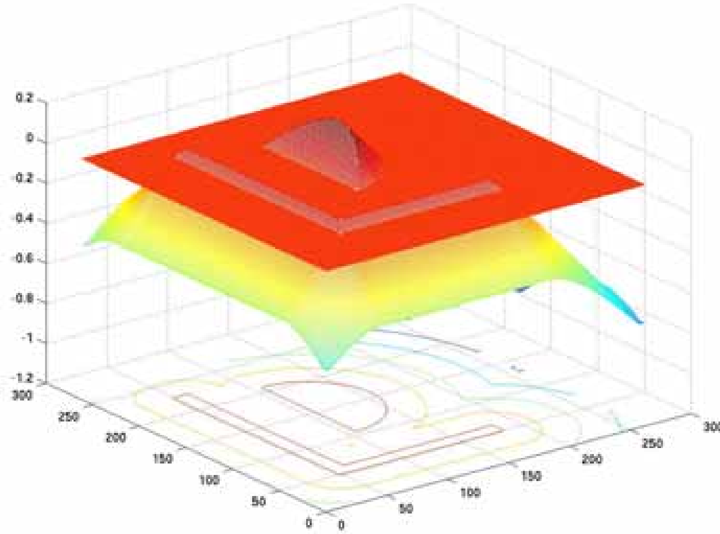
> $\psi(x, y) > 0$ "inside" a region (for example, those regions corresponding to the chrome portions of the mask);
> $\psi(x, y) < 0$, or is negative "outside" a region (for example, those regions corresponding to the clear quartz portions of the mask).

It is common to further define $\psi(x, y)$ such that it represents the distance to the contour (positive inside, negative outside), but this is not always necessary. The contours themselves are defined by the "level-set"[9], i.e. those values in the (x, y) plane such that $\psi(x, y) = 0$. As an example, consider the following simple pattern:

This could be represented by a mask function $\psi(x, y)$ as shown below:



In order to emphasize the contour-defining zero level set, we introduce a solid orange cut-plane to indicate the position of $\psi(x, y) = 0$.



From this illustration, one can see how the level-set $\psi(x, y) = 0$ defines the contours of the photomask pattern.

The above formulation of the problem has a variety of advantageous properties. For example, the level-set representation allows for contours to merge, break, appear, or disappear, in a consistent, mathematical representation. Various functions (for example, the wafer image) can be determined as closed form expressions. The mask function itself is an element of a Hilbert-space which is much larger than the two-dimensional space of the photomask, which allows for "more global" solutions to be found.

Formalizing the problem in this way implies that the Hamiltonian *H* is a functional: it maps the (infinite dimensional) Hilbert space of mask functions to a real number:

$$H : C(\Re^2) \to \Re$$

Our goal is to find the function $\psi(x, y)$ which minimizes the functional *H*. We know from the calculus of variations that the minimum occurs when

$$\frac{\delta H}{\delta \psi} = 0 \tag{4}$$

where $\frac{\delta H}{\delta \psi}$ indicates the Frechet derivative of the Hamiltonian with respect to the mask function.

The optimal mask is thus found as follows: We write down a closed form expression for *H* and calculate the Frechet derivative, also in closed form. We then solve the equation (4) above with our numerical method of choice. Because even a finite dimensional approximation to the full Hilbert space is quite large, it is most efficient to use iterative methods that converge quickly.

One important aspect of the minimization problem comes in the form of constraints. A variety of constraints are imposed by the realities of mask manufacturing; for example, two disjoint chrome regions must be separated by a minimum distance, and a chrome line must have a minimum thickness. We address these constraints by defining a sub-

space of the full Hilbert space of mask functions, and restricting our solution to this sub-space. We accomplish this restriction through the use of a projection operator applied periodically while solving the equation above.

RESULTS AND EXAMPLES

Figure 1 shows an example to demonstrate the power and flexibility of ILT [10]. The goal is to print a regular contact array with 110nm CD, 440nm pitch. The numerical aperture of the stepper is 0.78, the illumination source is a disk with a sigma of 0.3. In Figure 1 we show a continuous tone mask, an attenuated phase-shifting mask without mask constraints, and an attenuated phase-shifting mask with Manhattan constraints, all designed with ILT, to print the above described contact hole array. The process window of images produced by the Manhattan attenuated phase-shifting mask gives a DOF of 574 nm at 7% exposure latitude. The aerial image contrast is 0.85, well above 0.35, which is the rule-of-thumb minimum aerial image contrast required to print on wafer. It is quite remarkable that the combination of a single small central region, with four distinct surrounding lobes, should optimally print the contact pattern. Such remarkable patterns illustrate the power of ILT, which finds solutions that are often unexpected. Notice also the assist features found further away the contact, around the perimeter of the image.

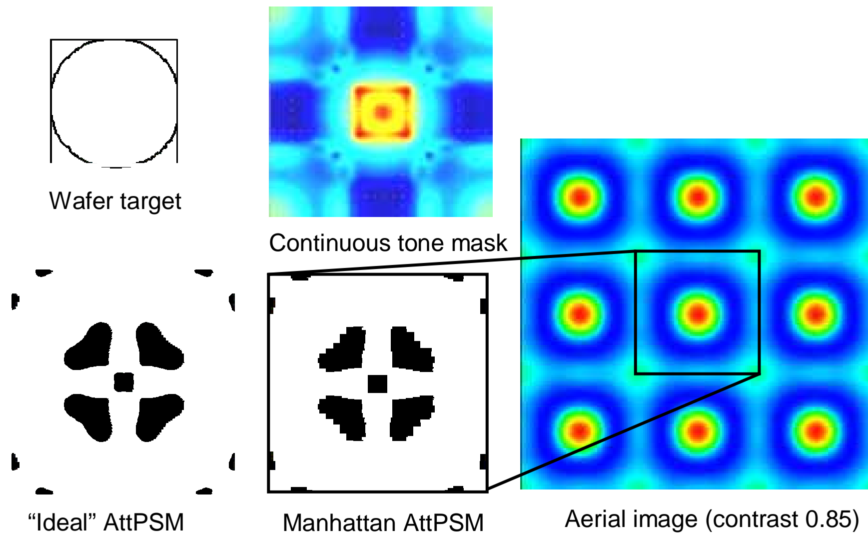


Figure 1. ILT contact array example: contact target of 110nm CD, 440nm pitch. Three types of masks were generated from ILT: continuous tone mask, "ideal" attenuated phase-shifting mask, attenuated phase-shifting mask with Manhattan constraints, and the aerial image simulated using Manhattan AttPSM.

By finding the more global optimized mask patterns, ILT brings an additional benefit of improved wafer pattern fidelity and process window. It also opens the possibility of using existing lithography equipments (e.g., scanner) into smaller geometries; in other words, extending the life of existing lithography equipments.

Using modern numerical methods and the latest processors, it is now possible to quickly solve the resulting minimization problem. Our implementation divides a large photomask into small regions called “work units”. These are distributed to a cluster of compute nodes, which can then process many work units in parallel. The solutions are then stitched together to form a complete mask. A large number of real full-chip designs have been processed this way through our system. The results can thereby be obtained quite quickly. In one recent example, a large 4 cm^2 die (wafer scale) was processed overnight.

Figure 2 shows wafer prints of clips of ILT optimized full-chip masks from world leading foundries. The result on the left (Figure 2(a)) is from a 65nm metal 1 layer[11]. In this case, both ILT and OPC used optical model. Two images for both ILT and OPC were captured at best focus, and at plus 60nm defocus. It shows wafer image using ILT mask has a better pattern fidelity through process window. From the zoomed-in pictures (insets at right), it is clear that ILT produces less line-end shortening and corner rounding issues. The result on the right (Figure 2(b)) is from another test mask and test wafer[12]. The design target is a square "donut" shape with very small dimension (CD 114nm on wafer). From the inner shape of the "donut" it is clear ILT can resolve smaller features better than OPC in this case.

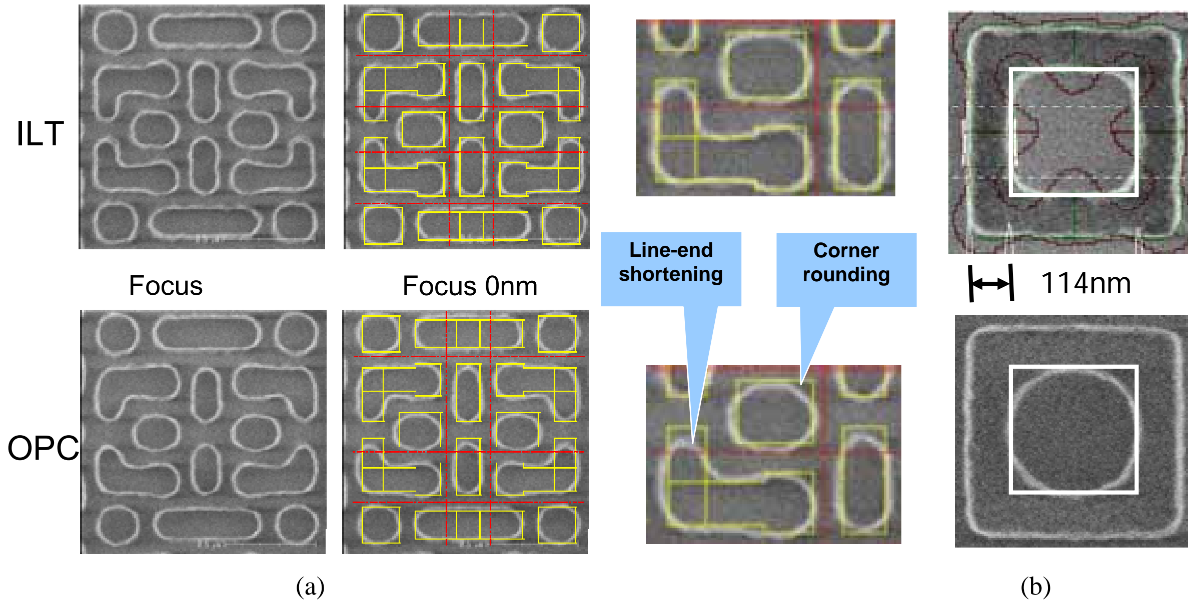


Figure 2. ILT and OPC wafer images. (a). Through focus wafer patterns of metal 1 layer, where outline is the design target. The insets at right are a magnification of a selected portion of the 0nm focus condition. (b). a square "donut" shape (CD 114nm on wafer).

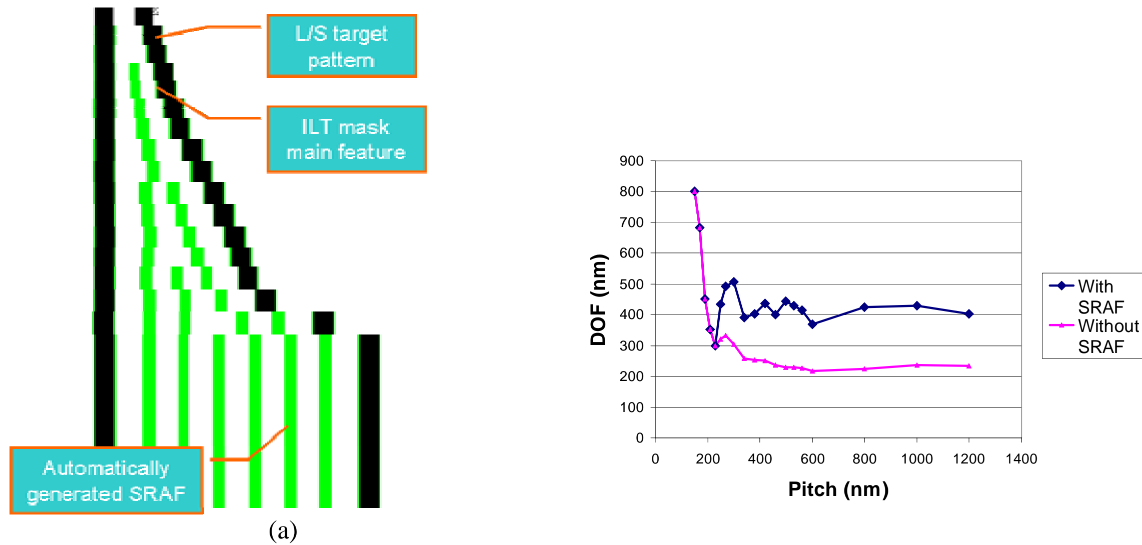
SRAF GENERATION

SRAF is commonly used in RET. There are three major problems in the current SRAF generation: 1) SRAF placements are primarily rule-based, and the rules are created using simple regular patterns, such as lines/spaces, or contacts with different pitches. Such rule might not applicable nor accurate for complicated geometries in a real design; 2) SRAF generation and OPC are two separate processes – first the SRAFs are generated and then OPC is run. This can be time consuming. 3) OPC only applies to main pattern. SRAFs usually are not optimized during OPC any more. 4) Side lobes which could print are difficult to detect and fix.

In ILT, SRAFs are automatically generated during the inversion calculation, and they are optimized simultaneously with the main features. Therefore, the SRAF generation becomes a straight forward, single step process. Since every pixel is considered in the computation; in the pixel-based ILT implementation, side-lobe-printing which has been a problem for edge-based OPC due to its fundamental edge-sampling-based approach does not occur.

To illustrate auto SRAF features described above, two examples will be shown in the following: one line/space sequence and one real contact layer pattern. In the first example, a simple line/space pattern with varying pitch is used as the target. The line CD in this sequence is kept as 70 nm, but the pitch is varied from 210nm to 800nm; therefore, this pattern covers dense, semi-densed, and nearly isolated conditions.

In the current OPC approach, to add SRAFs for line/space patterns, some simple rules are usually created; for instance, if the space between two adjacent line/space is smaller than a certain size, no SRAF; if the space is larger than certain size, add one SRAF with a certain size at a certain distance; if the space is larger than another distance, add two SRAFs, etc. In general, the number of SRAFs, when to add them, the SRAF location and dimension are all pre-defined based on simple rules.



(a) (b)

Figure.3. Simultaneous mask creation and SRAF placement using ILT for a sequence of pitches of clear field line/space patterns. (a) shows the line/space target (black) and ILT generated mask patterns (green/gray), including main features and SRAFs. (b) shows the DOF for through pitch patterns optimized with ILT with and without SRAF

Figure 3(a) shows this sequence of line/space target pattern and post-correction patterns, including SRAFs, to give the optimized wafer printing performance. Comparing with OPC approach, one can notice some distinguishing properties of SRAFs generated by ILT: the number of SRAFs depends on the pitch; and, the location and size of SRAFs change when pitch changes even for pitches where the number of SRAFs stays constant. Figure 3(b) shows the depth of focus (DOF) at 6% exposure latitude (EL) for this line/space pattern sequence. The DOF curve for line/space correction without SRAF shows the DOF drops as pitches increases; With SRAFs, DOF is approaching a constant after the space is large enough to place the first SRAF.

The example above contains only simple line-space, the next example (Figure 4) shows a real layout pattern, a 65nm SRAM contact layer. In this example, there are some regular rectangular SRAFs that may be equivalently generated using conventional OPC, there are also more complicated SRAFs, including the ring type of SRAF that is located in the center that are inconceivable by conventional OPC. Although not shown in the figure, no side lobes are present in operating ranges, despite the presence of quite large SRAFs.

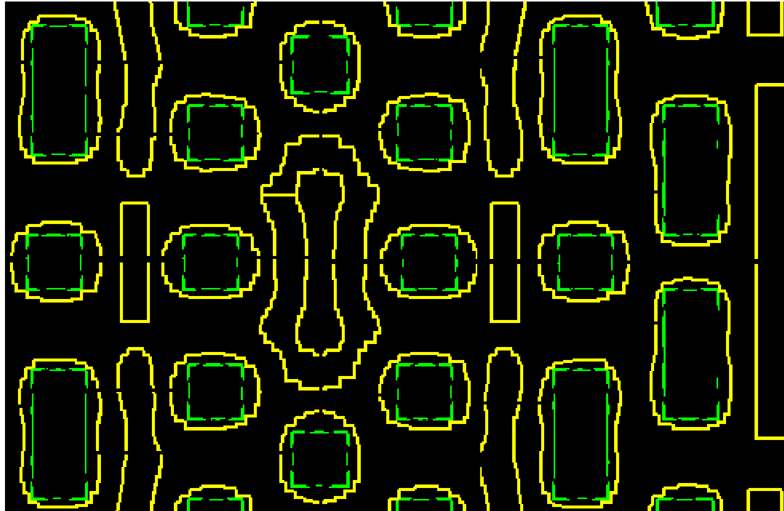


Figure 4. ILT automatically generated SRAF for a 65nm SRAM contact layer, where the green contour (dashed line) represents the drawn GDS target, and yellow contour (solid line) represents the mask pattern calculated by ILT including both main features and SRAFs.

PROCESS WINDOW ILT

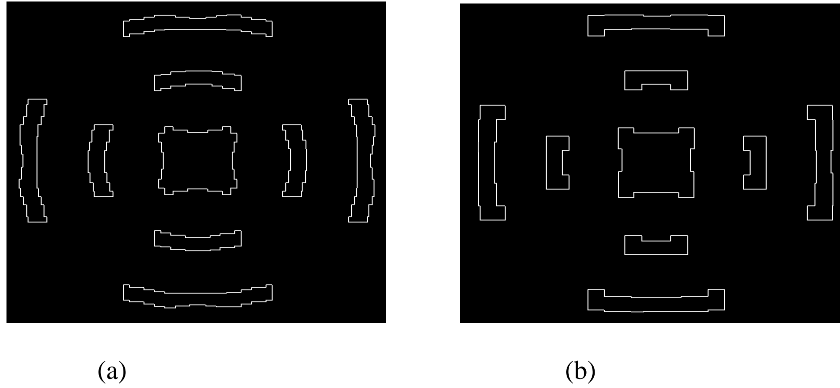
Most OPC software optimizes masks at the nominal process condition to seek the best pattern fidelity (the smallest edge placement error (EPE)) on a single image plane (best focus and nominal exposure). This worked well at 130nm and 90nm technology nodes when the process window, especially the depth of focus (DOF) was relatively large. This is not adequate, however, at the 45nm technology node and beyond. A good correction for nominal condition may cause bridging or pinching at slightly different exposure or focus. Such problems (called “hotspots”) are conventionally caught by verification systems after OPC and fixed manually at the OPC script level. Often, this process of detection and re-scripting must be repeated several times.

The process window “hot-spot” can be avoided in many cases, because there exist multiple solutions that give the same EPE at nominal. In the latest development of ILT, the process window margin is improved in two ways[13]: 1) Using multiple image planes, including the ones off the nominal condition, in the inversion calculation; 2) Adding terms directly related to DOF into the cost function.

MANHATTAN MASKS AND MRC IN INVERSION

In previous ILT implementations, the mask constraints were not considered, resulting in masks with curved geometry and many small fragments. Such masks are challenging for mask writing, inspection, metrology, and repair.

In Luminiescent’s ILT approach, the mask constraints are built into the inversion solver[14]. Users can specify mask rules, such as minimum CD, minimum space, minimum area, and minimum fragment length on the Manhattan shaped mask.

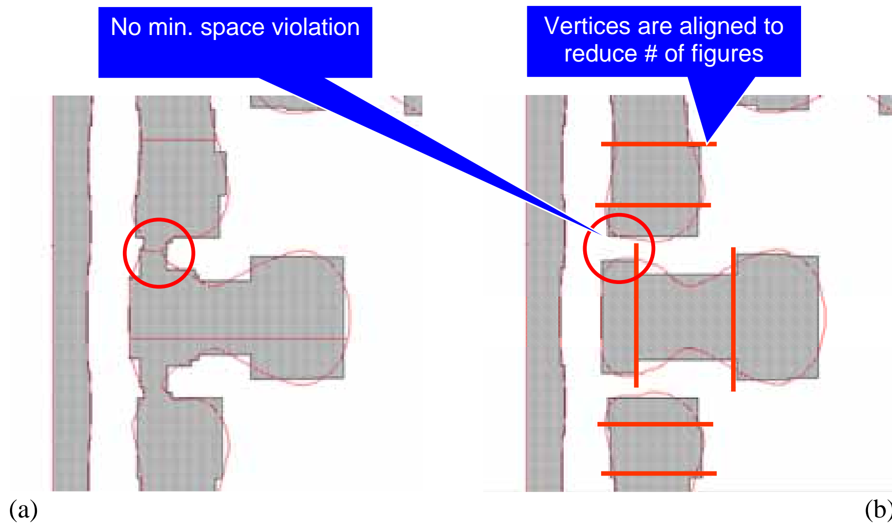


(a) (b)

Figure 5. ILT mask patterns for isolated contact with different user specified mask rules: (a) min. fragment length 20nm, min. line/space 25nm; (b) min. fragment length 50nm, min. line/space 50nm .

As shown in Figure 5, the user can specify different mask rules, in this case, the minimum fragment length 20nm in 4a, 50nm in 4b, and minimum line and space both 25nm at wafer scale. The ILT program will generate different patterns optimized based on the user specified settings.

Mask rules can also be enforced with ILT. During the inversion calculation, mask rules, especially minimum line, space, and area are checked. The resulting mask is constrained to satisfy the specified mask rules. For example, as shown in Figure 6, without mask rule enforced, the inversion may link two mask features together, creating a piece violating the minimum linewidth rule (Figure 6(a)). However, when mask rules are enforced during inversion, the algorithm won't allow such violation to happen, therefore, it generates different shapes of mask patterns that satisfy all mask rules (Figure 6(b)).



(a) (b)

Figure 6. An example showing (a) mask pattern calculated with minimum linewidth rule not enforced, and (b) mask pattern calculated with minimum linewidth and space rules enforced.

Mask data fracturing and mask writing time with VSB e-beam writers are also considered. As shown in Figure 6 (b), the vertices are aligned either in horizontal direction or vertical direction to reduce the number of e-beam figures.

SUMMARY AND CONCLUSIONS

Luminescent has developed the first commercial ILT solution, suitable for use in a production environment. It can rapidly solve for the optimal manufacturable photomask design, with turn-around times comparable or superior to other mask synthesis solutions such as OPC. By formulating the problem in a rigorous mathematical framework, we find highly optimal solutions which do not arise from traditional ad hoc approaches. The resulting masks often provide substantially improved depth-of-focus and exposure latitude, enabling geometries that may be otherwise unattainable.

REFERENCES

1. B.E.A. Saleh and S.I. Sayegh, “Reductions of errors of microphotographic reproductions by optical corrections of original masks”, Optical Eng. Vol. 20 pp 781-784 (1981)
2. K.M. Nashold and B.E.A. Saleh, “Image construction through diffraction-limited high-contrast imaging systems: an iterative approach”, J. Opt. Soc. Am.A, vol. 2 p. 635 (1985)
3. Y. Liu and A. Zachor, “Optimal binary image design for optical lithography”, Proc. SPIE Vol. 1264 pp 410-412 (1990)
4. Y. Liu and A. Zachor, “Binary and phase-shifting image design for optical lithography”, Proc. SPIE Vol. 1463 pp 382-399 (1991)
5. A. Rosenbluth et. al, “Optimum mask and source patterns to print a given shape”, JM3 vol. 1 pp 13-30 (2002)
6. Y-T Wang, Y.C. Pati, H. Watanabe and T. Kailath, “Automated design of halftoned double-exposure phase-shifting masks”, Proc. SPIE Vol. 2440 pp 290-301 (1995)
7. Y.H. Oh, and J-C Lee, Resolution enhancement through optical proximity correction and stepper parameter optimization for 0.12-um mask pattern, Proc. SPIE Vol. 3679 pp 607-613 (1999)
8. T. Fuhner and A. Erdmann, “Improved mask and source representations for automatic optimization of lithographic process conditions using a genetic algorithm”, Proc. SPIE Vol. 5754 pp 415-426 (2005)
9. S. Osher and J. A. Sethian, Fronts Propagating with Curvature-Dependent Speed: Algorithms Based o Hamilton-Jacobi Formulations, Journal of Computational Physics 79, pp 12-49 (1988)
10. D. Abrams and L. Pang, Fast Inverse Lithography Technology, 31st Internal Symposium of Microlithography, Proc. of SPIE Vol. 6154, San Jose, California, USA, Feb. 2006
11. C. Y. Hung, et al, Pushing the Lithography Limit: Applying Inverse Lithography Technology (ILT) at 65nm generation [6154-58], 31st Internal Symposium of Microlithography, Proc. of SPIE Vol. 6154, San Jose, California, USA, Feb. 2006
12. B. Lin, et al, Inverse Lithography Technology at Chip Scale, 31st Internal Symposium of Microlithography, Proc. of SPIE Vol. 6154, San Jose, California, USA, Feb. 2006
13. Y. Liu, et al, Inverse Lithography Technology Principles in Practice: Unintuitive Patterns, 25th Annual BACUS Symposium on Photomask Technology, Proc. of SPIE Vol. 5992, Monterey, California, USA, Oct. 2005
14. L. Pang, et al, Inverse Lithography Technology (ILT), What is the Impact to Photomask Industry? Photomask Japan, Yokohama, Japan, April, 2006