

Inverse Lithography Technology (ILT) and Its Readiness for Production in Advanced Technology Nodes

Linyong Pang, Yong Liu, Dan Abrams
Luminescent Technologies, Inc.

In this paper, we present the first ILT approach that can rapidly solve for the optimal photomask design and that is suitable for use in a production environment. We will discuss the latest development of ILT at Luminescent, in particular, in the areas of sub-resolution assist feature (SRAF) generation, mask rule compliance (MRC), and hardware acceleration. Results collected internally and from customers demonstrate that ILT is no longer just an R&D tool, but is in fact ready for production qualification at advanced technology nodes. By optimizing each element of the process, ILT can improve process windows while maintaining mask costs at a reasonable level.

ILT AND ITS TECHNICAL ASPECTS

Deep sub-wavelength lithography leads to smaller process windows, reduced yields and higher mask costs, impeding chipmakers' ability to keep pace with Moore's Law. Lithography is fundamentally constrained by basic principles of optical physics. At 45 nm, a line is less than one fourth 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 that is theoretically possible can be achieved by considering the design of photomasks as an inverse problem -- and then solving the inverse problem to find the optimal photomask for a given process, using a rigorous mathematical approach. Inverse Lithography Technology (ILT) has been explored for many years[1-8]. Although these early approaches to ILT often resulted in superb lithography, 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.

By ILT we mean the following: given the known forward transformation of a lithography process, ILT mathematically determines the optimized mask which produces the desired wafer target with the best pattern fidelity and largest process window. The forward transformation is modeled accurately, which 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. However, the strict inverse problem is ill-posed; because the forward operator is many-to-one (that is, many different masks will yield identical on-wafer results), and the function has no well-defined inverse. Moreover, for typical target patterns (e.g., a drawn layout with Manhattan geometry and sharp corners), there does not exist any mask function which

will produce the exact drawn wafer target. 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). This function is indicative of the quality of the solution, or the “goodness” of the mask. In a simple case, the Hamiltonian could be 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 normalized image log slope (NILS) of the image, the robustness against mask error enhancement factor (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.

Our approach to ILT is based on a branch of mathematics invented by Stan Osher at UCLA. Commonly known as level-set methods[9], these techniques have been applied to the solution of inverse problems in a wide range of engineering disciplines such as image processing and fluid dynamics. The mask and wafer contours are represented by level-sets in such methods. 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.

Another 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 width. 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.

A key distinctive feature of ILT is the absence of pattern-dependent heuristics, and the ability to broadly explore wide areas of solution space. ILT algorithms frequently lead to mask patterns which are unanticipated by a knowledgeable practitioner. As will be shown in the next example of the problem of placement of sub-resolution assist features (SRAFs), there are mask features that do not print on the wafer which are detached from the edges of the main mask patterns, 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. In contrast, ILT can determine optimal SRAFs simultaneously with the rest of the mask. The absence of segmentation scripts is a significant advantage because it

usually requires significant engineering resources to write such scripts for different patterns on different design layers.

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.

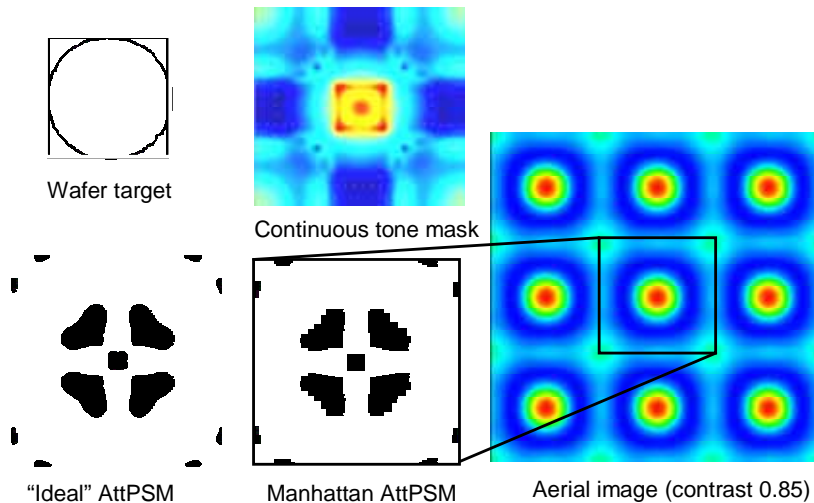


Figure 1. ILT contact array example: contact target of 110nm CD, 440nm pitch. Three types of masks were generated from ILT: continuous tone mask, "ideal" attenuated phase-shifting mask, attenuated phase-shifting mask with Manhattan constraints, and the aerial image simulated using Manhattan AttPSM.

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 cm² 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.

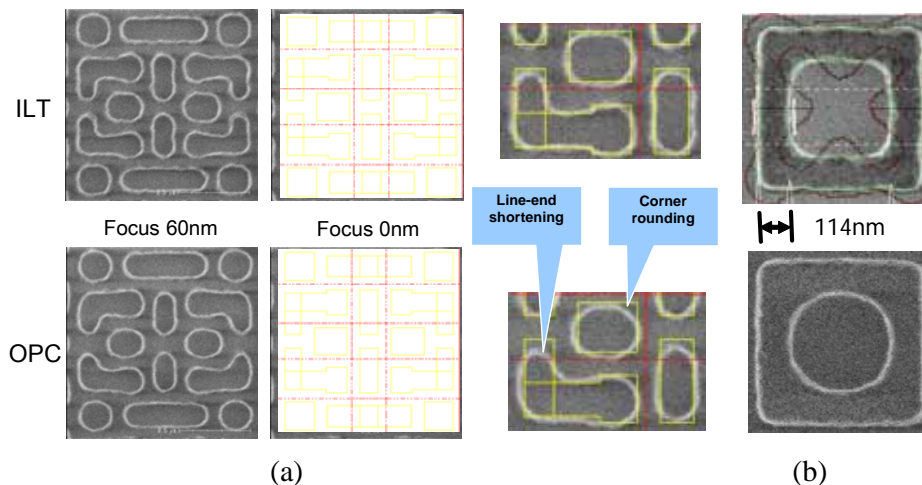


Figure 2. ILT and OPC wafer images. (a). Through focus wafer patterns of metal 1 layer, where outline is the design target. The insets at right are a magnification of a selected portion of the 0nm focus condition. (b). a square "donut" shape (CD 114nm on wafer).

LATEST DEVELOPMENT IN ILT

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-dense, 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.

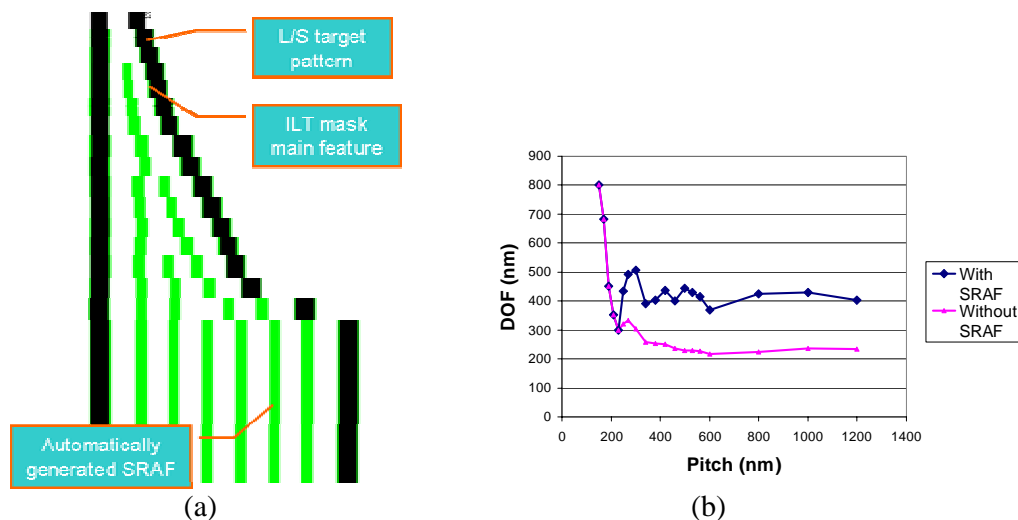


Figure.3. Simultaneous mask creation and SRAF placement using ILT for a sequence of pitches of clear field line/space patterns. (a) shows the line/space target pattern and ILT generated mask patterns (green/gray), including main features and SRAFs. (b) shows the DOF for through pitch patterns optimized 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.

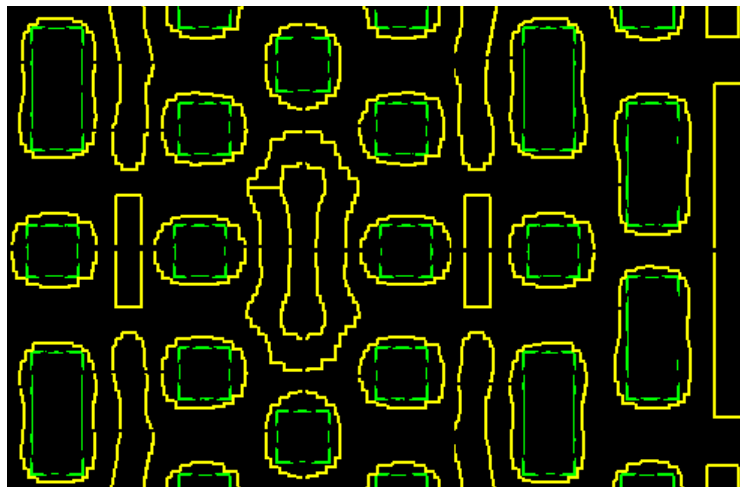


Figure 4. ILT automatically generated SRAF for a 65nm SRAM contact layer, where the green contour (dashed line) represents the drawn GDS target, and yellow contour (solid line) represents the mask pattern calculated by ILT including both main features and SRAFs.

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 Luminescent'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.

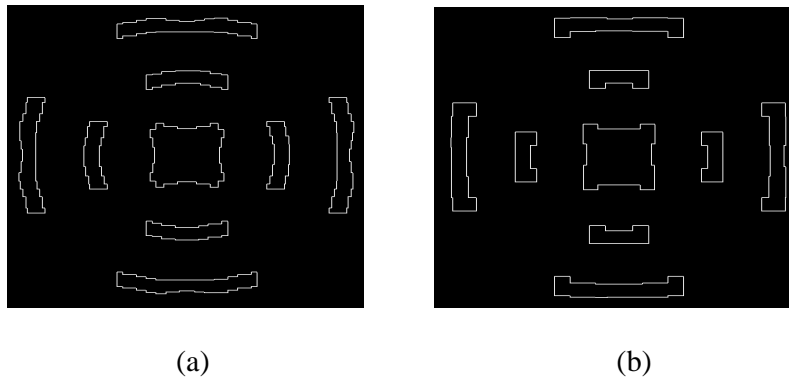


Figure 5. ILT mask patterns for isolated contact with different user specified mask rules: (a) min. fragment length 20nm, min. line/space 25nm; (b) min. fragment length 50nm, min. line/space 50nm .

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)).

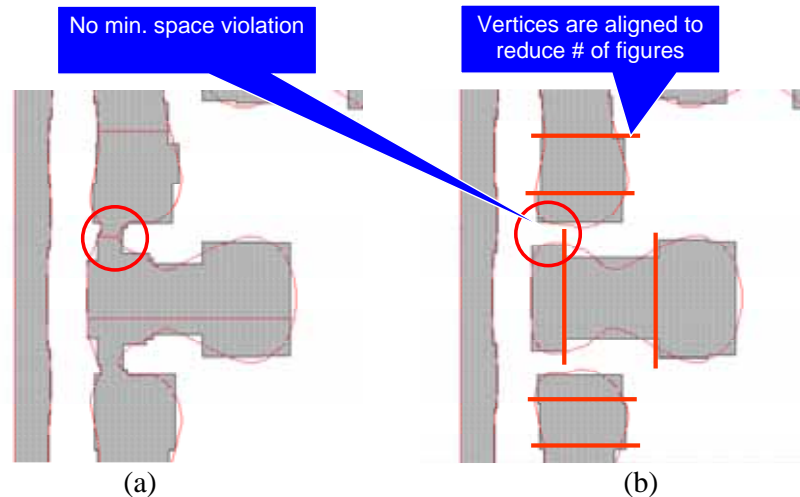


Figure 6. An example showing (a) mask pattern calculated with minimum linewidth rule not enforced, and (b) mask pattern calculated with minimum linewidth and space rules enforced.

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.

ILT ENABLING 45NM LITHOGRAPHY WITH DRY STEPPERS

Recent customer tape-outs show indications that ILT can enable the 45nm generation with dry steppers for certain layers that would otherwise require immersion steppers.

One of the common myths is that ILT cannot change the basic resolution physics and therefore cannot improve the fundamental resolution. This is not correct. First, the resolution is determined by how many orders of diffraction that a pattern can receive through the stepper optical system. If only zero order diffraction is received at the wafer image plane, then one would only see a constant background, in other words, the pattern cannot be resolved. That is the fundamental limitation of resolution. There are two common practices to increase the number of diffractions at the wafer image plane: using off axis illumination (OAI) or using phase shift mask. The applications of these approaches were extensively explored in the last two decades because the computation required is manageable by human beings. However, there is a third approach which is less popular and has only been explored recently – using the interaction of diffraction on mask patterns. SRAF can be considered as the first try in this approach for simple isolated line patterns. To make this approach applicable to complicated random patterns on the real chip designs, a fast inverse lithography computation engine is required. Second, for each generation, the resolution is actually not limited by the dense line/space patterns; in reality, it is the line-end and other features that limit the lithography capability. For example, in the SRAM core, resolving the line/space is not an issue; the issue is how to break line/space into small segments. ILT can help in such cases, because it requires a sophisticated consideration of shapes in surrounding area to properly manage the constructive and destructive interferences of higher order diffractions. Third, the K1 factor that a foundry feels comfortable to use is usually larger than 0.38, which is high; such K1 factor is

calculated from dense line/space patterns. The reason such high K1 factor is required is that the other geometries, such as line/end, have smaller K1 factor, and foundries have to consider all types of geometries to be safe. ILT is the best way to utilize higher diffraction orders for random geometries. With ILT, the lithography capability can be extended to its ultimate limits. .

We have seen ILT is able to improve the DOF of the most challenging contact layer in a 45nm DRAM core from about 200nm to roughly 400nm, and at the same time allowing the use of a less expensive stepper. In another example ILT has enabled 45nm SRAM critical layers (poly, diffusion) with single exposure, dry stepper annular illumination. These examples – both in simulation and on wafer – show that ILT can be used to extend the capabilities of dry lithography into situations which would otherwise have required immersion.

CONCLUSIONS

In this paper, we briefly described the frame-work formulation of ILT with some specifics of the implementation of ILT by Luminescent Technologies, and presented the latest developments of ILT. We showed ILT's unique ability to optimize main features and SRAF at the same time. In addition, we addressed some concerns about the manufacturability of ILT masks and mask rule compliance, specifically we showed with examples that ILT is adaptable and can generate masks with Manhattan edges that satisfy mask rule constraints. Masks with those constraints continue to exhibit unique features of ILT. Finally, we discussed the possibility of using ILT and dry stepper to print some 45nm critical layer patterns.

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. 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)
6. S-H Jang et. al, Manufacturability evaluation of model-based OPC masks, *Proc. SPIE* vol. 4889 p 520 (2002)
7. A. Rosenbluth et. al, Optimum mask and source patterns to print a given shape, *JM3* vol. 1 pp 13-30 (2002)
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 on Hamilton-Jacobi Formulations, *Journal of Computational Physics* 79, pp 12-49 (1988)
9. 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
10. 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
11. 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
12. 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
13. L. Pang, et al, Inverse Lithography Technology (ILT), What is the Impact to Photomask Industry? Photomask Japan, Yokohama, Japan, April, 2006

ACKNOWLEDGMENTS

The authors would like to thank the whole Luminescent team for creating the ILT implementation that made this work possible. The authors would especially like to thank Grace Dai for her help in preparing ILT mask patterns for this presentation, Andrew Moore and Paul Rissman for reviewing the manuscript.