32nm and below Logic Patterning using Optimized Illumination and Double Patterning

Michael C. Smayling*[a], Valery Axelrad[b]
[a]Tela Innovations, Inc., 655 Technology Pkwy, Suite 150, Campbell, CA, USA 95008
[b]Sequoia Design Systems, Inc., 137 Chapman Rd., Woodside, CA USA 94062

ABSTRACT

Line/space dimensions for 32nm generation logic are expected to be ~45-50nm at ~90-100nm pitch. It is likely that the node will begin at the upper end of the range, and then shrink by ~10% to a "28nm" node. For the lower end of the range, even with immersion scanners, the Rayleigh k_1 factor is below 0.32. The 22nm logic node should begin with minimum pitches of approximately 70nm, requiring some form of double patterning to maintain k_1 above 0.25.

Logic patterning has been more difficult than NAND Flash patterning because random logic was designed with complete "freedom" compared to the very regular patterns used in memory. The logic layouts with bends and multiple pitches resulted in larger rules, un-optimized illumination, and a poorly understood process windows with little control of context-dependent "hot spots."[1]

The introduction of logic design styles which use strictly one-directional lines for the critical levels now gives the opportunity for illumination optimization. Gridded Design Rules (GDR) have been demonstrated to give area-competitive layouts at existing 90, 65, and 45nm logic nodes while reducing CD variability.[2] These benefits can be extended to ≤32nm logic using selective double pass patterning.

Keywords: Low k_1, gridded design rules, restricted design rules, double patterning, lines and cuts

1. INTRODUCTION

The past 50 years have been the era of continual improvement of photolithography resolution. This has allowed integrated circuit designs to shrink by more than a factor of one hundred. Prior to sub-100nm technology nodes, the exposure equipment resolution improvement trend kept the k_1 value above 0.6, where k_1 is the fitting factor in the Rayleigh equation $CD = k_1 \lambda/NA$. Immersion lithography has been quickly introduced to extend the roadmap.[3]

k_1 has been decreasing for recent logic technology node as shown in Figure 1. To maintain pattern fidelity at k_1 values below ~0.6, resolution enhancement techniques (RET) such as optical proximity correction (OPC), off-axis illumination (OAI), and phase shift masks (PSM) have been introduced. Note that the "22S" point is not realizable with $k_1 < 0.25$.

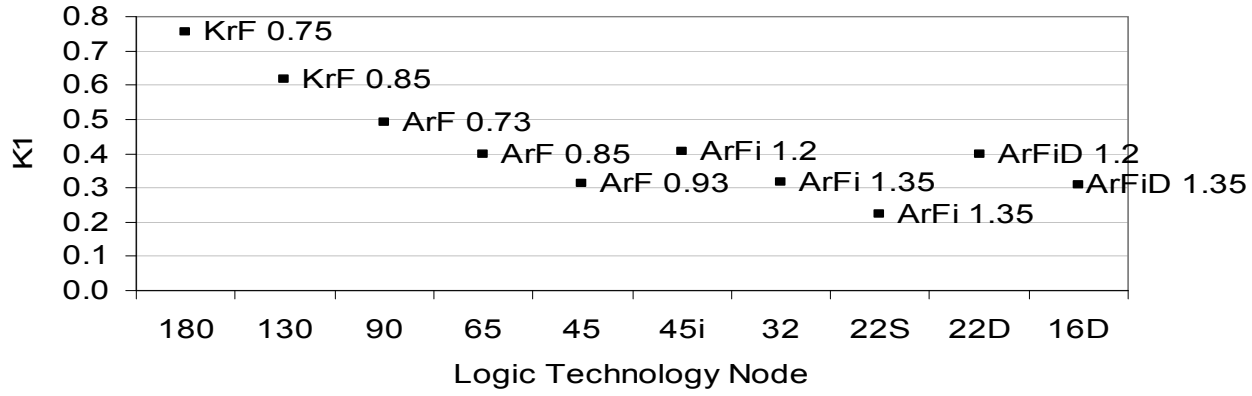


Fig. 1. k_1 trend for sub-200nm logic technology nodes.

*mike@tela-inc.com; phone +1 408 558-6321; fax +1 408 559-4600

Optical Microlithography XXII, edited by Harry J. Levinson, Mircea V. Dusa, Proc. of SPIE Vol. 7274, 72740K · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.814258

As k_1 decreases, "practical limits" are imposed by the design style.[1] 2D layouts with bent polygons are limited to ~0.35. A 1D layout style with parallel straight lines looking much like a grating pattern, and has a limit of ~0.28. Extensive efforts are being made to define "restricted design rules" which allow bends but with constraints on widths or spaces.[4] A 1D layout style with further requirements for keeping lines on a regular grid permits using a simplified set of design rules described as "gridded design rules."[5]

2. GDR CONCEPT

Gridded design rules can be better understood when compared to "complex design rules" (CDR). Figure 2 shows an example of two functionally equivalent layouts with the left side drawn with 2D CDR and the right side drawn with 1D GDR. Three problem areas are highlighted in the 2D CDR case. Site 1 represents a transistor gate line which is isolated in the x-direction from other lines; it will have a reduced process window as compared to the same gate line in the 1D GDR case. Site 2 indicates a gate line in a more dense environment, but it will also have a reduced process window since it is relatively isolated on one side. Finally, site 3 shows gate lines in a congested 2D environment; this site is susceptible to necking and bridging hotspots in addition to reduced process window.

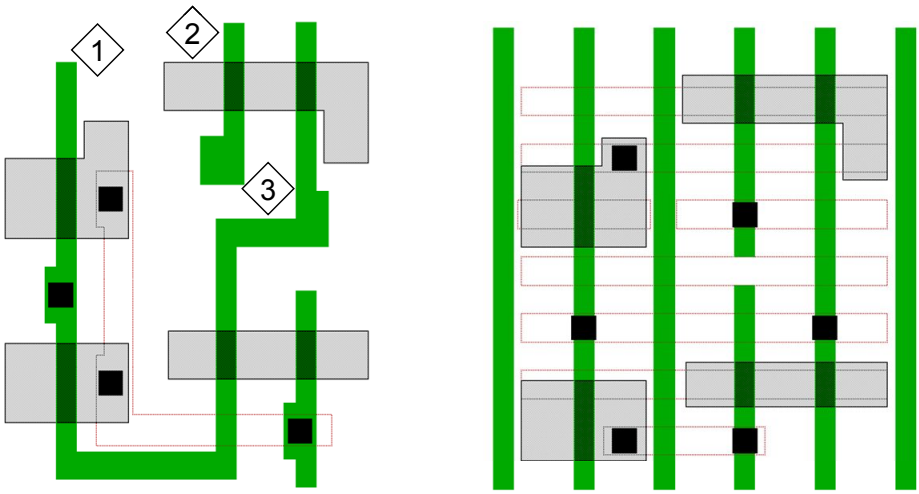


Fig. 2. 2D CDR layout (left side) compared to 1D GDR layout (right side).

As illustrated by the right side of Figure 2, the vertical gate lines are on a uniform pitch with dummy lines as needed. The horizontal first metal lines are also on a uniform pitch with circuit line segments separated by uniform gaps. Since the gate and first metal lines are on perpendicular grids, the diffusion and gate contacts are automatically located at intersections of the grid lines.

In a 1D GDR layout, the design rules are greatly simplified because they involve widths, spaces, and end-gaps. Overlap rules, like first metal end-overlap of a contact, are built into the first metal end-gap rule and hence are redundant. Similar logic applies to diffusion-to-contact, gate-to-contact, and metal-to-via overlaps. The Tela Canvas™ implementation of 1D GDR builds these constraints into the logic cell architecture to give "correct by design" layout.

One significant difference between 45nm layout and ≤32nm layout will be the active or diffusion layer. Whereas diffusion wires with 2D layouts were used at 45nm, as shown in Figure 2, for example, it is likely that only rectangular shapes will be allowed at 32nm and below. Diffusion jogs may be allowed to permit transistors with different widths on adjacent gate tracks at 32nm, but even this practice will be eliminated at 22nm. Of course, if 22nm technology uses FinFETs, then the diffusions will be 1D lines by definition and the pertinent rules will be pitch and end-gap.

An additional benefit of 1D GDR is that no pattern-specific restricted design rules (RDRs) are required. These rules were introduced to deal with increasingly complex interactions between 2D shapes, but could not completely guarantee manufacturable layouts. EDA tools for "hotspot" detection and correction were introduced on top of RDRs, but since 2D cells have context-dependent behavior, the design hierarchy can be broken by any hotspot fixing on any mask layer.

3. MODELING AND SIMULATION

The starting point for new technology development is modeling and simulation based on experience from previous technology nodes.[6] This is often problematic since models for new equipment and processes are usually not available until well into the development project. Fortunately for the foreseeable future, with λ/NA limited to 143nm, the available models are reasonably mature.

3.1 Lithography Optimization Strategy

The linkage between layers in 1D GDR layout requires that those linked layers be co-optimized. Intra-layer constraints include optical resolution, photoresist thickness and optical properties, pattern complexity, device and electrical requirements, and CD variability requirements.[7,8,9,10] Inter-layer constraints include alignment to perpendicular layers, alignment to hole layers, and the pitches of perpendicular layers. Each layer may have a different solution based on the weighting of these factors.

The variables available for the optimization include lithography, mask, RET, circuit design, and potential process extensions. Lithography factors include wavelength, numerical aperture, illuminator, polarization, photoresist and resist processing, and anti-reflective coatings. Mask factors include phase shift options, mask materials, and mask writing/inspection trade-offs. RET factors span several regimes, from mask data preparation such as OPC to illuminator conditions and mask type. Design factors include the layout style and the design purpose of different features. Process extensions such as self-aligned layers and spacer double patterning can dramatically shift the patterning burden from the exposure tool to other parts of the integrated process.

Source-mask-optimization (SMO) has been common for memory style layouts for many years, with more application to random logic in recent years. The problem with SMO for arbitrary 2D layouts has been that the solution either approaches a “lowest common denominator” with annular or at best quadrupole illumination, or the illuminator pattern becomes very complex and in some cases pixelated. Mask patterns based on an inverse transform method create patterns which may mathematically produce the desired aerial image but which may be costly to write and extremely difficult to inspect.

By including design factors and potential process extensions during the optimization, a completely different solution space can be explored. For example, the 1D GDR design style allows dipole illumination and OPC with 50% smaller output files. Extending 1D GDR to include lines / cuts (1D GDR-LC) allows even further illuminator optimization and the potential for no OPC on the line patterns. 1D GDR-LC is also extendable using SADP for the lines to at least the 16nm logic node.[11]

3.2 Lithography optimization tool

Lithography simulation and optimization of lithographic conditions are carried out using SEQUOIA Cell Designer (SCD).[12] SCD provides advanced physical simulation capabilities necessary for 32nm and 22nm technology nodes including high NA, aberrations, liquid immersion, chromatic distortion due to finite laser spectral distribution, and others. To achieve low k1 values indicated in Fig.1, optimization of all available parameters of the lithographic process is typically necessary and therefore general optimization functionality is a key component of SCD. Direct numerical optimization is used with arbitrary optimization variables (for example illumination shape, sigma, etc.) and general optimization cost function such as NILS, DOF, CD values, or even electrical simulation results.[8]

The regular structure of 1D GDR layouts provides an excellent opportunity for design-process co-optimization, in which lithographic parameters are optimized to achieve the best results for the design. In less regular conventional logic layouts many layout pitches and orientations are present, which forces the optimizer to choose an overall compromise. On the other hand, with a GDR layout it is very likely that a highly optimized lithography condition can be found which will provide a wide process window with measurably improved pattern fidelity using the same lithography hardware.

Lithography parameter optimization in SCD is therefore used as a process optimization tool (Manufacturing for Design) to get the best image quality for a given design and litho equipment set.

3.3 Simulation Results for 32nm logic

The Active layer is critical for SRAM memory cells and much less critical for other random logic like standard cells. Depending on the number of horizontal metal-1 lines in the logic architecture, the Active pitch can be two to four times larger than the Active pitch in the SRAM. Hence, the challenge for Active is to get good Active CD control (equivalent to transistor width) at semi-dense pitches while supporting a dense pitch in the SRAM bit cell.

The Gate layer is commonly fabricated using lines / cuts starting from the 45nm logic node where end-of-line pullback limited SRAM bit cell scaling.[13] Figure 3a shows the Gate layer pattern for a complex logic cell. Figure 3b shows how the pattern is split into two layout layers, the line pattern and the cut pattern. The Gate line pattern is expected to have a width of 30nm and a space of 90nm at the 32nm node. This is a relatively relaxed pitch due to the Active contacts placed between the transistor gates. No OPC is required for the lines, although dummy lines at the edges of memory and logic blocks can be increased in width to avoid lifting lines and the line pattern needs to extend past the blocks slightly to accommodate end-of-line pullback. The cut part of the Gate layer is more critical, and can have requirements approaching those of other "hole" layers like Contacts and Vias.

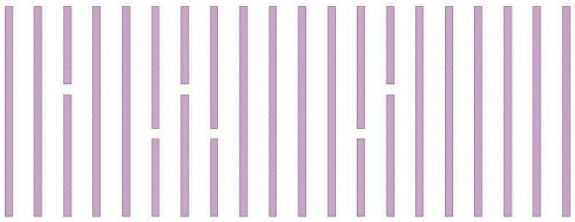


Fig. 3a. Designed Gate pattern showing final shapes for a complex logic cell.

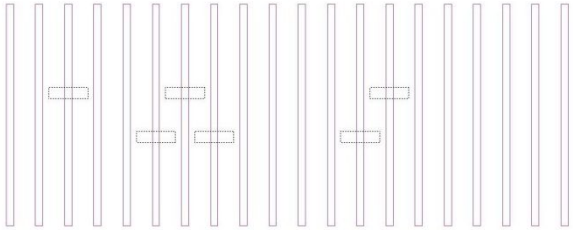
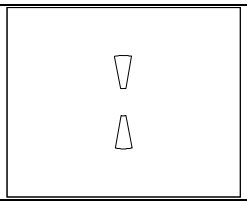
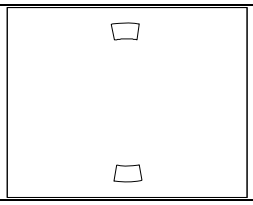
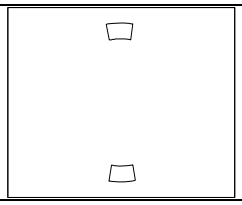
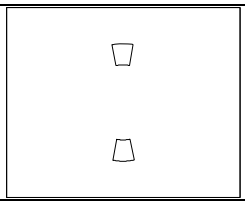


Fig. 3b. Gate line and cut layout of the logic cell.

The Gate line pattern was easily optimized with a dipole illuminator. The cut pattern took more effort because of the combinations of patterns to consider. The cut length also a variable, since the "x" direction results are not so critical as long as the final pattern has good fidelity at the intersection of the line to be cut, and does not overlap adjacent lines which are not to be cut. The cut width is the critical dimension, since it affects the final overlap of the Gate lines with Contacts or Active regions.

Gate cut optimization results are shown in Table 1. OPC was done for each illumination setting as part of the optimization loop. The optimization cost function was either CD (minimum distance between target and simulated CD) or NILS (normalized intensity log slope) used at selected cut locations. As can be seen in Table 1, the choice of the length of the cut feature impacts the optimal illuminator, especially for shorter cuts.

Table 1. 32nm Gate cut results for NILS optimization

Length, nm	120	140	160	180
Illuminator				
Phi-vert, deg	20.5	19.8	19.8	20.5
Phi-hor, deg	0.0	0.0	0.0	0.0
Sigma	0.59	0.98	0.98	0.59
Sigma-inner	0.10	0.78	0.78	0.47
Common parameters	NA = 1.35, water immersion, binary mask, no polarization			

The aerial image results are shown in Figures 4a and 4b. NILS optimization was used for both cases. Figure 4a has a cut with an initial length of 120nm, while 4b has a length of 160nm. The extra initial length gives a more rectangular final shape, but the case of adjacent cuts shows a tendency toward bridging. The NILS optimization maximizes the image intensity slope rather than just CD values and therefore should give a wider process window because at maximum NILS a smaller CD error results from the same exposure variation.

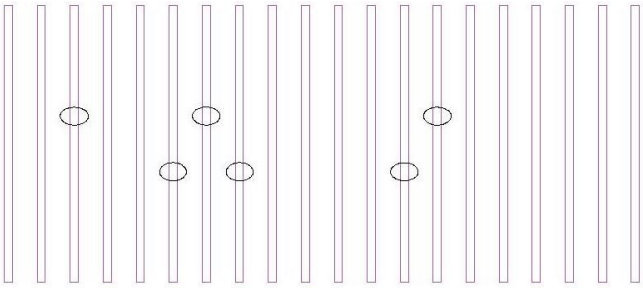


Fig 4a. Gate cut aerial image with 120nm length.

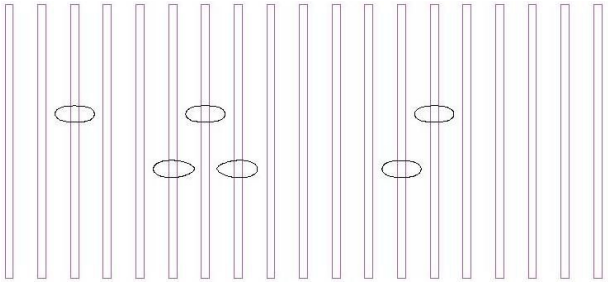


Fig 4b. Gate cut aerial image with 160nm length.

The Contact pattern is expected to have a hole size of 40x40nm and a minimum pitch of 116nm – 120nm. The Metal-1 pattern will have a pitch of 100nm to maintain 70% scaling from 45nm. Both of these layers should be suitable for single-pass patterning using Cquad or Quasar illumination.[6] Line-end shortening for the Metal-1 layer can be accommodated in the end-gap spacing rule, applied globally to the design.

3.4 Simulation Results for 22nm logic

The Active layer at 22nm will be limited by the needs of the SRAM bit cell. For either FinFET lines or planar FET rectangles, the Active pitches will be large compared to those of the other critical layers. A line / cut approach will probably be used for FinFETs for obvious reasons. Conventional lithography suitable for the bit cell will be adequate for the 1D GDR style Active shapes. If tight fin pitch is required, a line process like SADP can be used.[11]

The Gate layer will become more challenging, with a line pitch of 84nm and cut widths of 36nm. The line pattern can still be handled by a single pass of dipole illumination at NA=1.35; since the pitch is dictated by the contact-to-gate space, the line width and space can be patterned at 1:1 then trimmed to the desired final size.

The Gate cut pattern scales from 32nm, and has a width of 36nm and a length of 84 to 116nm. Since the width is below the rated resolution of current scanners, the patterned size can be enlarged and then shrunk with a process like SAFIER from TOK.[14] The same pattern was used as for the 32nm case, although pitches and end-gaps could have been easily adjusted.

As for the 32nm patterns, the 22nm patterns were sensitive to the cut length. To aid the optimization, not only were the cut widths included in the cost function, but the space between the two adjacent cuts was included to ensure adequate separation in the aerial image. The space was critical and optimization attempted without including the space ended up with bridging in most cases.

The Gate pattern results are shown in figure 5a, where the cut aerial image is combined with the line pattern. The line-end shape is not a problem since it is a gridded space away from a Contact or Active edge. The optimal illuminator shape is shown in figure 5b as a vertical dipole. The angle is 8.5 degrees, sigma is 0.59, and the inner sigma is 0.36. The cuts have enough of a horizontal major axis to allow the optimizer to select the vertical dipole over quadrapole, annular, or horizontal dipole alternatives.

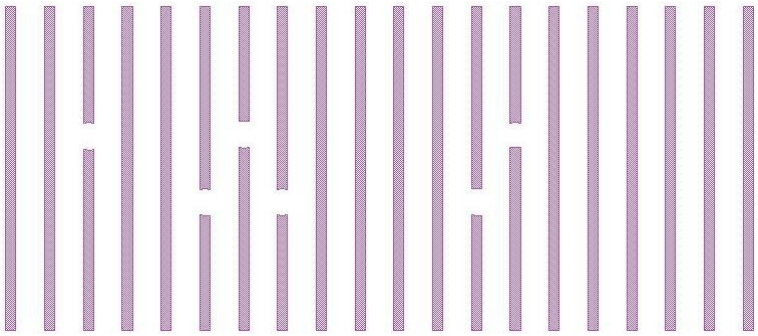


Fig. 5a. 22nm Gate final pattern.

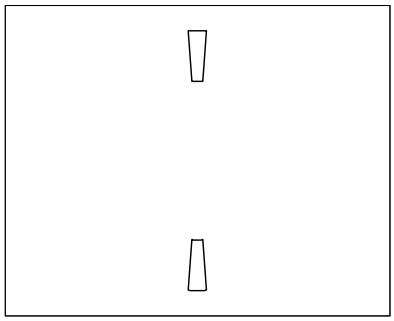


Fig. 5b. Illuminator shape after optimization.

The Gate cut simulations were also tried using polarization. As expected for the vertical dipole illuminator, the optimal polarization direction was perpendicular to the dipole axis, hence in the "x" direction. The NILS optimization was improved by 20% using polarization, and the run time actually decreased by 7%. Optimization of the mask, using MoSi instead of Cr, and different attenuation factors, was not included in this study but can give additional improvement.[15]

The Contact layer was not included in the work because it is not a line-cut double patterning layer. However, based on the results of the Gate cut optimization work, it is likely that the Contact layer will require a limited pitch range, it may be split into Active and Gate contacts, and it may require subresolution features.[16] Given the low exposed area, multi-e-beams or electron-projection-lithography may also be considered.[17]

The Metal-1 layer requires double patterning at 22nm because of the 70nm pitch. The line pattern can be created by techniques like SADP which allows patterning at twice the final pitch, then using sidewall spaces to double the spatial frequency.[11] An approach like SADP is suitable for either a subtractive-etch Metal-1 process or a single Damascene trench process.

The Metal-1 cut pattern is different from the Gate cut pattern in two factors. First, the Metal-1 cuts are either islands, using positive photoresist, or holes if negative photoresist is used. Second, the length of the cuts is limited since the Metal-1 line pitch is much smaller than the Gate line pitch. As in the case of the Gate cut pattern, the image contrast is very low, requiring careful optimization. Each of the cuts was included in the NILS optimization.

The Metal-1 optimized results are shown in Figure 6a, using the illuminator shown in Figure 6b. The cuts all have the same width, but may be multiple pitches long depending on the interconnect function. The optimal illuminator shown in Figure 6b is a horizontal dipole with an angle of 37 degrees, a sigma of 1.0, and an inner sigma of 0.9. The cut pattern originally had a width of 36nm, matching the line spacing, but this was below the resolution limit of the 1.35NA immersion lens. The drawn space was increased above 40nm to allow patterning; the final width can be reduced by photoresist trimming for islands or SAFIER for holes. The trench corners show pointed regions as a result of the combination of the line and cut patterns; advanced barrier/seed deposition processes should be able to completely coat the trench walls prior to electroplating.

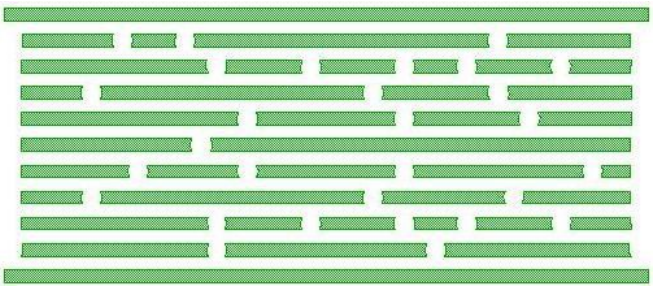


Fig. 6a. 22nm Metal-1 final Damascene trench pattern.

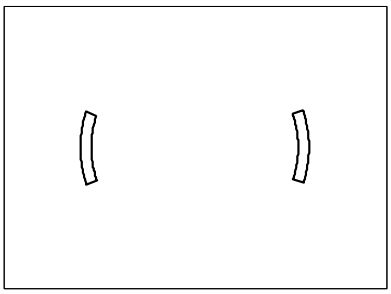


Fig. 6b. Illuminator shape after optimization.

The Metal-1 pattern in Figure 6a was a result of design adjustments and the photolithography optimization. By using the 1D GDR-LC layout style, a single parameter such as cut width could be globally changed and then a new optimization performed. 1D patterns also allow more flexible optimization than 2D patterns.[18] A design space can be evaluated to find design parameters and photolithography settings which are compatible.

4. CONCLUSIONS

The photolithography process for 32nm and 22nm logic cells has been optimized for layouts using 1D GDR-LC. 193nm immersion lithography is suitable for 32nm technology, with a line-cut double patterning step at the Gate layer. ~70% scaling from the 45nm node can be supported. At the 22nm node, high NA immersion lithography is still able to meet the patterning needs for Gate and Metal-1 cuts with a 70% shrink from 32nm. While the Gate line pattern can still be handled optically, the Metal-1 line pattern will need a process assist like SADP to meet the pitch requirements. Hole patterns such as Contacts and Vias will benefit from gridding; at 22nn, double patterning and/or assist features will improve process margin.

The 22nm optimizations were much more difficult to achieve than the 32nm cases. Lower contrast and operation at lower k_1 values at 22nm resulted in sensitivity to the simulation environment. Much like "gridding" in a classical TCAD problem, the choice of surrounding environment, resolution, point sources, and OPC parameters could artificially change the outcome.

We would like to thank the staff at Sequoia Design Systems and Tela Innovations for their help in creating layouts and running the OPC and simulation jobs. We appreciate the continued support of our executive management. Special thanks go to Michael Duane for his review and suggestions.

REFERENCES

[1] W. Arnold, "Lithography for the 32nm Technology Node," IEDM 32nm Technology Short Course (2006).
[2] M. C. Smayling, H. Y. Liu, L. Cai, "Low k_1 logic design using gridded design rules," Proc. of SPIE vol. 6925 (2008).
[3] B.J. Lin, "Immersion lithography and its impact on semiconductor manufacturing," Proc. of SPIE vol. 5377 (2004).
[4] L. Capodieci, "From Optical Proximity Correction to Lithography-Driven Physical Design (1996-2006): 10 years of Resolution Enhancement Technology and the roadmap enablers for the next decade," Proc. of SPIE vol. 6154 (2006).
[5] M. C. Smayling, "Gridded Design Rules – 1-D Design Enables Scaling of CMOS Logic," Nanochip Technology Journal, vol. 6(2), (2008).
[6] S. Mimotogi, et al, "Patterning Strategy and Performance of 1.3NA Tool for 32nm Node Lithography," Proc. of SPIE vol. 6924 (2008).
[7] M. Smayling, "Cell-based aerial image analysis of design styles for 45 nanometer generation logic," Proc. of SPIE vol. 6521 (2007).
[8] V. Axelrad, A. Shibkov, G. Hill, H-J Lin, C. Tabery, D. White, V. Boksha, R. Thilmany, "A Novel Design-Process Optimization Technique Based on Self-Consistent Electrical Performance Evaluation", Proc. of SPIE vol. 5756 (2005).
[9] R. Pack, "Physical and timing verification of subwavelength-scale designs: I. Lithography impact of MOSFETs," Proc. of SPIE vol. 5042 (2003).
[10] K. Monahan and Brian Trafas, "Design and Process Limited Yield at the 65nm Node and Beyond," Proc. of SPIE vol. 5756 (2005).
[11] M. C. Smayling, C. Bencher, H. D. Chen, H. Dai, M. P. Duane, "APF pitch halving for 22nm logic cells using gridded design rules," Proc. of SPIE vol. 6925 (2008).
[12] Cell Designer User's Manual, SEQUOIA Design Systems, USA
[13] T. W. Houston, R. A. Soper, T. J. Aton, "Double pattern and etch of poly with hard mask," US Patent 6,787,469 (2004).

[14] http://www.tok.co.jp/en/tech/safier/index.html
[15] T. Sato, A. Endo, A. Mimotogi, K. Sato, S. Tanaka, "Impact of polarization on an attenuated phase shift mask with ArF hyper-numerical aperture lithography," J. Microlitho., Microfab., Microsyst., vol 5, 043001 (2006).
[16] T. Chen et al, "A Single-Exposure Approach for Patterning 45nm Flash/DRAM Contact Hole Mask," Proc. of SPIE vol. 6283 (2006).
[17] M. Yamabe, "Status and issues of electron projection lithography," J. Microlitho., Microfab., Microsyst., vol 4, 011005 (2005).
[18] T. Fühner, A. Erdmann, S. Seifert, "Direct optimization approach for lithographic process conditions," J.Micro/Nanolith. MEMS MOEMS vol 6, 031006 (2007).