Programmable Lithography Engine, ProLE™, Grid-Type Supercomputer and Its Applications

John S. Petersen, Mark John Maslow, David J. Gerold and Robert T. Greenway
Petersen Advanced Lithography, Inc., 12325 Hymeadow Drive, Suite 2-201, Austin, TX 78750 USA

(512) 241-1100/jpetersen@advlitho.com

Abstract

There are many variables that can affect lithographic dependent device yield. Because of this, it is not enough to make optical proximity corrections (OPC) based on the mask type, wavelength, lens, illumination-type and coherence. Resist chemistry and physics along with substrate, exposure, and all post-exposure processing must be considered too. Only a holistic approach to finding imaging solutions will accelerate yield and maximize performance. Since experiments are too costly in both time and money, accomplishing this takes massive amounts of accurate simulation capability. Our solution is to create a workbench that has a set of advanced user applications that utilize best-in-class simulator engines for solving litho-related DFM problems using distributive computing. Our product, ProLE™ (Programmable Lithography Engine), is an integrated system that combines Petersen Advanced Lithography Inc.'s (PAL's) proprietary applications and cluster management software wrapped around commercial software engines, along with optional commercial hardware and software. It uses the most rigorous lithography simulation engines to solve deep sub-wavelength imaging problems accurately and at speeds that are several orders of magnitude faster than current methods. Specifically, ProLE uses full vector thin-mask aerial image models or when needed, full across source 3D electromagnetic field simulation to make accurate aerial image predictions along with calibrated resist models;. The ProLE workstation from Petersen Advanced Lithography, Inc., is the first commercial product that makes it possible to do these intensive calculations at a fraction of a time previously available thus significantly reducing time to market for advance technology devices.

In this work, ProLE is introduced, through model comparison to show why vector imaging and rigorous resist models work better than other less rigorous models, then some applications of that use our distributive computing solution are shown. Topics covered describe why ProLE solutions are needed from an economic and technical aspect, a high level discussion of how the distributive system works, speed benchmarking, and finally, a brief survey of applications including advanced aberrations for lens sensitivity and flare studies, optical-proximity-correction for a bitcell and an application that will allow evaluation of the potential of a design to have systematic failures during fabrication.

Keywords: lithography, simulation, EDA, TCAD, RET, yield, systematic failure, DFM, PSM, GRID computing

Introduction

The ability of IC designers to simply send a completed layout to a fab and expect high production yields ended with the advent of the sub-wavelength era. As designs reached 130nm feature size and below, design and manufacturing complexity dramatically increased: the addition of non-printing features such as optical proximity correction (OPC) and phase shifters required new levels of simulation to insure functionality and yields. The gap between circuit layout and actual results on silicon grows exponentially as circuits shrink. **Semiconductor design costs are skyrocketing, from approximately $0.5M at 350nm feature size and $5M at 180nm, to $50M at 90nm for complex designs.**

Without new simulation tools:

- IC tapeouts will require more custom adjustment time to allow manufacturability, significantly increasing time to market
- New device yields will decline and take longer to ramp to acceptable levels
- Pressure to reduce design iterations will mount even more due to rising photomask costs, which now exceed $1M below the 130nm device node, and which may more than double as reticles become more specialized and use more process layers

As the industry transitions from 130nm, to 90nm and then 65nm device nodes, the geometric increases in complexity will drive computer modeling as a key tool for maximizing yield as early as possible and minimizing design cycle time and iterations. We feel that tomorrow's leadership in device density and the benefits it provides such as high computing power, low dissipated power, small footprint, and lower production

cost will be only attainable by those who exploit accurate and fast computer simulation to adjust circuit layout for actual silicon results *before* trying it in a Fab. Failure to develop more effective simulation will put Fab-less design firms at worsening disadvantage to integrated device manufacturers (IDMs). When a device requires many design iterations to develop acceptable yield, an IDM has advantage over a Fab-less producer because it controls access to critical Fab resources.

Further, unlike earlier technologies where random errors from particles dominated, systematic within-die failures hinder production device yield at the 130nm technology node. This yield loss is due to errors that arise from the convolution of the design and the process. Further these systematic failures grow in magnitude with each smaller device technology. To a significant extent, these failures are within chip, and are image and pattern transfer related mismatches to design at both the device-cell and the place-and-route portion of the design cycle. Therefore making lithography aware chip designs reduces *non-reoccurring engineering* (NRE) costs, and raises *average selling prices* (ASP) by getting product to market faster, with fewer physical product-learning cycles, lower associated reticle and material cost, and improved product performance. When it is considered that the design cost increases 36% every generation, excluding the cost of reticle sets which doubles every generation, and that the estimated costs to design a full logic chip at the 90nm node is $50M, it becomes imperative that NRE be reduced and ASP improved for companies to not only succeed in the sub-wavelength era but for fabless semiconductor companies, to even stay in business.[1]

Succeeding will take rigorous simulators that provide accurate predictions with few assumptions, about how a circuit needs to be imaged and what shape the circuit should take to provide the best ASP. There are simulation products that can be modified to help reduce systematic failures but their imaging models are largely based on Hopkins's scalar theory of aerial image formation and, due to vector effects that result in loss of interference efficiency from polarization in high numerical aperture imaging of small features, do not accurately comprehend image formation in the photoresist material. Further, the models fall short because they typically make the Kirchhoff approximation that assumes an infinitely thin opaque material and as a result, do not take into account how the electric field is affected by the topography and dielectric constants of the pattern on the reticle. Nor do these products comprehend the affects of the photoresist chemistry, such as reaction-diffusion, reactant/product loss and development to form the final three-dimensional image. Finally, no products take into account the post-imaging pattern transfer effects. The reason for this is that the problems are computationally daunting, requiring a supercomputer or equivalent to solve even the simplest problem in a timely manner. We have developed the Programmable Lithography Engine (ProLE™) product that allows us to overcome the computational constraints, but to also solve them economically.

For the use of Kirchhoff approximations of infinitely thin masks, simulators like PROLITH™ from KLA-Tencor's FINLE Division, Austin, Texas, have the rigor needed to provide accurate solutions of the resist and etch. When needed, these simulators can circumvent their scalar mask simulation inaccuracies by using full Maxwell electric field solvers like TEMPEST from UC Berkeley, commercially available from Panoramic Technologies, Berkeley, California. However, these simulators can only handle one relatively small simulation volume, with magnitude of 1000nm by 1000nm by 300nm, at a time. Therefore these simulators are ineffectual at solving full-scale imaging and design problems.

To remedy, we have developed a Programmable Lithography Engine (ProLE™) product that allows us to overcome the computational constraints of these commercially available lithography simulators by distributing unique simulation jobs across many computers, thus creating a "cluster" or "GRID" type supercomputer. A grid-supercomputer can leverage the core competencies of commercially available products, but with the cycle times to make it commercially viable to tackle large practical imaging problems. It is this system combined with a systematic approach to problem solving and model support services that make it possible to not only solve the inverse problem of creating the right mask to produce the desired image in silicon, see Figure 1, based on the most rigorous and numerically daunting simulators,[2, 3, 4, 5, 6, 7, 8] but to also provide the shape of the final etched structure on the wafer that can then be used to do parametric test and find systematic failures before building expensive reticles and product development wafers. With ProLE we believe that we can reduce design NRE costs by 10% to 20%, reduce retooling of reticles by 25% to 50%, pull in product release by one to two months, and improve the Average-Selling-Price by 10% to 20%. Beyond improved efficiency, it is reasonable that improved quantification of the process will allow improved design optimization and thus device performance, resulting in still more ASP increase.

In this work we will introduce ProLE, and use comparison to show why we need vector imaging and rigorous resist models. Then we will discuss some applications that use our distributive computing solution.

Integrated Imaging Systems

There are more than sixty-nine variables that can affect lithographic yield, and of those listed in the Figure 2 fishbone diagram, some like resist and dissolution chemistry are groupings of many other critical variables.[9] It is not enough to make optical proximity corrections (OPC) based on the mask type, wavelength, lens, illumination-type and coherence as indicated with ellipses on the fishbone. Resist chemistry and physics along with substrate, exposure, and all post-exposure processing must be considered too. Only a holistic approach to finding imaging solutions will accelerate yield and maximize performance. Since experiments are too costly in both time and money, accomplishing this takes massive amounts of accurate simulation capability. To gain accuracy, rigorous models with accurately measured and calibrated parameters are required along with the computational methodology to provide useful question-to-answer cycle times.

Image Process Integration is the process of taking all of these process variables from figure 2 into account and optimizing them with the device layout to build an integrated imaging system (IIS). This is what people have practiced either actively or passively for years and is the subject of many papers throughout the history of conferences that cover macro- and micro-imaging. But with the advent of imaging features smaller than the wavelength of the actinic energy, along with movement to two-beam imaging so as to extend the frequency of allowed imaging, the need for a formalized process of IIS-design within the infrastructure has become mandatory for extending production resolution and profitability at the same time.[10]

IIS formation is a large component of the design-for-manufacturing (DFM) domain and as such requires cross-disciplinary collaboration between diverse communities for it to be effective. DFM requires tools and knowledge from EDA (Electronic-Design-Automation), Litho-TCAD (Technical-Computer-Aided-Design), RET (Resolution-Extension-Technology) and the mask and wafer fabrication areas, see Figure 3. The large numerical aperture (NA), state-of-the-art exposure tool and the ability to use two-beam imaging is rapidly outpacing the capabilities of the mask fab and RET disciplines. The mask fab lacks the equipment and know-how to build masks that would fully utilize today's exposure tools. RET falls short because these models do not treat the energy in the diffraction pattern as a vector quantity, and so miss the phase-error induced loss of interference efficiency in the p-polarized electric field that arises at large diffraction angles.[11, 12, 13] This will be discussed later in the "Model Comparison." Further, the lithography simulators do not take into account the change in the phase error as the energy enters the resist nor the subsequent smudging of the recorded image with post processing and etch. This weakness is being worked on by the RET companies but the need is becoming more and more apparent for immense amounts of more rigorous lithography simulation from TCAD, simulation that takes all these factors and more into consideration. Furthermore this trend will continue as Moore's law drives minimum feature and pitch smaller and smaller. There will come a point where RET will require the same rigor as the simulators used in TCAD on at least a portion of each design. As discussed previously, EDA has problems too. The gap between expected performance and actual is growing with each jump to a smaller technology node. This is in part due to the growth of systematic failures and overly conservative guard-banding. Again TCAD can help here by taking a layout, correcting it for imaging, simulating how the pattern looks on the wafer under typical set of process variations, and outputting this shape for 2D- or 3D-parasitic extraction. This will help designers to identify failures prior to making expensive masks and wasting valuable fab and design cycle time. But, like RET, this requires immense amounts of computational power and finesse to make this a valuable tool. Enter ProLE.

Description of the Programmable Lithography Engine ProLE

Our solution is to create a workbench that has a set of advanced user applications for solving litho-related DFM problems using distributive computing. Our product, ProLE™ (Programmable Lithography Engine), is an integrated system that combines Petersen Advanced Lithography Inc.'s (PAL's) proprietary applications and cluster management software wrapped around commercial software engines, along with optional commercial hardware and software. It uses the most rigorous lithography simulations to solve deep sub-wavelength imaging problems accurately and at speeds several orders of magnitude faster than current methods. Currently, this system uses a workbench environment where the simulationist creates batch jobs for PROLITH 3D that are distributed across the cluster to be run by PROLITH-based ProLE engines. Once the

simulations are complete the results are sent back to the workbench for subsequent analysis with applications like ProDATA, Microsoft EXCEL or SAS Institute's JMP.

A conceptual description of the ProLE distributive computing system is shown in Figure 4 and in more detail in Figure 5. Using a streamlined GUI that reads a PROLITH *.pl2 file's parameter list, the interface allows selection of parameters for study. After selection, the simulation variable ranges and increments are defined, including input files from directories not resident in the PROLITH database such as mask files, aberration files, or feature measurement coordinates. Then sample plans are determined, for instance choosing to run the diagonals of some simulation matrix and not others and/or to use Monte Carlo methodology for process sensitivity studies. Next, the workbench interface is used to define compute optimization, how the jobs are to be created, and how results are to be saved. Finally; jobs are made and posted in a directory for the cluster engines. In the cluster, the ProLE Engine is looking for work. When it sees a job it takes it and runs it.. Both workbench and engine are used to monitor and manage the cluster; in most applications each engine is its own master thus making it possible to average overhead across the cluster. This boosts efficiency of the overall system and makes it massively scalable in a GRID-type computing environment. As figure 5 shows, ProLE Engines continue to look for work until none are left. Then they close PROLITH and keep looking for work. The workbench collects and sorts the results as defined by the user. Security is maintained in proprietary fashion to insure that only workbench-defined jobs are run and no cannibalizing of an engine's license for single user operation is permitted. In this way PAL is able to make ProLE run its applications as if it had a single, supercomputing-fast, copy of the simulator at a cost point that is reasonable to the marketplace while protecting the intellectual property of the commercial engines that PAL uses to power ProLE applications.

Model Comparison

Now we will briefly discuss the advantages of using ProLE to solve lithography problems. As mentioned during the DFM discussion, it has become important to use the most rigorous models possible for predicting a given response. In this section we will use PROLITH to compare the results of simulators of varying rigor in both aerial image formation and in resist image formation. Etch will not be discussed in this section.

In the comparison shown in Figure 6, we evaluated the process window of a 70nm binary mask line on a 250nm pitch imaged with a 0.85 numerical aperture exposure, 0.85 partial coherence and 193nm exposure. Three different PROLITH Kirchoff, thin-mask aerial image models were used to simulate this mask's aerial images: High NA Scalar, Full Scalar and, the most rigorous, Vector (unpolarized). The aerial image models vary as follows: High NA scalar is a solution that takes into account focus-into-the-resist, radiometric corrections, and the energy from the pattern's zero-diffraction-order convolved through the range of angles of the source NA. The full scalar model is the same but adds the angular affect of the higher diffraction orders, not only the zero-diffraction-order. The vector (unpolarized) model simulates the loss of parallelism of the coherently linked electric field vectors with the angular distance of the diffraction pattern from the center of the lens. For more detail see Chris Mack's book Inside PROLITH™: A Comprehensive Guide to Optical Lithography Simulation, pub: FINLE Technologies, Austin, TX, USA (1997).[13] The linewidth through focus-exposure response was compared for each aerial image model against the following models: (a) aerial image threshold model; (b) threshold resist model response for a PROLITH lumped parameter model using very high contrast and very thin resist with no acid diffusion; (c) A lumped parameter model tuned to behave like the full resist and imaging models; (d) full chemically-amplified positive resist models (the most rigorous).

Now, before getting into any detail, note that the process windows for (a) form smile shapes instead of the frowns of (b), (c) and (d). This is because purely sampling the aerial image does not take into account any of the imaging biases associated with the resist imaging process, and as a result, poorly reflects the process. Because they don't reflect how the aerial image is sampled by the resist as the image goes through variations in focus, IIS decisions made using only the aerial image are doomed to failure. Within (a) (as well as in (b) the vector (unpolarized) curve), the plot does not overlay the other two models that do overlay each other. Also note that between (a) and (b) the working depth of focus (DoF) appears slightly better for the vector model, possibly as a result of lower contrast image sampling of the aerial image closer to the isofocal point. Now moving on to (c), in this curve the interference effects were not taken into account because a fixed index was assumed between the substrate and the resist. Also, focus-into-the-resist was not taken into account. Because of these model assumptions and because of the isolated nature of the feature, the responses for all three models are similar, but most similar with the two more rigorous models. Finally, in (d) significant differences are observed. In this vector unpolarized model graph, as in (c), the feature sizes at slightly lower dose, suggesting a

lower contrast image. Furthermore in (d), both the full scalar and the vector models show a different center of focus than the high NA scalar model, because focus-into-the-resist and full diffraction pattern thin-film interference are taken into account. If we were to change pitch we would see that the center of focus would change too. This complication means it is imperative to properly emulate the resist patterning step to make proper choices about how to form the desired image.[14]

The next three figures examine optical proximity corrections based on the resist models in figure 6 for the three aerial image models and their predictions for different imaging situations. In each case where percent error is given it is the difference between the prediction made by the reference and the lesser models. E.G, if both models predict 10nm correction necessary, than the error of prediction would be 0.0%. The reference was taken to be the full resist/vector (unpolarized) model.

Figure 7 shows the correction applied to a 70nm line on a 250nm pitch so that it sizes at the same dose as a 70nm line on a 1000nm pitch with 20nm scattering bars. With this pattern, the threshold resist model showed consistently the worst predictions of around 60% compared to reference correction of -7nm. The simplified resist model predictions varied from –21%, -71% and –43% for high NA, full scalar and vector image models, respectively. The full resist model without the best images models showed errors too: 43% for high NA and 14% for full scalar.

In Figure 8, predictions for correction of a 70nm chromeless phase lithography (CPL) pattern are made. In this example the dose was optimized for a 70nm CPL line on 1000nm pitch, with half-toned chrome assist bars having an effective 20nm width. Bias was then applied to a 70nm Binary line on 200nm pitch. Again the threshold resist model consistently made the worst predictions of 33%, 43% and 48% error compared to the 10.5nm correction predicted by the reference model. Next, for this pattern the simplified resist model varied from reference correction by –14%, 14% and –5% for high NA, full, and vector image models. And while the full resist model is 14% off for both image models from the reference correction, it is still in error by +1.5nm.

Figure 9 looks at line-end shortening of 100nm butted lines with a 150nm gap, where the dose was calibrated to linewidth and mask gap space was varied to achieve 150nm gap on wafer. Dimensions listed are the required 1x gap widths. In this case the threshold model, as in the other examples, shows an extreme over-correction relative to the other resist models. The other two resist models were similar to each other, with the full resist being the best when non-vector-aerial image models are considered.

To summarize figures 7, 8 and 9, all combinations of resist and aerial image models show error from the reference. However, except for inaccuracy of the threshold model, there are no apparent trends in the error. This is because there are so many variables at play, and unless all the chemistry is understood the best matches will be essentially acceptable point solutions making it unlikely to extrapolate to the set of features found in a chip from a limited set of test cases. In addition, figures 7-9 were at a single focus, the method widely used today for OPC and RET decisions. Most users don't know where they are in the focus-expo process window given one focus point based on one simulation and over-simplified models. This situation is untenable from an engineering standpoint, and is unnecessarily risky for making multi-million dollar OPC and RET decisions.

Speed Benchmarking

In terms of computational power, in theory we are only limited by the number of engines at work solving a given problem. While this is generally true in practice, it is rare for multiple computers to run a job 100% as efficient as running it on a single CPU. The trick is to be efficient enough to make it cost effective to split jobs and run them. In this section, speed and efficiency benchmarks for two kinds of simulations were determined. The first case is where ProLE distributes a set of simulation variables so that they are run in subsets that are optimized for using the PROLITH engine most effectively. In this situation the individual engines across the cluster had 90.5% efficiency relative to a single CPU running the same job. In another case, where instead of simply varying PROLITH settings, we loaded multiple custom mask files, the forty mask simulation completed with an efficiency of 96%. In this instance the simulation area was 200nm to 520nm on a side depending on the mask parameters; while the simulation areas change the simulation volume was kept constant with x- and y-grid=300 sim-grid points and z-grid= 26 for a simulation volume of 2.34 X 10^6 sim-points. In this instance, one engine finished the mask set in 5999 seconds and eight engines ran the same set in 780seconds, effectively 7.7 times faster than the single computer.

Based on these types of efficiencies and on results of a 1,200 individual mask-screening study we did using PROLITH v7.1.1 with a sixteen unit 1GHz cluster of Pentium™ III's, we found that the ProLE workstation could solve a fifty mask layout problem with a total 1,256,850 unique focus-exposure simulations

in 5.03 days when using x- and y-sim-grid= 300 sim-grid points and z-grid=100 for a simulation volume of 9 X 10^6 sim-grid points.

Then more recently during a beta-test of ProLE with the newer PROLITH v7.2.2, we ran 11,388 individual contact hole masks with 17 focus X 41 exposure conditions for each mask for a total of 7,937,436 individual simulations. The simulation volume for this work was the same as in the previous 1,200 mask screening of 9 X 10^6 sim-grid points (x- and y-sim-grid=300 and z-grid=100). We did the test two times, first running as stated and then using a prescreening approach. With seven 700MHz engines and nine 1000MHz engines the job took 14.5 days. The raw uncorrected speed of v7.2.2 was 2.2 times faster and, corrected for MHz speed, 2.5 times faster than with v7.1.1. Then we again ran the contact hole study using the prescreening technique. This time we did the work in less than fourteen hours by using a limited focus-exposure matrix, semi-automatically sorting the "working" designs from the non-working designs, where "working" means they exhibited acceptable focus-exposure process windows, and then finishing by running the good 169 masks with the 17 focus X 41 exposure conditions.[15]

ProLE Applications and Uses

With the compute power of ProLE, we are developing applications that would not have been cost effective in the past. We are currently developing products that can be used as Litho-TCAD, RET, EDA tools and are planning applications for mask and wafer fab.

For TCAD we use Monte Carlo methods for doing aberration studies using 136 Zernike polynomial terms. This is useful for doing lens sensitivity,[16] flare studies[17] and aberration parameter extraction.[18] Typical responses in these types of studies are how features change shape and position with varying aberrations, illumination conditions and with changes in focus-exposure, but it is certainly not limited to just those variables. Figure 10 and 11 shows screen shots of the advanced aberration application. Figure 10 shows the ProLE TCAD application setup screens for aberration studies. The window on the left is the ProLE parameter selection screen and the screen on the left is the variable set-up screen. For this example four different Zernike terms were selected to be varied Z5 (upper left pupil map), Z46 (lower left), Z71 (upper right) and Z129 (lower right). Figure 11 shows the ProLE TCAD application aberration Selection Screen. The pupil map of Z5 is on the upper right corner and the horizontal gray line indicates its position in the selection table. The five columns of the table are from left to right: term selection, fringe term name, aberration type, normalization factor and formula. The table is laid out based on term order and frequency. On the left are, again, the pupil maps for Z46, Z71 and Z129.

For RET, figure 1 discussed briefly solving the inverse problem for developing OPC. In practice, we first determine what needs to be corrected with the help of the customer. A process audit determines the physical chemical attributes of the resist chemistry by studying deprotection chemistry relative to dose and PEB time and temperature, as well as the resist's dissolution properties. Then we set ProLE up to generate focus-exposure matrices for multiple sample points for subsequent common corridor and overlay analysis using ProDATA.

The results to follow were previously published.[8] In Figure 12, the original design was non-manufacturable as shown by no overlapping area in the process windows of metrology samples of various critical places in the SRAM bitcell. No combination of process settings would make this design work in manufacturing. The design, although it was decorated with OPC, didn't account for lithography effects that PAL later predicted and corrected for. After the initial simulation work shown in figure 12, mask files were generated using varying OPC decorations of the bitcells. As shown in figure 13, ProLE simulation found a solution by screening the features of each mask at sixteen different measurement locations. Doing this, we were able to optimize the design for manufacturability. Note the large area of overlapping process windows of 0.8 µm DoF with 10% exposure latitude, giving a manufacturing-capable process region. But we were not yet finished. In Figure 14 we show that we checked further for the stability and the quality of our solution as it interacted with other device layers across varying combinations of focus, exposure and alignment. Doing this helps insure that the answers determined in a given optimization or OPC validation will also minimize the occurrence of systematic failures due to improper correction of the features. As for the results of this work, in side-by-side comparisons the PAL-corrected bitcell yielded 60% versus zero for the uncorrected control, and versus 20% for the bitcell corrected by someone other than PAL. In addition, production was pulled in three months and no mask revisions were needed.

Today, we use ProDATA with responses like CD displacement and critical-shape-error, but we envision other applications such as shown in Figure 15. In this figure, we start with GDSII active and poly

design layouts, convert the poly layer to a PROLITH mask file (*.msk) using SLAM, import it to PROLITH, simulate imaging and etch, and then export the processed image converted to a GDSII file (again using SLAM) for subsequent analysis by a third-party parasitic extraction tool or for overlay and shape analysis in ProDATA. Figure 15 shows the GDSII layout of the original design, the post-develop resist image, and post-etch poly pattern referenced to the ideal active layer.

To effectively study systematic device failures due to real-world distributions of process inputs, Monte Carlo techniques are especially useful. For this reason, ProLE is designed to easily vary process inputs according to desired statistical distributions.

Other applications that we are developing will help lithographers study line-edge-roughness by using Monte Carlo techniques to generate varying linewidths along a feature edge or for doing most other process sensitivity studies. To our advanced aberration application we are adding chromatic aberrations.[19] Another application is to generate scatterometric and overlay targets for subsequent electromagnetic field simulation of broadband exposure sources so that researchers can quickly study the effects of placement errors and shape changes. These same EMF techniques can be used for studying deep-subwavelength and EUV imaging or for doing other topography studies.

Summary and Conclusion

The cost of designing and taking to market an advanced semiconductor product is becoming prohibitively expensive. Delaying delivery of production quantities of a new chip can cost millions of dollars in lost margins. Having to retool mask sets can cost more than $1 million at the 130nm technology node and beyond. Lost productivity of a $2 billion factory can cost in the neighborhood of $1 million per day, and use of these facilities, as laboratories must be reduced to a critical minimum. When considering that it costs each integrated-device-manufacturer (IDM) and foundry $300 to $500 million to develop a new manufacturing technology-node, and in addition that a state-of-the-art chip like a microprocessor at the 130nm node costs $18 million to design and at the 90nm node costs around $50 million to design with a significant portion being non-recurring-engineering costs (NRE), it becomes imperative that product gets designed so that it not only achieves yield entitlement quickly but does so with the desired performance specifications.

To further complicate things, new yield issues have emerged at sub-180nm process technologies: Performance problems are driven by interconnect wiring efficiency and structure, and gate uniformity and balance. Marginalities give rise to systematic failure where yield loss is no longer driven by random defects but from design and process interactions, of which a significant portion can be traced to litho. Because of this, more conservative design practices are employed to insure greater probability of success; therefore designers are conservatively guard-banding because of their lack of understanding of the inherent process.

In reality, beyond the fact that the full power of a process technology isn't being exploited because of excessive guard-banding, there are still manufacturing issues that the current set of EDA tools do not adequately comprehend because of today's increasing process complexity. This means the design community is not fully exploiting the technological capabilities of the IC factory and thus they are not reaping the return-on-investment they otherwise could.

With intelligent simulation setup and sufficient compute power, we can use the rigorous vector and full electromagnetic field models required for considering layout, illuminator, mask, resist and pattern transfer in combination to accurately predict how a device pattern is going to image, and whether it will show systematic failure or not.

Our solution is to create a workbench that has a set of advanced user applications that utilize best-in-class simulator engines for solving litho-related DFM problems using distributive computing. Our product, ProLE™ (Programmable Lithography Engine), is an integrated system that combines PAL's proprietary application and cluster computing management software with optional commercial hardware and software. It uses the most rigorous lithography simulations to solve deep sub-wavelength imaging problems accurately and at speeds several orders of magnitude faster than current methods. We believe our solution, and solutions like it, are required to design and produce integrated circuit products that make companies profitable in the era of deep subwavelength lithography.

Acknowledgements

Thanks to Will Conley of Motorola for the beta-test simulation masks.[15] We would also like to thank Ken Rygler of Rygler Associates and Mark Craig of HPL/TestChip for many useful discussions. Finally, we would like to thank KLA-Tencor, especially Jeff Byers, Mark Smith, Chris Mack, Chris Sallee and Ed Charrier.

FIGURES

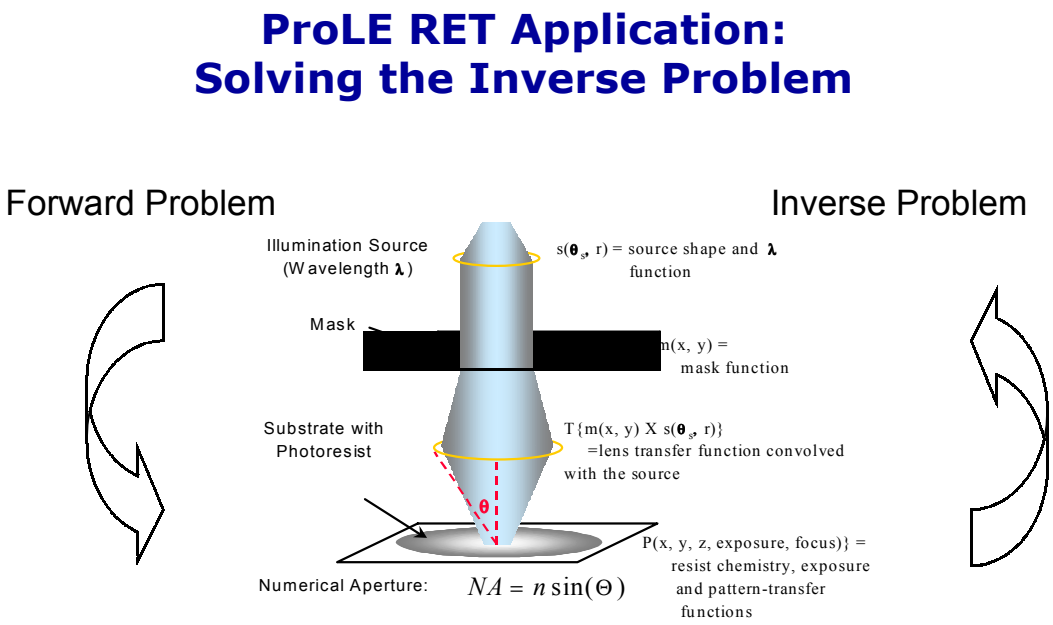


Figure 1: The final image function, f(x, y, z) is a function of the source, mask, lens, exposure-focus condition and resist chemistry and pattern transfer function. To solve the inverse problem do an error analysis using the proper propagation of error methodology to minimize the difference between the final image and the ideal image:

error function = F{f(x, y, z)Actual- f(x,y,z) Ideal}

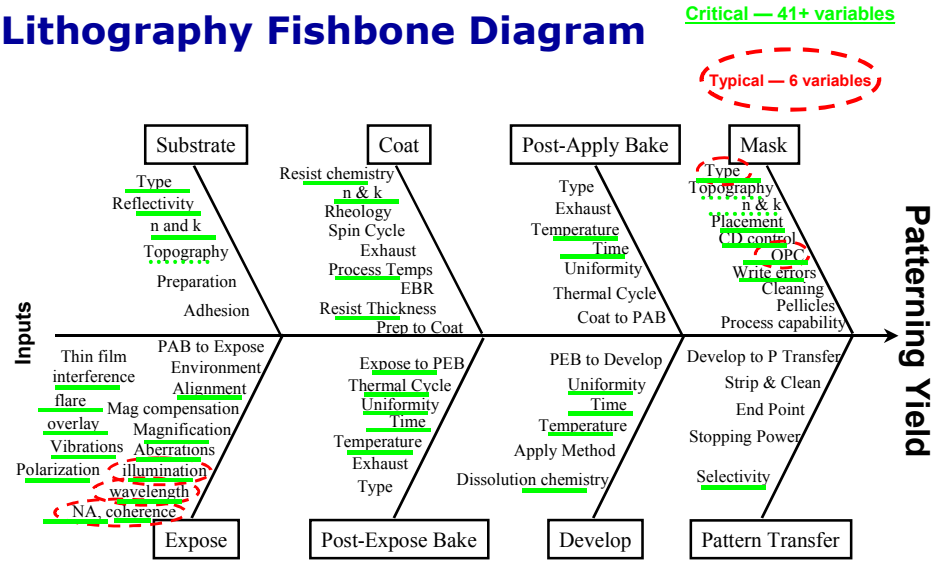


Figure 2: There are more than sixty-nine variables that can affect lithographic yield, some like resist and dissolution chemistry are groupings of many other critical variables.

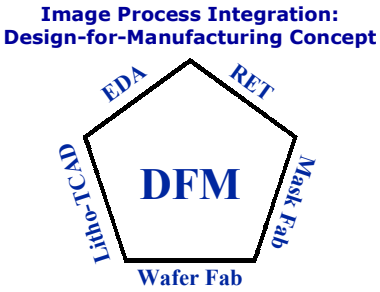


Figure 3: IIS formation is a large component of the design-for-manufacturing (DFM) domain and as such requires cross-disciplinary collaboration between diverse communities for it to be effective. DFM requires tools and knowledge from EDA (Electronic-Design-Automation), Litho-TCAD (Technical-Computer-Aided-Design), RET (Resolution-Extension-Technology) and the mask and wafer fabrication areas.

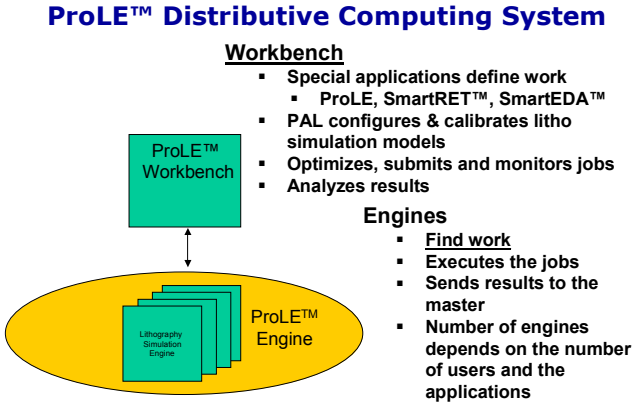


Figure 4: A conceptual description of the ProLE distributive computing system.

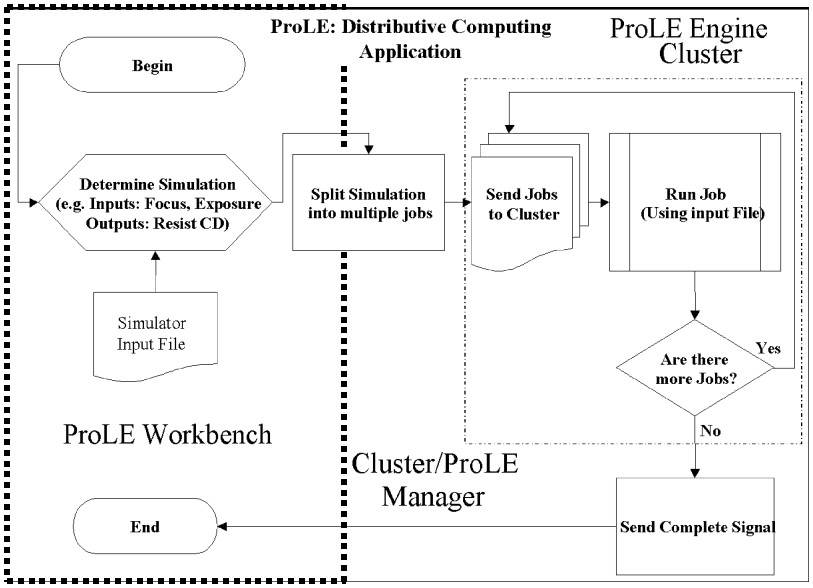


Figure 5: ProLE process flow schematic.

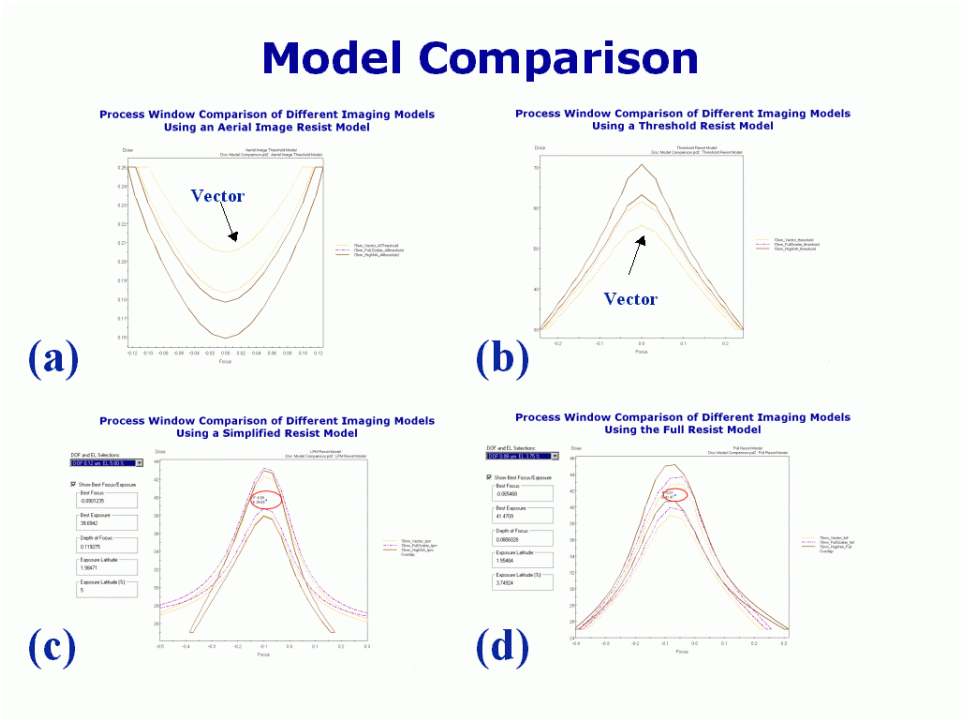


Figure 6: The process window of a 70nm binary mask line on a 250nm pitch imaged with a 0.85 numerical aperture exposure, 0.85 partial coherence and 193nm exposure and three different PROLITH aerial image models, High NA Scalar, Full Scalar and, the most rigorous, Vector (unpolarized). The linewidth response based on (a) aerial image threshold model; (b) threshold resist model response for a PROLITH lumped parameter model using very high contrast and very thin resist with no acid diffusion; (c) A lumped parameter that had been tuned to behave like the full resist and imaging models; (d) full chemical amplified positive resist models (the most rigorous).

	High NA Scalar Image Model	Full Scalar Image Model	Full Vector Image Model	Ref for Fig 6
Threshold Resist Model	64.29% -11.5nm (58.5nm)	64.29% -11.5nm (58.5nm)	57.14% -11nm (59nm)	(b)
Simplified Resist Model	-21.43% -5.5nm (64.5nm)	-71.4% -2nm (68nm)	-42.9% -4nm (66nm)	(c)
Full Resist Model	42.9% -10nm (50nm)	14.29% -8nm (62nm)	Reference -7nm (63nm)	(d)

Figure 7: Shows the correction applied to a 70nm line on a 250nm pitch so that it sizes at the same dose as a 70nm line on a 1000nm pitch with 20nm scattering bars. (0.67nm=x-grid simulation)

	High NA Scalar Image Model	Full Scalar Image Model	Full Vector Image Model	Ref for Fig 6
Threshold Resist Model	33% +14nm (84nm)	42.9% +15nm (85nm)	47.6% +15.5nm (85.5nm)	(b)
Simplified Resist Model	-14.3% +9nm (79nm)	14.3% +12nm (82nm)	-4.76% +10nm (80nm)	(c)
Full Resist Model	14.3% +12nm (82nm)	14.3% +12nm (82nm)	Reference +10.5nm (80.5nm)	(d)

Figure 8: Shows predictions for correction of a 70nm chromeless phase lithography (CPL) pattern are made. In this example the dose was optimized for a 70nm CPL line on 1000nm pitch, with half-toned chrome assist bars having an effective 20nm width. Bias was then applied to a 70nm Binary line on 200nm pitch. (0.67nm=x-grid simulation)

	High NA Scalar Image Model	Full Scalar Image Model	Full Vector Image Model	Ref for Fig 6
Threshold Resist Model	-9nm (73nm)	-13nm (69nm)	-11nm (71nm)	(b)
Simplified Resist Model	+4nm (86nm)	+2nm (84nm)	+4nm (86nm)	(c)
Full Resist Model	+2nm (84nm)	-1nm (81nm)	Reference (82nm)	(d)

Figure 9: Looks at line-end shortening of 100nm butted lines with a 150nm gap, where the dose was calibrated to linewidth and mask gap space was varied to achieve 150nm gap on wafer. Dimensions listed are the required 1x gap widths. (0.67nm=x-grid simulation)

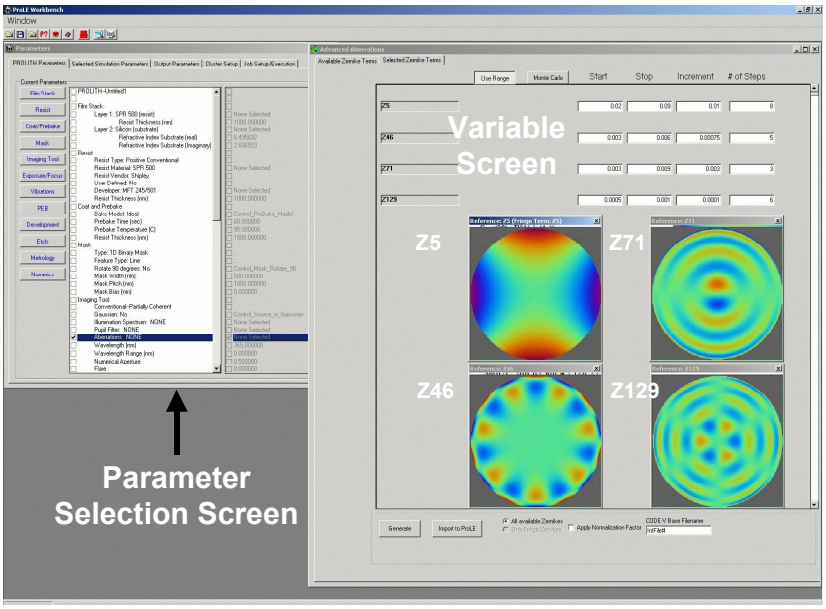


Figure 10: ProLE TCAD application showing setup screens for aberration studies. The window on the left is the ProLE parameter selection screen and the screen on the left is the variable set-up screen. For this example four different Zernike terms were selected to be varied Z5 (upper left pupil map), Z46 (lower left), Z71 (upper right) and Z129 (lower right).

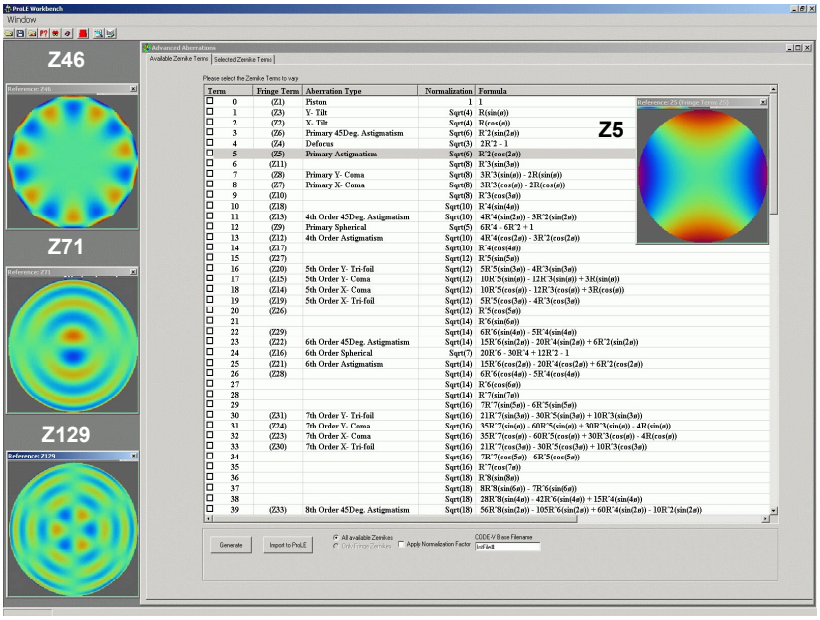
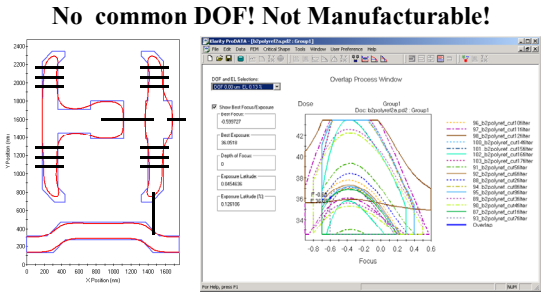
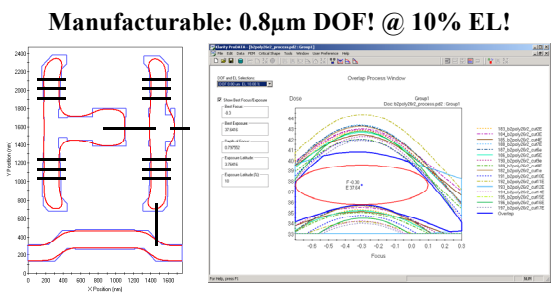


Figure 11: ProLE TCAD application showing the aberration Selection Screen. The pupil map of Z5 is on the upper right corner and the horizontal line indicates its position in the selection table. The five columns of the table are from left to right: term selection, term name, aberration type, normalization factor and formula. On the left are the pupil maps for Z46, Z71 and Z129.



Craig, Mark J., Petersen, John S., Lund, Joshua, Gerold, David J., Chen, Nien-Po, Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 380-389 (7/2002)

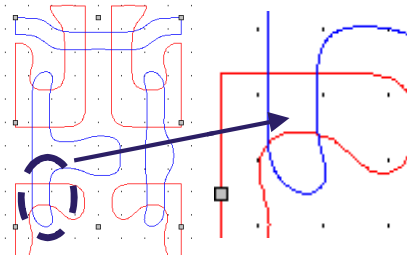
Figure 12: The original design was non-manufacturable as shown by no overlapping area in the individual feature process windows. No combination of process settings would make this design work in manufacturing. The design, although it was decorated with OPC, didn't account for lithography effects that PAL later predicted and corrected for.



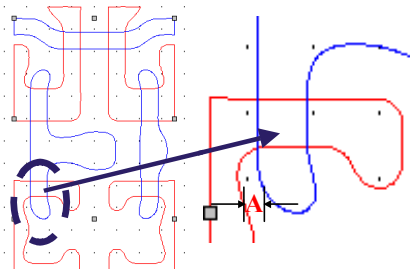
Craig, Mark J., Petersen, John S., Lund, Joshua, Gerold, David J., Chen, Nien-Po, Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 380-389 (7/2002)

Figure 13: After the initial simulation work shown in figure 12, mask files were generated using varying OPC decorations of the bitcells and ProLE simulation found a solution by screening the features of each mask at sixteen different measurement locations. Doing this, we were able to optimize the design for manufacturability. Note the large area of overlapping process windows of 0.8 µm DoF with 10% exposure latitude = manufacturing capable region.

Uncorrected Active/Corrected Poly
Low Yield and Poor performance



Corrected Active/Corrected Poly
Good Yield and Enhanced performance



•The best working designs will not have a curved active area (red) under the poly area (blue), so the uncorrected active on the left yields poorer performance than the one in the figure on the right.

•From an overlay standpoint the tip of the poly cannot fall off the field nor can it encroach on the active area that parallels it, area A.

Craig, Mark J., Petersen, John S., Lund, Joshua, Gerold, David J., Chen, Nien-Po, Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 380-389 (7/2002)

Figure 14: To get yield it is not enough to do OPC for one layer and another without checking how they interact across varying combinations of focus, exposure and alignment so to make sure that the answers determined also minimize the occurrence of systematic failures due to improper correction of the features.

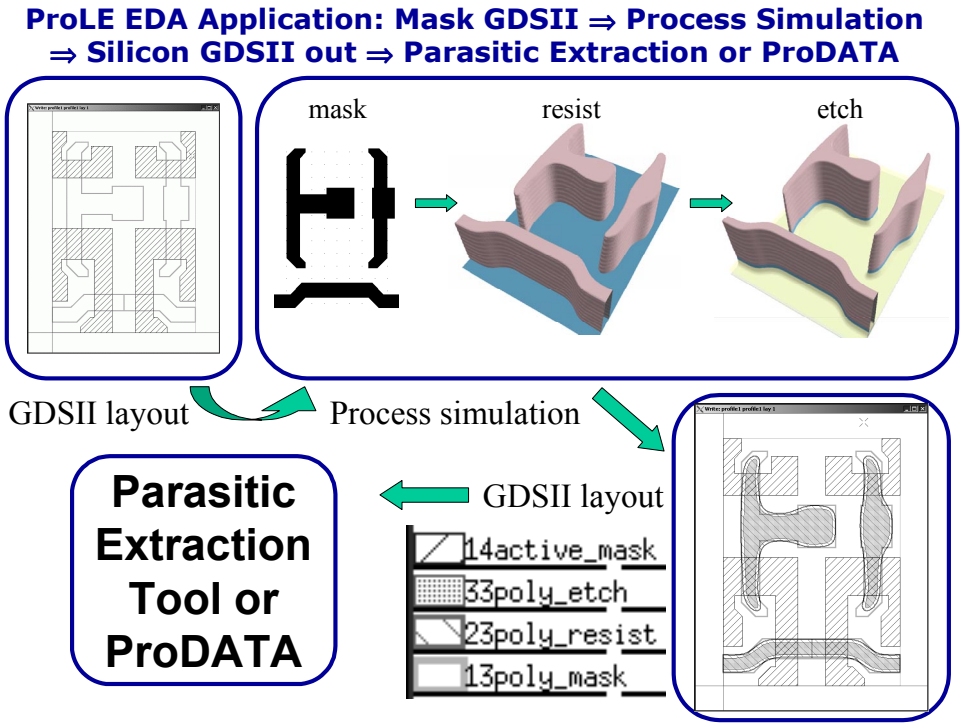


Figure 15. We start with a GDSII of a design layout, convert it a PROLITH mask file (*.msk), import it to PROLITH, simulate imaging and etch, then we export the processed image and convert it to a GDSII file for subsequent analysis by a third-party parasitic extraction tool or in ProDATA. In the GDSII layout of the original design, after develop and after etch patterns are shown referenced to the active layer.

References

1 Wilson, Ron, "Statistical techniques attack process variation", EE Times, (Mar 6, 2003 (12:03 PM)), URL: http://www.siliconstrategies.com/story/OEG20030306S0042.

[2] Petersen, John S., Beach, James V., Gerold, David J., Maslow, Mark J., "Developing an integrated imaging system for the 70-nm node using high numerical aperture ArF lithography", Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 274-287, (7/2002).

[3] Beach, James V., Petersen, John S., Maslow, Mark J., Gerold, Dave J., McCafferty, D. C., "Evaluation of SCAA mask technology as a pathway to the 65-nm node, SPIE conference 5040 [paper: 5040-107], to be presented 2/2003.

[4] Petersen, John S., Conley, Will, Roman, Bernard J., Litt, Lloyd C., Lucas, Kevin, Wu, Wei, Van Den Broeke, Douglas J., Chen, J. Fung, Laidig, Thomas L., Wampler, Kurt E., Gerold, David J., Socha, Robert J., van Praagh, Judith, Droste, Richard, "Development of a sub-100-nm integrated imaging system using chromeless phase-shifting imaging with very high NA KrF exposure and off-axis illumination", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen; Ed. p. 515-529 (7/2002).

[5] Petersen, John S., Conley, Will, Roman, Bernard J., Litt, Lloyd C., Lucas, Kevin, Wu, Wei, Van Den Broeke, Douglas J., Chen, J. Fung, Laidig, Thomas L., Wampler, Kurt E., Gerold, David J., Socha, Robert J., van Praagh, Judith, Droste, Richard, "Development of a sub-100nm integrated imaging system using chromeless phase-shifting imaging with very high NA KrF exposure and off-axis illumination", Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 298-311 (7/2002).

[6] Chen, Jang Fung, Inc.; Petersen, John S., Socha, Robert J., Laidig, Thomas L., Wampler, Kurt E., Nakagawa, Kent H., Hughes, Greg P., MacDonald, Susan S., Ng, Waiman, "Binary halftone chromeless PSM technology for quarter-lambda optical lithography", Proc. SPIE Vol. 4346 Optical Microlithography XIV, Christopher J. Progler; Ed., p. 515-533 (9/2001).
[7] Van Den Broeke, Douglas J., Chen, J. Fung, Laidig, Thomas L., Hsu, Stephen, Wampler, Kurt E, Socha, Robert J.; Petersen, John S., Petersen, "Complex 2D pattern lithography at lambda/4 resolution using chromeless phase lithography (CPL)", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen; Ed., p. 196-214 (7/2002).
[8] Craig, Mark J., Petersen, John S., Lund, Joshua, Gerold, David J., Chen, Nien-Po, "Robust methodology for state-of-the-art embedded SRAM bitcell design", Proc. SPIE Vol. 4692, Design, Process Integration, and Characterization for Microelectronics, Alexander Starikov; Kenneth W. Tobin; Eds., p. 380-389 (7/2002).
[9] Petersen, John S. and Byers Jeffrey D., "Examination of isolated and grouped feature bias in positive acting, chemically amplified resist systems", SPIE Vol. 2724, Advances in Resist Technology and Processing XIII, Roderick R. Kunz, Ed., p. 163-171 (1996).
[10] Petersen, John S., McCallum, Martin, Kachwala, Nishrin, Socha, Robert J., Chen, J. Fung, Laidig, Tom, Smith, Bruce W., "Assessment of a hypothetical roadmap that extends optical lithography through the 70nm technology node", SPIE Vol. 3546, 18th Annual Symposium on Photomask Technology and Management, Brian J. Grenon and Frank E. Abboud, Eds., p. 288-303 (1998).
[11] Brunner, Timothy A., Seong, Nakgeuon, Hinsberg, William D., Hoffnagle, John A. Houle, Frances A., Sanchez, Martha I., "High NA lithographic imagery at Brewster's angle", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen; Ed., p. 1-10 (7/2002).
[12] Smith, Bruce W. and Cashmore, Julian, "Challenges in high NA, polarization, and photoresists", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen, Ed., p. 11-24 (7/2002).
[13] Mack, Chris A., Inside PROLITH™: A Comprehensive Guide to Optical Lithography Simulation, pub: FINLE Technologies, Austin, TX, USA (1997). (Available from Petersen Advanced Lithography, Inc., sales@advlitho.com)
[14] Kachwala, Nishrin, Petersen, John S., Chen, J. Fung, Canjemi, Mike, McCallum, Martin, "Imaging contrast improvement for 160nm line features using sub resolution assist features with binary, 6% ternary attenuated phase shift masks with process tuned resist", SPIE Vol. 3679, Optical Microlithography, XII, Luc Van den hove; Ed., p. 55-67 (1999).
[15] Thanks to Will Conley of Motorola for the masks to simulate, results to be published later
[16] Progler, Christopher and Wheeler, Donald, "Optical lens specification from the user's perspective", Proc. SPIE Vol. 3334, Optical Microlithography XI, Luc Van den hove, Ed., p.256-268 (1998).
[17] Lai, Kafai, Gallatin, Gregg, Rosenbluth, Alan E., Fonseca, Carlos, Liebmann, Lars, Progler, Chris, "Methodology for generating exposure tool specification for alternating phase shift mask application for 70nm node", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen; Ed., p. 336-347 (7/2002).
[18] Smith, Bruce, Conley, Will, Garza, Cesar, Meute, Jeff, Miller, Daniel, Rich, Georgia, Graffenberg, Vicki, Dean, Kim, Patel, Shashi, Ford, Arnie, Foster, Jim, Moers, Marco, Cummings, Kevin, Webb, Jim, Dewa, Paul, Zerrade, Azzedine, McDonald, Susan, Hughes, Greg and Dirksen, Peter, "Aberration determination in early 157nm exposure systems", Proc. SPIE Vol. 4691, Optical Microlithography XV, Anthony Yen; Ed., p. 1635-1643 (7/2002).
[19] Kroyan, Armen, Bendik, Joseph, Semprez, Olivier, Farrar, Nigel, Rowan, Chris, Mack, Chris A., "Modeling the effects of excimer laser bandwidths on lithographic performance", SPIE Vol. 4000,. Christopher Progler Ed., p. 658- 664 (2000).