Optimization Criteria for SRAM Design - Lithography Contribution

Daniel C. Cole,[b] Orest Bula,[a] Edward W. Conrad,[a] Daniel Coops,[a] William C. Leipold,[a]
R. W. Mann,[a] and Jeffrey H. Oppold[a]
[a]IBM Microelectronics, Essex Junction, Vermont 05452
[b]Dept. of Manufacturing Engineering, 15 St. Mary's Street
Boston University, Boston, Massachusetts 02215

July 19, 2002

ABSTRACT

Here we discuss the use of well calibrated resist and etch bias models, in conjunction with a fast microlithographic aerial image simulator, to predict and "optimize" the printed shapes through all critical levels in a dense SRAM design. Our key emphasis here is on "optimization criteria," namely, having achieved good predictability for printability with lithography models, how to use this capability in conjunction with device and circuit design considerations, not just to achieve "best printability," but rather to achieve the combination of best electrical performance, yield, and density. The key lithography/design optimization issues discussed here are: (1) tightening of gate width variation by reducing spatial curvature in the source and drain regions, (2) achieving sufficient contact areas, (3) maximizing process window for overlay, (4) reducing leakage mechanisms by reducing contributions of stress and strain due to the printed shapes of oxide isolation regions, (5) examining topological differences in design during the optimization process, (6) accounting for mask corner rounding, and (7) designing for scaleability to smaller dimensions (here to 0.18 μm) to achieve optimal design reusability. The last item requires excellent lithography models for examining scaleability issues without hardware.

Lithography, simulation, SRAM, proximity correction, CAD, phenomenological model, photoresist model, optimization, microchip

1 INTRODUCTION

Specific aspects of an optimization study for an embedded dense random access memory (SRAM) design are presented here. These points will be used as a vehicle for a general discussion on the need for the merging of computer aided design (CAD) methods in microelectronics with process and technology simulation methods, particularly those in lithography. We believe that a new era has been, and is, emerging, namely, where these two disciplines become far more tightly coupled. The reasons for this merger are many-fold, but largely arise from a combination of improved phenomenological physical models and methods,[1-7] enormously improved numerical and computational methods,[8-11,3] faster computers with larger memories, as well as other technology enablers, combined with the recognized continued need for moving beyond the idealized CAD geometrical designs as the microelectronics industry forges on into yet smaller and smaller device sizes. As will be argued, considerable opportunity exists for the continued emphasis on developing methods to enable this merger. The specific study described here, on an advanced embedded, dense SRAM, illustrates some of the key advantages and issues that

can be addressed by these methods.

The SRAM optimization design work described here is certainly new. However, some of the key lithography concerns, such as the use of proximity corrections in the mask design, is certainly not a new topic for microlithographers. Indeed, work presented by technologists for the past several years at microlithography conferences, such as the annual Society for Photo Engineers (SPIE) on microlithography, have strongly emphasized this point, and some of the original ideas on this subject, such as controlling serif structures, goes back much farther in time.[12,13] Moreover, many of the CAD issues, such as the use of "design rules," are certainly not new to microchip designers. Nevertheless, we believe that much of the merger of these disciplines, particularly regarding the emphasis on what is important to optimize, is new, or at least very current and key to further significant advances in microelectronics.

The outline of our article is as follow. Section 2 provides a brief historical perspective on the areas of CAD and process simulation. Section 3 turns to a general discussion on what is typically optimized in these disciplines. Section 4 explains why the SRAM study is a good vehicle for exploring new lithography/design issues, while Sec. 5 describes the degree of success in applying the lithography and etch models in this study. Next, Sec. 6 turns to a discussion of the specific simulation steps followed in this study. Section 7 goes over the more detailed examples of what was important to concentrate on in this study in terms of optimizing the design and associated manufacturing processes. Finally, the article ends with some concluding remarks and recommendations for future developments in Sec. 8.

2 BRIEF HISTORICAL PERSPECTIVE ON TWO DISCIPLINES

2.1 Interactions

Microlithographers and microchip designers are increasingly becoming more familiar with each other's domain, but the gap between them remains large for leading edge technologies. There are a lot of reasons for this separation, with most of them being traceable back to the original practices in microlithography. Early on there was much less of a need, as compared with today, to worry too greatly about the process limitations resulting in corner rounding on patterns and deviations from Manhattan-like structures. Moreover, technological methods did not exist at the time to enable correcting such imperfections without resorting to the very expensive option of whole new generations of lithography tooling and materials. Consequently, microchip designers would design with straight-edge polygon patterns, and process technologists would have some requirements on the structures that they would need to meet, and there was little interaction beyond that.

As designs became more aggressive, thereby pushing the processing capabilities to greater demands, interactions between the groups became more common. Still, however, the interaction consisted solely of a set of written down "design rules" that designers could not violate, such as minimum spacing between structures in their rectilinear-like designs, or minimum area of intersection of the box-like structures drawn in overlapping design levels. Similarly, process developers were required to meet additional criteria, beyond the usual specification of being able to print a critical dimension within some specified process tolerance. A common practice appeared to develop in the industry, although it was called by different names within each company, namely, occasional meetings between designers and process technologists would take place to agree on possible "waivers" to design rules. Designers would typically push for loosening of the restrictions imposed on them, by these "design rules," since if such structures could be made, then faster circuitry performance could be obtained. In contrast, manufacturing technologists would tend to act more conservatively, since of course they were required to commit to being able to manufacture these smaller sized structures, and to ensure that these structures would hold up well under a fairly wide range of operating conditions. Occasionally waivers could be granted, since, for example, designers might be able to accept more variability on dimensions of some structures in exchange for smaller dimensions.

As one might imagine, particularly in this age of computers and high technology, such negotiations seem incredibly slow, and fairly imprecise. The imprecision has often arisen because requests for waivers are inevitably over design situations that have not yet been built. Without specific experiments run, one's ability is typically poor to gauge accurately whether a waiver is reasonable or not. Consequently, the decisions on most waivers have been, and continue to be, erring on the side of conservatism in design, simply because of this difficulty in being able to make a more precise call on the matter. Clearly, a strong need exists to improve this situation.

2.2 Tool Sets

The main tool set for designers consists of a large host of CAD software capabilities, including: (1) 2-D shape drawing tools, (2) the means for connecting these shapes easily using wiring programs ("place and route tools") that range in ability from manual to fully automatic in nature, (3) representing these structures in terms of their electronic components ("device extraction programs"), (4) running post-processor checking programs to flag design rule violations, as well as (5) a large set of circuit simulation programs, based on "compact models" for microelectronic devices and the largely rectilinear set of patterns of interconnecting circuitry.

In contrast, the process technologist's "tool set" has been far less oriented around simulation and computer methods. Instead, an organization might have some technology computer aided-design (TCAD) programs that are run, but such programs would normally simulate only a very small section of a circuit. Indeed, process and device TCAD simulation has typically concentrated on a cross-section of a single FET device, or, in lithography, on a cross-section of a set of lines in photoresist. Three-dimensional (3-D) process and device simulations, although certainly possible, and although certainly reported in the literature since as long ago as the early 1970s,[14] nevertheless were, and still are, extremely computationally and memory intensive. Lithography 3-D simulation capabilities have had an even later history.[15−18]

Consequently, less computationally intensive phenomenological methods have been pursued in recent years,[1−7] with the aim of merging their predictions with CAD methods, rather than concentrating on more traditional lithography predictions involving things like photoresist sidewall structures. Indeed, progress has been made, and is expected to progress further, at not only predicting the printing of photoresist structures, but also the etching of the underlying semiconductor materials. These models now seem sufficiently accurate[7] and fast[8−11,3] that they can readily fit in with more traditional CAD design methods, at least in regard to directly checking the integrity of design rules. Typical agreement between simulation and measurement, through the etch process beyond photoresist formation, can now be achieved to 10 to 40 nm on all points of nominal 0.25 μm 2-D shapes, including defocus conditions.[7]

At the very least, these simulation methods can be used in "waiver negotiations" between designers and manufacturing engineers to help better pinpoint whether waivers should be allowed. On a more aggressive scale, design rules may actually be directly based on the simulation of various test cases, or, yet in the most aggressive method, design rules may simply be replaced by the simulations themselves, where the "coding of the rules" is actually now viewed as being directly encoded within the calibrated phenomenological models that predict how the structures will print.

For well controlled processes and calibrated models, certainly these methods go well beyond the accuracy of previously used simple design rules (for example, see Sec. 5), which can't possibly take into account the myriad proximity conditions that can occur in an aggressive design. Consequently, designers can now push their layouts to within the limits of the model predictions; manufacturing engineers can also feel more confident that "waivers allowed" are legitimate, and will safely yield manufacturable structures. Of course, process windows should still be checked and verified, such as the printing of these structures for out of focus conditions.

Thus, the "tool sets" of CAD designers and lithographers are now beginning to merge (a number of examples are presented in Sec. 7), enabling communications to open up well beyond the simple waiver negotiations regarding

the typical opposing demands on tighter or looser design rules.

3 GENERAL DISCUSSION ON OPTIMIZATION

We can now turn to the next level of discussion that really has not been possible previously, or, at least has only been possible within very limited domains: namely, aiming to optimize on more general conditions than what the separate camps of microchip circuit designers and semiconductor technology processors traditionally pursue. As will be described here and in subsequent sections, we believe that this area is only beginning to open up, and will probably be an active area of research and development for many years to come.

Traditionally, microchip CAD designers' main concern has been circuit speed and reliability, in the case of logic designs, and maximum memory and access speed in the case of memory chips. Speed can typically be improved by making smaller circuit layouts and reducing signal delays along conduction paths, while ensuring that the overall integrated set of electrical signals still work in proper conjunction and sequence with each other. A large number of techniques exist to accomplish these aims, which is one area that has been very aggressively pursued within the microchip CAD community. Indeed, many doctoral theses each year are written on sophisticated means for automatically, and semi-automatically, optimally laying out CAD designs for microelectronic circuits. In addition, reliability issues, such as due to electrical noise, are an important aim of a microchip designer, to ensure that adequate ratios of signal level to noise are maintained during operation.

In contrast, traditionally process technologists have largely concentrated on jointly minimizing semiconductor structure sizes while maintaining or maximizing to whatever extent is possible, the process windows associated with making these structures. For the manufacturing engineer, the main aim is to maximize the "yield" on the number of chips that are produced that meet agreed upon circuit performance goals, and that do not fail over prescribed operating conditions that aim to stress the chip's capabilities (temperature ranges, high and low voltage operating conditions, electrical noise introductions, etc.). Typically each set of process steps, including the steps of lithography, etching, deposition, oxidation, ion implantation, and diffusion, has its own set of agreed upon process windows that must be met; the development technologists for each of these process steps aim to maximize their individual process windows, with the technology integrators attempting to coordinate and arrive at the needed overall process window to achieve the required "yield."

Thus, although the goals of the two groups are not totally opposed to each other, in general, they are quite different, and, yes, are often opposed to each other. The fast, phenomenological process models earlier discussed[7] will help to improve the communication between the groups, and may actually become a main focus of communication.

However, the industry needs to go beyond even this level of interaction. Specifically, each group needs to become more familiar with the goals of the other groups, and more combined and integrated goals need to be obtained. Some of the examples in Sec. 7 illustrate these points and, we think, begin to show some indication of more general goals, which are a combination of factors. Yes, overall circuit speed is desired, but so also is the maximum percentage of "good chips" made from each semiconductor wafer. Both factors enter into the profit that a semiconductor manufacturer can obtain, since the faster the circuit speed, the more that can be charged for the chip, while the higher the yield, then the more chips that can be produced for the same number of process steps.

Tools that integrate and help predict the result of these goals can only help to provide more common and uniform aims. Indeed, the term, "design for manufacturability"[19] has certainly become more common in recent years, which seems to very well represent the point of view being expressed here.

4 SRAM - IDEAL VEHICLE FOR TESTING MERGER OF DISCIPLINES

The example we will use to better illustrate much of the qualitative discussion in the earlier sections is work that involved the close interaction of designers and technologists over a relatively short period of intense work, followed by less intensive, but longer range interactions. An SRAM design seems, at this point, nearly ideal for this sort of interaction; typically an SRAM is a high volume part, and so requires improved techniques on combining methods for optimizing the design and processing capabilities simultaneously, beyond the simple expedient of "design rules."

This study involved an embedded SRAM design, so that the processing steps for the implants, heat cycles, oxidation steps, etc., all were dictated by the already existing logic process into which this dense SRAM structure was being embedded. A second goal of this project was that despite the processing step limitations, the aim was to make the dense SRAM structure have smaller critical dimensions than the surrounding circuitry, since it was reasoned that the extra proximity corrections being implemented on this smaller region, should enable this goal to be achieved.

In this study, six distinct designs were submitted by designers for various proximity correction analyses. They ranged in area from 12.0 μm^2 down to 6.0 μm^2. Here we note that many of these designs were topologically quite different from each other, meaning, that one could not in general simply move and smoothly change single shapes to evolve naturally from one cell design to another. Rather, just as a doughnut and a sphere are topologically distinct in nature, so were some of these designs, thereby preventing the single treatment of one design by advanced, but practiced, optimization methods to cover the entire set of designs.

Each cell had different attributes and disadvantages, ranging of course from the size of the cell, to the potential yield that might be expected for manufacturing the cells. The simulation work was intended to aid in the selection of the best cell candidate for manufacturing, after of course, all potential proximity corrections and area intersection and overlay improvements were made to each cell.

For each design, four critical levels were treated to the proximity correction, overlay, and intersection analysis that will be described here. These mask levels were those used for patterning the following SRAM levels: (1) gate, (2) diffusion, (3) first contact, and (4) first metal interconnect levels. Three different resist types, as well as both positive and negative masks, were used in the lithography steps. The calibration methods discussed in Ref. 7 were used for characterizing each of these processes. For two of the levels, namely, the gate and first metal levels, simulations were carried out using both Micrascan II and III in order to ascertain the potential benefit of moving these levels to Micrascan III in the manufacturing line. Both in-focus and a range of out of focus conditions were examined for the printability of each mask level. All critical dimensions considered ranged in value from 0.22 μm to 0.305 μm, for the photoresist printed structures. A constant etch bias was used as the approximate model for the difference in underlying etched structure from the printed photoresist structures. As discussed in Ref. 7, although this very rough model worked reasonably well for us, it is certainly possible for much more sophisticated treatment of this effect. Nevertheless, four different constant etch biases were taken into account in this work, as characterized for four different etch process conditions; these biases ranged from -0.0175 μm to $+0.02$ μm.

Thus, as will be noted from (1) the multiple process conditions needing to be examined and taken into account, (2) the large number of possible SRAM designs needing to be considered, and (3) the aim to maximize the process windows associated with any one cell, there was a sufficient multitude of conditions to examine that the SRAM turned out to be an excellent test of the potential of this modeling approach.

5 AGREEMENT BETWEEN EXPERIMENT AND SIMULATION

Reference 7 contains many more details on the issues of calibration, metrology, and general fits to experiment that can be expected from the fast phenomenological photolithographic models described there. The models described in Secs. 3.4 and 4 in Ref. 7 were precisely the ones used here. These phenomenological models seemed sufficient for describing the general behavior of each of these processes, although at the time of this application work, we had considerably more SEM data to calibrate the models for UV2HS than for the other two resist processes used. Nevertheless, it should be noted, 2-D characterization from even one sample is actually fairly powerful, certainly far more so than the same situation for a 1-D characterization from a single sample, simply because of the large number of points involved in the fitting procedure for the 2-D versus the 1-D case.

In addition, the mask fabrication process had been well characterized in terms of knowing what were the smallest sized serif-like features that could be fabricated, and how the mask structures differed from the intended design structures.[20] Figure 1 gives an indication of the importance for making this characterization in the mask fabrication process. Mask corner rounding was not taken into account directly in this particular set of simulation runs, although in subsequent work since then, these considerations have become more standard in the simulation work.

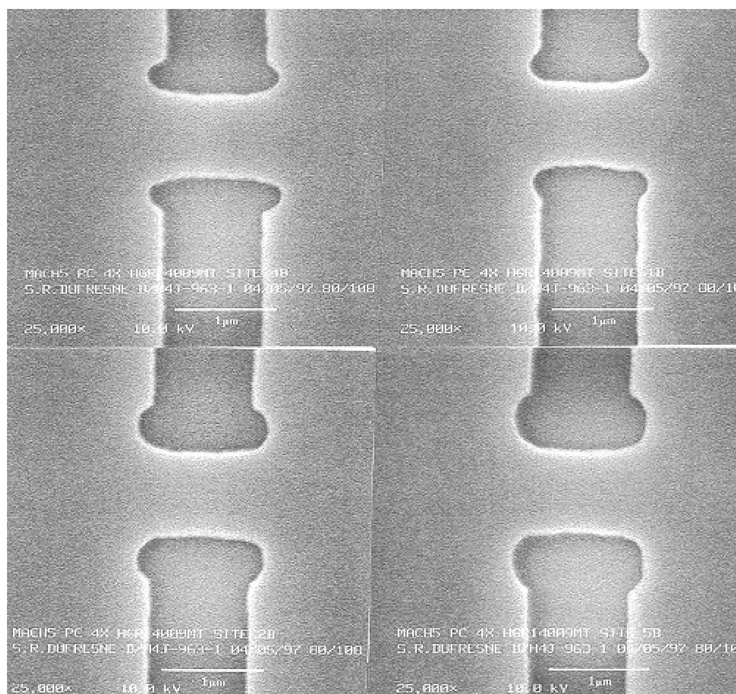

Figure 1: Four SEM pictures are shown here, one in each corner, of four different sized "anchor" serif structures that would be added on to the ends of lines to make the lines print closer to the intended design dimensions. The anchors can be seen to be increasing in size as one proceeds from the top left, to top right, to bottom left, to bottom right, of these four pictures. As can be seen, the serif structures have rounded corners.

Shortly after the designs were optimized by simulation and by interacting with the team of designers, a subset of the designs were submitted for processing. After processing the wafers through the diffusion level, a small set of samples were taken from the manufacturing line so as to get an early check on the shapes and sizes of printed patterns for the first two critical levels as compared with model prediction. Figure 2 shows a section of one of the SRAM cells for the gate and diffusion levels, where the latter was indicated by the presence of the oxide isolation region bordering the diffused regions. The SEM measurements were calibrated in the horizontal and vertical

directions using known pitch dimensions from the cell. Both the gate and diffusion SEM shapes were adjusted for a nonorthogonality component existing in the data. As can be seen, the agreement between simulation and experiment was quite good and seemed reasonably consistent with the predictions as discussed in Ref. 7.[21] We emphasize here that the shapes shown in Fig. 3 represent the final printed structures on the wafer, meaning that gates structures and oxide isolation regions were fully formed as marked on the SEM micrograph. Also, we note that the end shortening of the two center vertical bars in Fig. 2, which represented two gate structures, was predicted very accurately; end shortening of lines has been a particularly difficult quantity to predict accurately with simple models, so this agreement was quite encouraging. Figure 4 shows what the original CAD design looked like for the four levels; comparing Figs. 2 and 4, one can readily see the often significant difference that exists between the CAD design that microchip designers work with, and the final manufactured shapes.

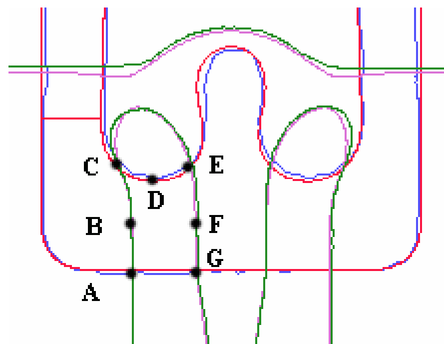

Figure 2: Section of one of the SRAM cell designs with total cell area $\sim 7\ \mu\text{m}^2$. The two vertical shapes and the top horizontal shape represent the gate level, while the shape that looks like a "W" indicates the diffusion region. Each shape has two curves, one being the pattern obtained from an SEM micrograph, while the other is the prediction from the phenomenological models for printing as described in Ref. 7. The SEM shapes are the final shapes on the wafer (*i.e.*, not simply the photoresist structures used in the patterning).

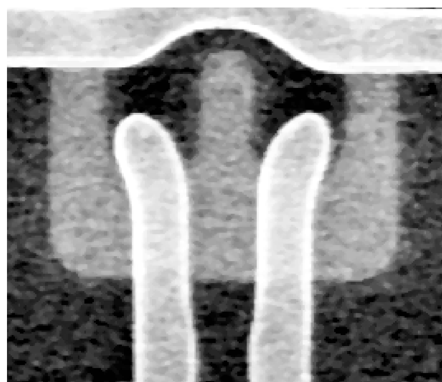

Figure 3: SEM micrograph of region of SRAM cell corresponding to the section in Fig. 2. The gate and diffusion regions are shown here, as formed on the semiconductor wafer. The smallest critical dimension shown here equals 0.19 μm.

6 STEPS FOLLOWED IN SIMULATION

Here we note that in order to turn a large series of simulation runs out in a fairly short span of time, for the six cells with all their possible variations, it was quite critical to have the models well calibrated ahead of time, and

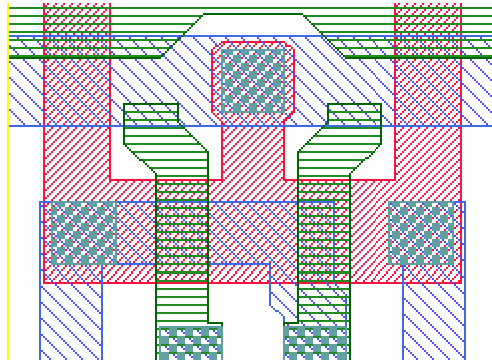

Figure 4: CAD layout of section of SRAM cell corresponding to Figs. 2 and 3. Note that the four design levels of gate, diffusion, contact, and first metal interconnect are superimposed here.

all parameters and conditions fully organized. Thus, of course the initial CAD designs needed to be obtained, but then beyond this, the manufacturing specifications for the sizes of each of the critical dimensions associated with each of the four levels needed to be precisely known. A table was made that contained the critical CAD dimension for each level, and the precise corresponding dimension for (1) the mask, (2) the photoresist developed structure, as well as (3) the final etched structure associated with either the gate, diffusion, contact, or first metal level. These dimensions between the CAD design, the mask, the photoresist structure, and the etched structure, differed either in the positive or negative direction, depending on the resist and etch processes used. The difference in dimensions could range as much as ± 0.04 μm between each step. Indeed, for one of the levels, the difference between the CAD dimension and the final etched dimension on the wafer was 0.09 μm for a 0.19 μm final etched dimension, nearly 50% of the final dimension. Thus, accurately accounting and keeping track of these biases and directions (negative or positive), was quite critical. The table construction of dimensions and model parameters enabled simulation jobs to be easily run by automatically extracting the required information for each run from the table.

A range of focus conditions was run for each level. Predictions for printability and process windows were used to improve the initial designs. Some of the optimization work to correct for specific proximity process effects was done semi-manually, and some in an automatic mode, using a combination of either internally developed IBM software, or the program EOPTIMASK.[22,23] Figure 5 illustrates typical output of these predicted results, overlaid on the original design data. The region in Fig. 5 corresponds with the same region in Figs. 2-4. Output like this was generated for enabling the printed structures for the photoresist as well as the etched structures on the wafer to be either viewed separately, or jointly, and compared level by level.

7 SPECIFIC OPTIMIZATION POINTS

Having illustrated the close agreement one can expect between the phenomenological models discussed in Ref. 7 and the actual printed shapes, we now turn to a series of points illustrating various optimization steps followed between the SRAM designers and the technologists for achieving a highly optimized and manufacturable cell design, prior to fabrication. Because there are so many points that one needs to consider, we simply list a number of them here, with the intent of having them represent the general sort of issues involved.

First, clearly, any issue that a designer might normally check, such as the intersection area between the contact structures and the levels that the contacts must connect, can be done far more accurately and realistically with these phenomenological models than could possible be done by only examining the intersection of the straight-

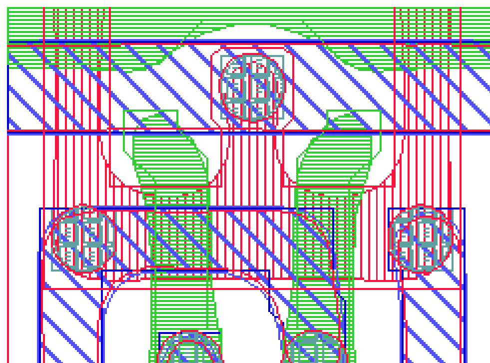

Figure 5: The predicted shapes from the phenomenological models described in Ref. 7 are shown here for the following four levels: gate, diffusion, contact, and first metal interconnect. The predicted shapes are overlaid with the original CAD design. As normally viewed by designers via a CAD tool, where each level can be assigned a different color and toggled off or on, or made more transparent or opaque, the identification of each level becomes significantly easier than as presented in the present black and white format.

edge polygon shapes in the CAD design. For example, if one looks at the design data for the diffusion level at the point where it intersects with the contact in Fig. 4 (see the square region in the top middle of the figure), and compare that to the predicted printed shapes in Fig. 5, which is emphasized by the overlay in Fig. 5 of the design data to the predicted printed shapes, one sees an enormous difference. Clearly, working with the design data alone to judge how levels will lie above and below each other, is far less accurate than doing so with the physical simulated data. For a designer wanting to aggressively minimize cell area, yet still wanting to ensure that levels adequately connect appropriately, then the physical predictions become essential.

Of course, as described earlier, much of this can be, and traditionally has been done by giving designers a set of "design rules," that roughly reflect the differences to expect between design and printed shapes. Information such as what is contained in Fig. 5, however, contains far more information than what is contained in a normal set of simple design rules.

To more specifically illustrate this point, Fig. 6 shows two contacts at the center of the figure, overlaying the ends of two metal lines. The rounding of the ends of the metal lines emphasizes to the designer that the contacts need to be moved away from the ends to ensure adequate intersected area between the two structures. Moreover, knowing the average overlay error associated with aligning one mask level to another, enables one to also take this factor into consideration when constructing the initial design. Thus, as can be seen in Fig. 6, as the metal line pattern is displaced vertically up or down by a small amount, or even horizontally, the two patterns still maintain a reasonable amount of intersected area.

Thus, two specific points have been mentioned so far regarding the use of the physically simulated structures to improve the general design, namely, the improved capability for predicted intersection area when optimal overlay exists, and the related topic of doing the same practice, but when some misalignment between levels is taken into account, as inevitably occurs during fabrication. We note that programs can be, and often are, written and used to ascertain the latter, as expressed in terms of the average intersection area due to a normally distributed random misalignment. Making use of the physically simulated shape enables this calculation to be far more realistic.

A third point has to do directly with proximity effects, but as generalized by examining the requirements between levels. Figures 7 and 8 show one of the full SRAM cells we analyzed. This cell had a size of 7.0 μm^2; it had the same basic design as that presented in Ref. 24, but with proximity corrections included. The "hammer

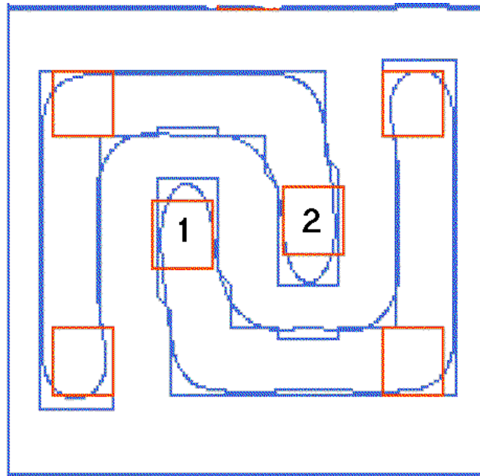

Figure 6: A section of one SRAM cell design is shown here, containing the contact level and the metal interconnect level. Both the CAD design and the simulated physical structure are displayed for the metal level; only the CAD design is shown here for the contact level.

head" additions to the ends of long structures were added in many places, as can be seen by a close examination of Fig. 7. The mask construction of four different sizes of these "hammer heads" was shown earlier in Fig. 1. As can be seen in Fig. 7, very different sizes and shapes of these hammer heads were added to different lines; some even have a left and right asymmetry associated with them. These differences were done on purpose, making use of the predictive power of how they would print so as to move structures as close together as possible, yet not result in bridging between structures during process window variation (defocus), and yet also ensure sufficient contact area between overlapping levels.

A fourth point that will be made here, which has more to do with the specific FET devices formed in the cell, is that one needs to design the FETs so that the gate length varies very little along the width dimension of the device. Figure 2 helps to illustrate this point. Again, the gates in this figure are the two center vertical shapes, while the region shaped like a "W" is the electrically active region containing dopant atoms. To the left of point B and to the right of point F, are the source and drain regions associated with the left-most gate. The FET "gate length," from source to drain, varies somewhat, namely, the distance from A to G, versus from B to F, versus from C to D to E, vary somewhat. The device turn-on characteristics are sharpest if this gate length has a very small variation. The situation shown here is under the condition where all structures were printed in best focus. Of course, out of focus conditions, or under or over exposed conditions, also contribute to variations in these structures. Figure 9 illustrates a corresponding section of a different SRAM cell where some serif structures were added on the diffusion level corresponding to points C, D, and E in Fig. 2, in order to improve the printability of this edge of the FET structure.

Of particular concern, is that misalignment between the gate and the diffusion pattern do not become too great, as that factor can have a very severe effect on the FET that is formed. For example, displacing the left-most gate to the left or right in Fig. 2 can have a big effect on the FET structure, and if the extreme case is considered where the gate is significantly lowered relative to the diffusion, the FET source to drain region could actually become electrically shorted. The extension of the gate above the point D was put in for this purpose, namely, to guard against the possible shorting of the device due to gate to diffusion misalignment.

Some of the FETs in an SRAM cell are more important than others in terms of their electrical performance and characteristics; consequently, different emphasis may be placed on different devices in order to enhance the

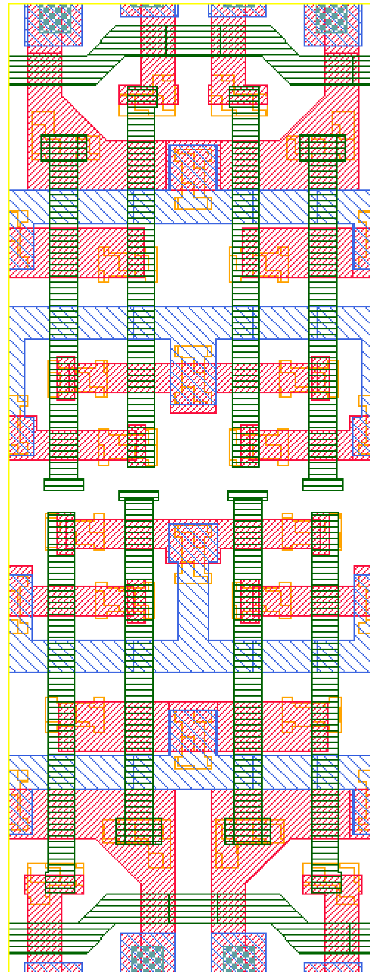

Figure 7: One of the six SRAM cell CAD designs. The four levels corresponding to the gate, diffusion, contact, and first metal levels are displayed here.

capabilities and process tolerances of the most critical ones. This same statement holds for semiconductor devices in general within a logic circuit.

A fifth point that was considered in the SRAM design stage was the reported evidence in the literature[25] of electrical leakage in an SRAM cell, apparently due to defects occurring in the underlying crystalline structure caused by stress and strain from the oxide isolation regions surrounding the active diffusion regions. We used the predicted fabricated shapes of these oxide isolation regions to estimate the effects of these shapes on the stress induced in the surrounding semiconductor regions.[26] Effects such as these are only possible to consider in the design phase if one has sufficient predictive power to accurately simulate the printed shapes of fabricated structures.

A sixth point we wish to make is that with realistic physical shapes predicted, more realistic estimates on signal delays could be made. In our work, we did not pursue the latter, but it is clear that such capabilities are possible. Of course, such determinations would need to be made in a computationally efficient manner, which is an extremely difficult task for a large circuit. Nevertheless, we note this point, as there are techniques that could take some advantage of the improved shape predictions shown here, such as, at the very least, by carrying

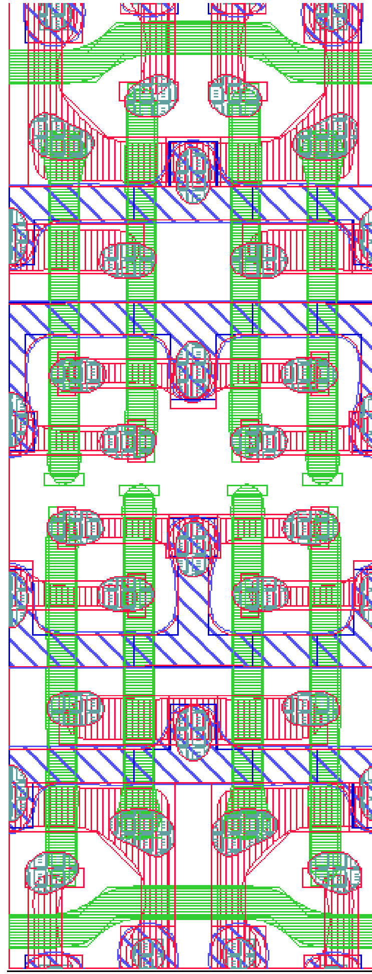

Figure 8: Physical predictions of printed shapes for the SRAM cell design in Fig. 7, overlayed with the CAD design.

out specific simulation experiments on selected regions to create a table of delay rules, that could be used more generally by a circuit simulator.

Finally, we wish to end with one final "optimization" point, which we view as the most significant one: namely, an enormous amount of work goes into creating a circuit library. To get the best return on this development activity, one wants the designs to be as reusable as possible when migrating the designs to a semiconductor technology with smaller dimensional fabrication capabilities. Migrating the designs easily can only be done if the original design was created with "scaleability" in mind from the very beginning. However, without at least some predictive capability to know how structures will print at smaller dimensions, then designing circuits to make them scaleable becomes incredibly difficult, if not impossible. Here we submit that models such as the ones reported in Ref. 7 can be used to gain at least some reasonable predictive capability to examine future scaling issues, thereby enabling present circuit designs to be made as reusable as possible. We note here that the present SRAM set of designs with nominal 0.25 μm critical dimensions, was designed to be scaleable to 0.18 μm critical dimensions.

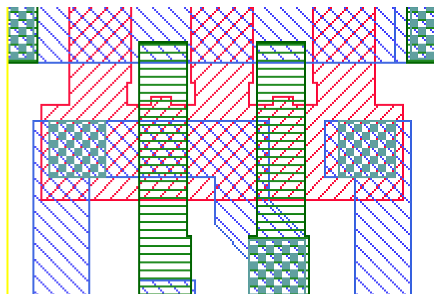

Figure 9: Serif structures were added in this section of an SRAM design to help the diffusion level print more like a horizontal line across the ends of the two vertical gates.

8 CONCLUDING REMARKS

A number of examples were illustrated here involving the use of fast, phenomenological models for predicting the printed patterns of photoresist structures in microlithography, as well as the final fabricated structures on a semiconductor wafer. An optimization study was qualitatively described that involved the close interaction of microchip designers and semiconductor technologists to create a highly optimized and manufacturable embedded SRAM cell in a logic technology, where the critical dimensions of the SRAM were smaller than those of the corresponding logic process, yet the same process steps for all were followed. This specific study was used as a vehicle to help point out means by which more general optimization work can be done for memory and circuit designs by combining the goals targeted by semiconductor developers, manufacturing technologists, and microchip designers.

The SRAM study described here involved what we consider to be a specific, hard optimization on a cell design. The extra effort on such a project is justifiable, and required, due to the projected high volume of such parts. However, for general random logic designs, of course more automated means need to be developed. Even here, however, we feel that the methods described in this article can be applied by making use of the fast simulation models to create a numerical laboratory where experiments can be run to generate, at the very least, design rules that can be applied either in a post-processor manner, or perhaps in yet more advanced ways. Using studies, such as the one described here, to look beyond single level optimization criteria, should be helpful for improving overall microchip performance, reliability, and yield. Moreover, it seems essential that if the large intellectual investment in circuit library generation is to be significantly extended via design reuse, so that circuit libraries can be scaled in reasonable manners as semiconductor technologies continue to advance to smaller dimensions, then making use of predictive models such as those used here, is required.

ACKNOWLEDGMENTS

We thank Anne McGuire for the SEM pictures of hammer head serifs shown in Fig.1. Also, we thank Jeff Johnson and Jim Slinkman, in particular, for their development of a means for using the lithography shape output discussed in Ref. 7 to calculate the stress and strain behavior on active diffusion regions due to oxide growth regions. D. C. Cole acknowledges work supported in part by DARPA and AFOSR.

9 REFERENCES

[1] T. A. Brunner and R. A. Ferguson, "Approximate Models for Resist Processing Effects," in *Optical Microlithography IX*, G. E. Fuller, ed., *Proc. SPIE* **2726**, pp. 198-207, 1996.

[2] J. Li, "Model-based Optical Proximity Correction Including Effects of Photoresist Processes," in *Optical Microlithography X*, G. E. Fuller, ed., *Proc. SPIE* **3051**, pp. 643-651, 1997.

[3] N. Cobb, A. Zakhor, and E. Miloslavsky, "Mathematical and CAD Framework for Proximity Correction," in *Optical Microlithography IX*, G. E. Fuller, ed., *Proc. SPIE* **2726**, pp. 208-222, 1996.

[4] N. Cobb and A. Zakhor, "Experimental Results on Optical Proximity Correction with Variable Threshold Resist Model," in *Optical Microlithography X*, G. E. Fuller, ed., *Proc. SPIE* **3051**, pp. 458-468, 1997.

[5] J. Zhao, J. Garofalo, J. Blatchford, E. Ehrlacher, and E. Nease, "Applications of Enhanced Optical Proximity Correction Models," in *Optical Microlithography XI*, Luc Van den hove, ed., *Proc. SPIE* **3334**, pp. 234-244, 1998.

[6] C. N. Ahn, H. B. Kim, and K. H. Baik, "A Novel Approximate Model for Resist Process," in *Optical Microlithography XI*, Luc Van den hove, ed., *Proc. SPIE* **3334**, pp. 752-763, 1998.

[7] E. W. Conrad, D. C. Cole, D. P. Paul, and E. Barouch, "Model Considerations, Calibration Issues, and Metrology Methods for Resist-Bias Model," in *Metrology, Inspection, and Process Control for Microlithography XIII*, B. Singh, ed., *Proc. SPIE* **3677**, (in print) 1999.

[8] D. C. Cole, E. Barouch, U. Hollerbach, and S. A. Orszag, "Derivation and Simulation of Scalar Aerial Images for Higher Numerical Apertures," *Jap. J. Appl. Phys.* **31**, pp. 4110-4119, 1992.

[9] E. Barouch, D. C. Cole, U. Hollerbach, S. A. Orszag, "Vector Aerial Image with Off-Axis Illumination," in *Optical/Laser Microlithography VI*, J. D. Cuthbert, *Proc. SPIE* **1927**, pp. 686-708, 1993.

[10] Y. C. Pati and T. Kailath, "Phase-Shifting Masks for Microlithography: Automated Design and Mask Requirements," *Journal of the Optical Society of America A*, **11** (9), pp. 2438-2452, 1994.

[11] D. A. Bernard, J. Li, J. C. Rey, K. Rouz, V. Axelrad, "Efficient Computational Techniques for Aerial Imaging Simulation," in *Optical Microlithography IX*, G. E. Fuller, ed., *Proc. SPIE* **2726**, pp. 273-287, 1996.

[12] A. Starikov, "Use of a Single Size Square Serif for Variable Print Bias Compensation in Microlithography: Method, Design, and Practice," in *Optical/Laser Microlithography II*, *Proc. SPIE* **1088**, pp. 34-46, 1989.

[13] C. C. Fu, T. Yang, and D. R. Stone, "Enhancement of Lithographic Patterns by Using Serif Features," *IEEE Trans. Electron Devices* **ED-38** (12), pp. 2599-2603, 1991.

[14] D. C. Cole, E. M. Buturla, S. S. Furkay, *et al.*, "The Use of Simulation in Semiconductor Technology Development," *Solid-State Elec.* **33**, pp. 591-623, 1990.

[15] F. Jones and J. Paraszczak, "RD3D: Computer Simulation of Resist Development in Three Dimensions, " *IEEE Trans. Electron Devices* **ED-28** (12), p. 1544-1552, 1981.

[16] E. Barouch, B. D. Bradie, U. Hollerbach, G. Karniadakis, S. A. Orszag, "Comprehensive 3-D Notching Simulator with Nonplanar Substrates," in *Optical/Laser Microlithography III*, V. Pol, ed., *Proc. SPIE* **1264**, pp. 334-342, 1990.

[17] E. Barouch, U. Hollerbach, S. A. Orszag, C. R. Szmanda, J. W. Thackeray, "Simulations of Bar Printing Over a MOSFET Device Using I-line and Deep-UV Resists," in *Optical/Laser Microlithography IV*, V. Pol, ed., *Proc. SPIE* **1463**, pp. 464-474, 1991.

[18] J. J. Helmsen, E. G. Puckett, P. Colella, M. Dorr, "Two new methods for simulating photolithography development in 3D," in *Optical Microlithography IX*, G. E. Fuller, ed., *Proc. SPIE* **2726**, pp. 253-261, 1996.

[19] See, for example, C. Yu, "Integrated Circuit Process Design for Manufacturability Using Statistical Metrology," Ph.D. Dissertation, 5/96, University of California at Berkeley.

[20] W. Leipold, F. Baker, O. Bula, R. Leidy, and A. McGuire, "Polysilicon Line End Resolution Enhancement: Improving Yield on IBM's Current and Future Technologies," *IBM MicroNews* **4** (3), pp. 19-23, 1998.

[21] Since the shapes shown in Fig. 2 were the final *etched* shapes on the wafer, for two levels, as compared with the predictions of our resist bias model plus etch bias model, then having somewhat worse agreement than pure resist structure results on a single level, as mainly studied in Ref. 7, seemed very reasonable.

[22] E. Barouch, U. Hollerbach, R. R. Vallishayee, "OPTIMASK: an OPC Algorithm for Chrome and Phase-Shift Mask Design," in *Optical/Laser Microlithography VIII*, T. A. Brunner, ed., *Proc. SPIE* **2440**, pp. 192-206, 1995.

[23] M. Garza, E. Jackson, W. P. Shen, N. Eib, et al., "Achieving Subhalf-Micron I-Line Manufacturability Through Automated OPC," in *Emerging Lithographic Technologies*, D. E. Seeger, Ed., *Proc. SPIE* **3048**, pp. 382-395, 1997.

[24] S. Subbana, P. Agnello, E. Crabbé, et al., "A High Density 6.9 sq. μm Embedded SRAM Cell in a High Performance 0.25 μm-Generation CMOS Logic Technology," IEDM Tech. Dig., pp. 275-278, 1996.

[25] K. Ishimaru, F. Matsuoka, M. Takahashi, et al., "Mechanical Stress Induced MOSFET Punch-through and Process Optimization for Deep Submicron TEOS-O_3 Filled STI Device," IEDM Tech. Dig., pp. 123-124 (1997).

[26] The development of a simulation capability for this effect was due to the efforts largely of J. B. Johnson and J. A. Slinkman.