1

INTRODUCTION

COPYRIGHTED MATERIAL

1.1 VALUE OF DESIGN FOR MANUFACTURABILITY

Many designers are still unclear as to the value of having design for manufacturability (DFM) inserted into their design flow, simply because DFM requires additional resources (tool cost and design resources), design delays, and so on. No designer is willing to sacrifice schedule while increasing the resources required—just to achieve a better yield. Designers are always seeking ways to improve performance, power, and die size while minimizing design margins and eliminating the need to rework the design as a result of circuit bugs. To be attractive to designers, DFM must offer avenues to achieve these goals as well.

The ultimate reward for using DFM is an economic one. A design with higher performance, lower power, and smaller die size translates to a higher average selling price (ASP) and lower manufacturing costs. An improved and predictable yield during manufacturing results in a reduced time to market, higher profits, and a longer product lifetime. Nowak [1] describes this economic concept, which bridges the return on investment (ROI) gap between design and manufacturing. The concept is well illustrated in Figure 1.1, where the dashed line shows a possible life cycle of a design in an advanced technology node without deploying DFM, and the solid line shows a life cycle of similar design incorporating DFM. Figure 1.1 suggests that a design with DFM will result in a faster, more predictable yield ramp, thus less time to market, higher profits, and a longer product life cycle.

Nano-CMOS Design for Manufacturability: Robust Circuit and Physical Design for Sub-65 nm Technology Nodes
By Ban Wong, Franz Zach, Victor Moroz, Anurag Mittal, Greg Starr, and Andrew Kahng
Copyright © 2009 John Wiley & Sons, Inc.

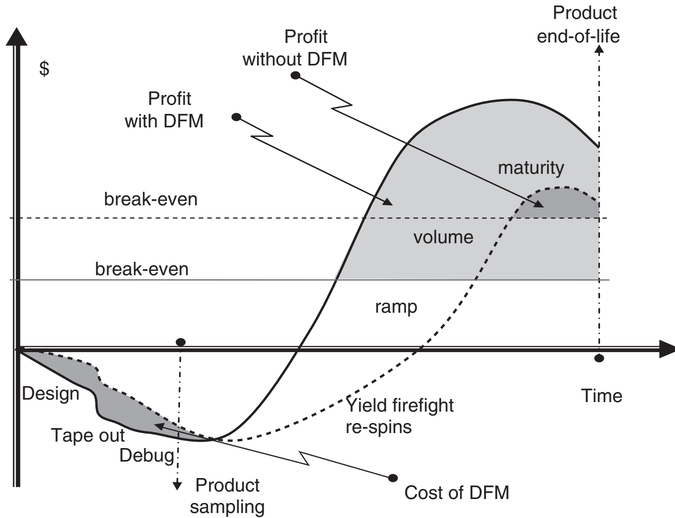

Figure 1.1 Economics of DFM. (Adapted with permission from Nowak/Radojcic, ref. 1.)

As the technology advances, the critical dimensions (CDs) of many critical layers are well into the subwavelength regime, and film thickness has approached atomic layer dimensions. The result is greater variability. For example, the gate dielectric of a typical 65-nm node is on the order of four atomic layers thick. It would be impossible for any process to place four atomic layers precisely on the gate oxide of every transistor on an entire wafer. Also, with gate length CDs in the sub-40-nm range, 4 nm of variability would represent a 10% change in the CDs. This is a meager CD budget, yet would have to be shared among the various processes: lithography, etching, optical and etching proximity effects, mask error, and mask error enhancement factor (MEEF) [2, Chap. 3]. Without designer intervention, it would be impossible to achieve the CD budget. A large polysilicon ("poly") CD variation would result in a large spread in product performance, power, and leakage.

Dopant fluctuation is unavoidable at these dimensions, where the number of dopants for a transistor in a typical standard cell in the 65-nm node is less than 100 dopant atoms, resulting in a large $\sigma\Delta V_t$ (threshold voltage) value. It has been determined that the dopant location is also important at these dimensions [3]. Figure 1.2 shows that the V_t of two transistors with the same number of dopant atoms can still be very different, depending on the location of the dopants. The dopant location is a result of the difference in energy that each dopant acquired as it was being propelled toward the silicon wafer. No one has yet been able to cut the energy tail of the implanters. Process variability has greater impact now, and if not designed for, reduces parametric yields [3–5].

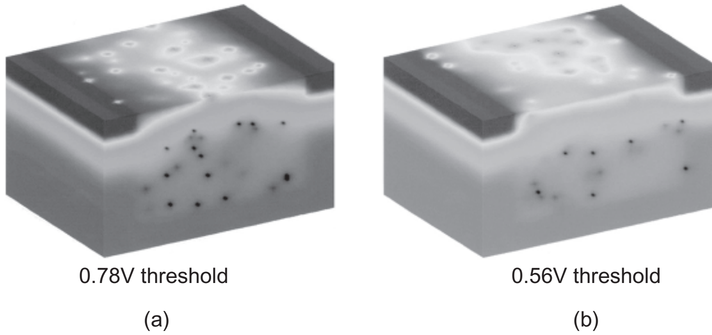

Figure 1.2 Random dopant location-induced V_t fluctuation (a) 0.78-V threshold; (b) 0.56-V threshold. Both devices have 170 dopants in the channel depletion region. (From ref. 3.)

1.2 DEFICIENCIES IN BOOLEAN-BASED DESIGN RULES IN THE SUBWAVELENGTH REGIME [6]

With older technology, Boolean-based design rules worked well and have been the design sign-off to guarantee a manufacturable design. As technology scales, problems arise. Figure 1.3a shows a failure under certain process conditions for a design that meets design rules. In the subwavelength regime, the limitations of Boolean-based design rules are beginning to show. What is more critical is the case shown in Figure 1.3b, where the lines are not yet shorted but are already so close that they pose a reliability hazard. How many such reliability hazards are lurking in the design? That will be difficult to quantify or detect without a model-based tool. Figure 1.4 shows another typical failure as a result of pattern distortion due to the optical proximity effect in the subwavelength regime. The two landing pads are equidistant from the line above. The only difference within this context is the tab on the horizontal line that is above the right landing pad (pad B). This difference in proximity caused the right landing pad to be pulled toward the horizontal line above, resulting in shorts under certain process conditions. Incidentally, the short occurred between structures that are not minimally spaced relative to each other. In the absence of the proximity effect as in landing pad A, no short is seen.

For the 90-nm node, the first-order proximity effect is to the structures immediately adjoinly the polygon of concern. At the 45-nm node the proximity effect influence is as far as a few structures away from the polygon. When the proximity effects are so far reaching, it is very difficult to code Boolean-based rules to describe this effect so that designers can design for it. At these advanced nodes a model-based approach would be inevitable to fully describe this effect to designers so that it can be avoided in their design.

For each technology generation, manufacturers attempt to deal with the problem by resolving the densest pattern that the design rules allow. However, this does not mean that they can print the full chip of any design. Most

Figure 1.3 Failure at a process corner for a design that met design rules.

Figure 1.4 Example of the need for a model-based solution.

problems are the result of two-dimensional patterns, such as line-end patterns, where the diffraction patterns are much more complicated than the simple dense line–space pattern that the manufacturers have developed for the lithography process. The line ends are also most susceptible to etch distortion and proximity effects, as well as a different response to poly CD trim. These types of issues will get worse as the technology advances, resulting in lots of opportunities for designer intervention on designs of the future. Model-based tools will be required to help find and fix these hotspots. Boolean-based design rules alone will no longer be sufficient. Without model-based tools, the Boolean-based design rules will have to be very complex as well as there being too many to be practical. Even then, they will still not be well enough developed to address all the new and newly exacerbated design–process interactions.

1.3 IMPACT OF VARIABILITY ON YIELD AND PERFORMANCE

At the process and device levels, we are seeing line edge roughness (LER) contributing more significantly to channel-length variability at the CDs of nano-CMOS devices. This results in higher device OFF-current (I_{off}) variability, as shown in Figure 1.5. Chips designed for used in hand held devices will be affected by this higher I_{off} variation and must be designed for, or this will present a yield issue that delays product introduction (see in Figure 1.1.)

Poly CD control is getting difficult, but the criticality for many circuits is not abating, as shown in Figure 1.6, where 6% lower poly CDs result in an unusable product. For microprocessors the speed versus average selling price is nonlinear; in fact, it increases exponentially with speed. There is a huge financial motivation to make the poly CDs as narrow as possible and still have a good product yield. The lower the poly CDs, the higher the speed of the part will be. The better the CD control, the lower the CDs can be pushed. Figure 1.7 shows that the margins drop very quickly with reduced poly CDs. There are a lot of opportunities for a designer to participate in improving the poly CD control to produce the best-performing part with the least yield loss. Figure 1.8 shows the V_t response to poly CDs. When the CDs are larger than the target values by 10%, we see that the V_t spread improves, whereas a –10% value for the poly CDs results in a greater change in V_t from the target as well as a larger CD σ value. Many of the devices have V_t values below zero, which means they will never shut off, resulting in massive current leakage. Design opportunities to improve poly CD control are covered in more detail in Chapter 6.

Figure 1.9 shows the delay distribution of two similar designs with different layout styles. It is clear from these data that the design style can have a large effect on the σ value of the delay distribution.

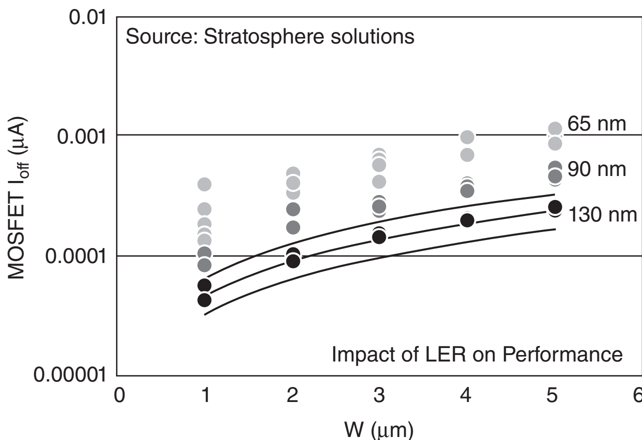

Figure 1.5 I_{off} variability as a result of line-edge roughness. (Courtesy of Stratosphere Solutions.)

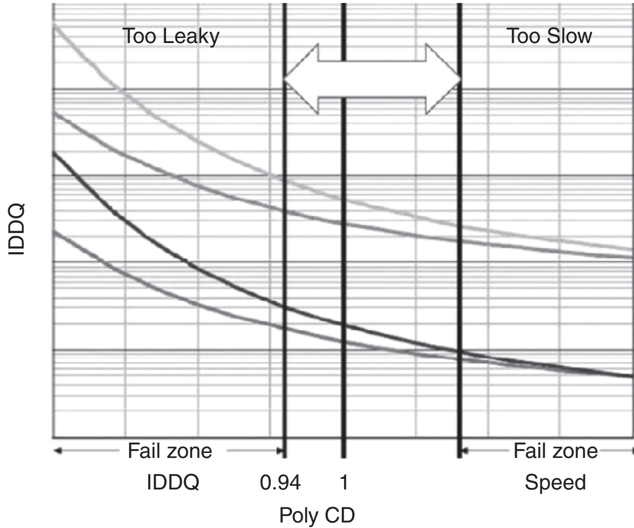

Figure 1.6 IDDQ response to poly CD.

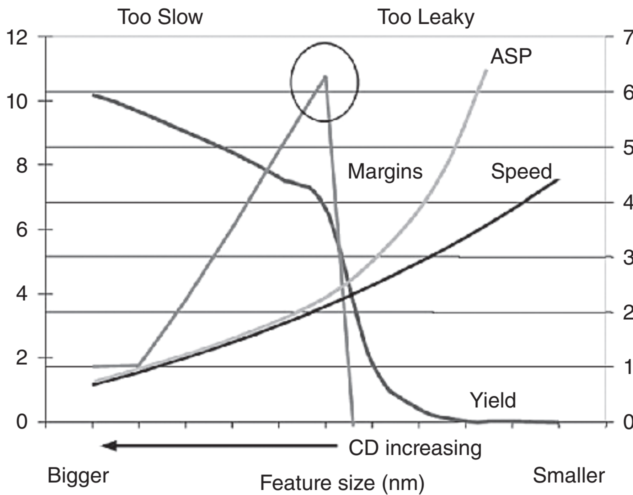

Figure 1.7 Design margin as a function of poly CD.

Patterning reproducibility will be a challenge, as shown in Figure 1.10. The two poly lines shown in the figure are supposed to be identical. As a result of patterning-reproducibility limitations as well as LER, they look different under high magnifications. Transistor matching will be difficult at the nano-CMOS nodes.

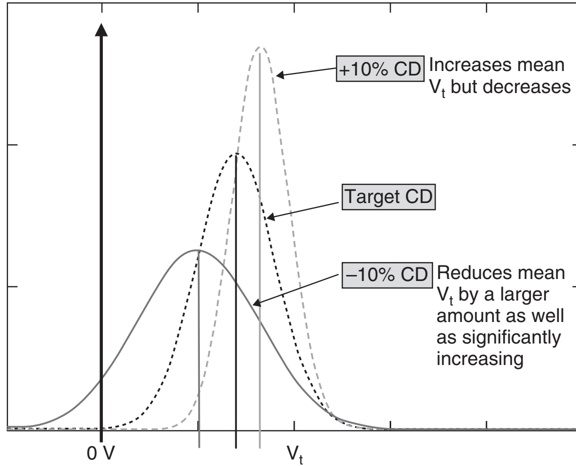

Figure 1.8 V_t spread versus poly CD.

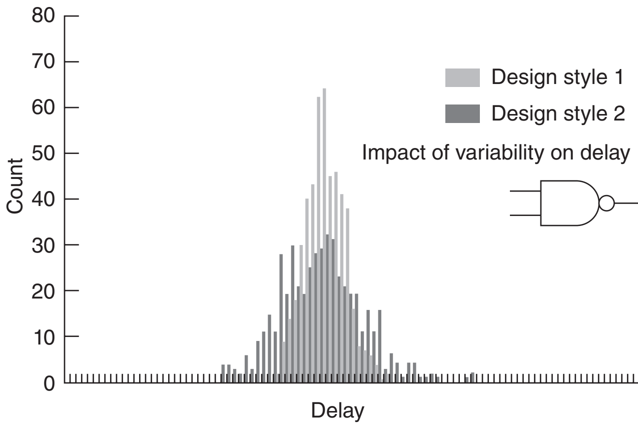

Figure 1.9 Layout style and design context affect variability. (Courtesy of Stratosphere Solutions.)

Process-Related Variability

- Line-edge roughness, random dopant fluctuations and location, poly CD control, and so on: have greater impact on electrical properties
- Mask error and mask error enhancement factor in a low-k_1 lithography process: affect electrical properties and device matching
- OPC and patterning reproducibility: a matching issue for analog, memory, and clock distribution designs

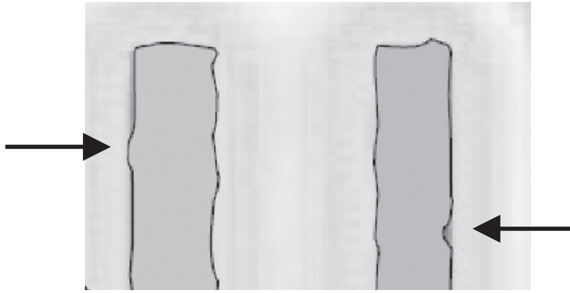

Figure 1.10 Challenge of patterning reproduceability at nano-CMOS dimensions. (From ref. 4.)

Device-Level Variability

- V_t, device ON current (I_{on}), and I_{off} variability: increase
- Resistance and capacitance variability: increases for vias, contacts, and metal
- Transistor, resistor, and capacitor: mismatch increases
- NBTI and other V_t shift mechanisms
- Stress and strain variability: film thickness and proximity
- Lithography distortions: OPC, optical proximity, and lithographic condition dependent

Cell- or Circuit-Level Variability

- Delay distributions: much broader
- Cell (block) placement and rotation: induce variability
- Layout- and context-dependent stress and strain variability: results in I_{on} variability
- Well proximity effect
- Layout effect on mobility

Chip-Level Variability

- Increase in *di*/*dt*: can result in timing and functional failures
- Signal integrity: has a major impact on timing and functionality
- Greater spread in chip standby power
- Voltage and temperature gradient: affects device performance
- Substrate noise (triple well isolation may not work for radio frequencies)
- Jitter

1.4 INDUSTRY CHALLENGE: THE DISAPPEARING PROCESS WINDOW

Since the 65-nm technology node was developed, manufacturers have been working below a robust manufacturing level. Even when the process has been extended by all known techniques, with the numerical aperture (NA) pushed as high as feasible using the latest scanners, manufacturers are unable to keep the process above the robust manufacturability line (Figure 1.11). The result is that the process window is shrinking as the optical lithography process is falling behind Moore's law. The illumination wavelength has been stuck at 193 nm, due to difficulties in bringing extreme ultraviolet scanners on line, and to materials and integration concerns with the 157-nm wavelength (Table 1.1) [5–7]. The knobs that are available are NA and immersion lithography to make greater than unity NA work with the depth of focus (DOF) margin in these

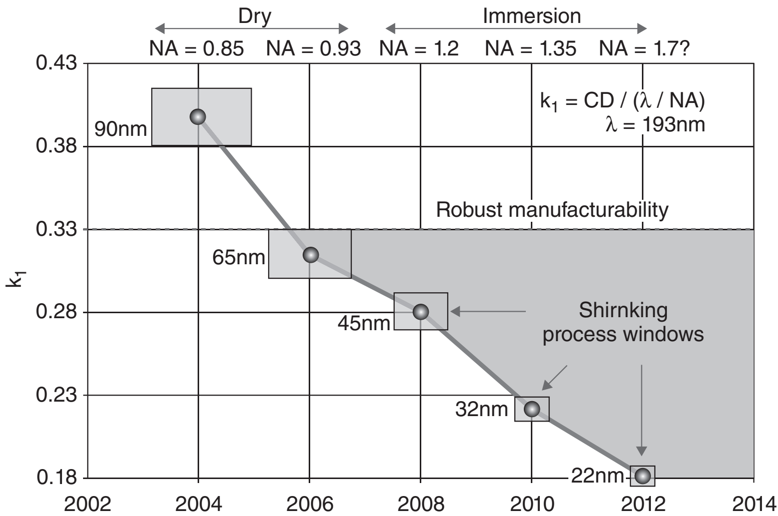

Figure 1.11 Industry's challenge of disappearing process windows. (Courtesy of Invarium.)

Table 1.1 Illumination sources available for optical lithography

Wavelength (nm)	Illumination Source	Year Introduced
436	g-line	1980
365	i-line	1989
248	KrF laser	1995
193	ArF laser	2002
193i	ArF laser	2006
157	F_2 laser	Failed
13.4	EUV	?

processes. Aggressive resolution enhancement techniques (RETs), phase-shift masking (PSM), and illumination optimization are the other knobs available. Even when all these techniques are deployed to the fullest, the process window is still shrinking [7].

Therefore, designers must now do their part to bridge the process–design gap to make their designs manufacturable within these technology nodes. In Chapter 2 we go into more detail on the challenges of the lithographic process and the solutions available to manufacturers as well as designers.

1.5 MOBILITY ENHANCEMENT TECHNIQUES: A NEW SOURCE OF VARIABILITY INDUCED BY DESIGN–PROCESS INTERACTION

To remain competitive, all foundries have resorted to some sort of stress engineering using stress memory techniques (SMTs), contact etch stop layer (CESL) stressing film, or recessed source–drains using embedded silicon–germanium (eSiGe). CESL tensile and compressive film is used to enhance nMOS and pMOS drive current, respectively (more details in succeeding chapters).

Figure 1.12 shows all the layout parameters that affect transistor carrier mobility. The mobility is affected by the stress applied to the transistor channel by the stressing film or recessed eSiGe source–drain and shallow trench isolation (STI). The stress applied to the channel by the CESL film is proportional to the volume and proximity of the film to the channel. The volume and proximity of the film to the channel are modulated by poly pitch, contact pitch, contact-to-poly (channel) space, and contact critical dimensions (CDs). The number of contacts on a transistor affect the degree to which the CESL stressing film is perforated, which relaxes the film, resulting in reduced stressing film

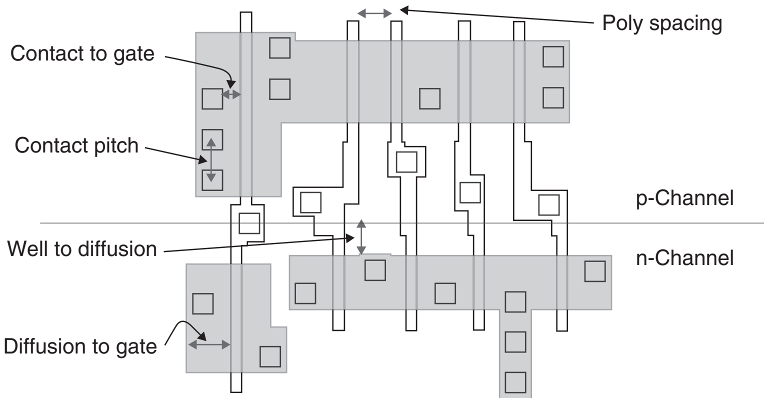

Figure 1.12 Effect of layout parameters on carrier mobility. (Courtesy of Synopsys.)

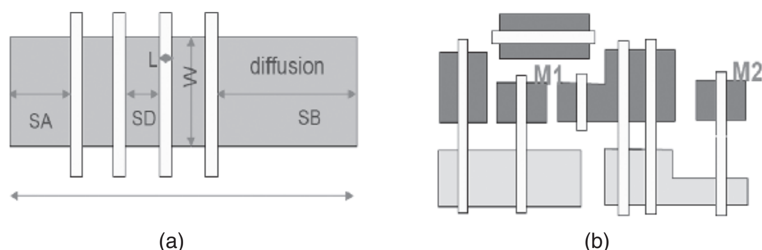

Figure 1.13 Stress proximity–device drive strength modeling issue. (a) BSIM4 LOD model is indifferent to context (the surrounding layout patterns). (b) M1 and M2 have identical *W*/*L* and LOD, but different stress and drive strength. (Courtesy of Synopsys.)

effectiveness. This results in lower carrier mobility, hence lower transistor drive.

STI in the 90- and 65-nm technologies exerts a compressive stress on transistors. As the distance of the edge of the diffusion (where STI begins) to the transistor channel is increased, the effect of the STI stress as seen by the transistor channel diminishes. Therefore, varying that distance will change the transistor drive current as the stress level changes. Since STI exerts compressive stress, it degrades nMOS drive current but improves pMOS drive.

Figure 1.13 illustrates modeling issues with the current BSIM length of diffusion (LOD) model. The parameters available to the current LOD model include SA, SB, and SD. These parameters cannot differentiate transistors of differing context and can seriously misrepresent the stress effects experienced by a similar transistor with a different context. For example, M1 and M2 in Figure 1.13b have similar LODs but very different contexts and so will experience significantly different channel stress, which will not be captured by the current BSIM4 LOD model. The result is that a designer using the BSIM4 LOD model will find that his or her simulations will defer from the behavior of the circuit on silicon. We discuss this issue and its solutions more in subsequent chapters.

Figure 1.14 shows the effect of eSiGe recess depth on the stress in the channel—hence the drive strength. The greater the recess depth, the greater is the stress on the channel. However, at minimum poly-to-poly spacing we see a steep change in the stress level with a very slight change in the poly-to-poly spacing. Therefore, at minimumal poly-to-poly spacing, the drive strength of the device changes a lot if the process varies the spacing. Poly-to-poly spacing variation can be a result of poly CD variation. Poly CD variation in such a context affects the transistor drive by varying the transistor threshold voltage as well as the mobility of the carriers as the stress-level changes with the poly-to-poly spacing. Therefore, it is prudent not to design critical circuits with minimum poly-to-poly spacing to move into the flatter portion of the curve in Figure 1.14.

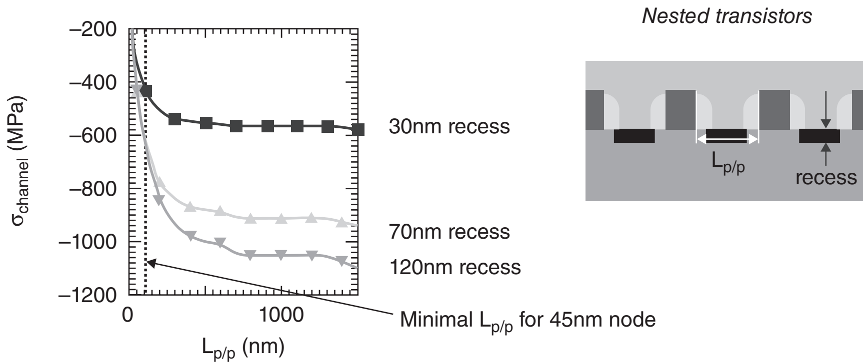

Figure 1.14 Layout process interaction. (Courtesy of Synopsys.)

1.6 DESIGN DEPENDENCY OF CHIP SURFACE TOPOLOGY

Ever since the industry switched from aluminum (Al) to copper (Cu) interconnects, the planarization and copper clearing technique used is chemical–mechanical polishing (CMP). Due to the differences in the materials' (copper and dielectric) resistance to wear during the CMP process, the topology of the chip is affected by differing copper density on the chip, which in turn is dependent on the design. Manufacturers use various techniques to normalize the copper density, but these techniques are not perfect and on some designs it may be difficult to normalize the density without changing the design characteristics and intent. The result is that the topology of most chips is far from being flat (Figure 1.15). This can cause catastrophic failures, due to the inability of CMP to clear the copper on the higher levels of interconnect, as shown in Figure 1.16. Figure 1.17 shows that additional interconnect layers compound the topological impact, which eats into the already narrow DOF margin in the lithographic process and can create lithographic hotspots as well.

Chip surface undulation also results in parametric variation, as it causes varying copper loss on interconnects at different locations. The interconnect resistivity and capacitance will vary so that as-drawn extraction will not reflect reality and will result in overdesign to allow for the unmodeled variations. There is therefore a need for a tool to help model these effects and to correct for them as much as possible.

1.7 NEWLY EXACERBATED NARROW WIDTH EFFECT IN NANO-CMOS NODES

It has been observed that the effective channel length of devices at the STI edge will be longer than the channel length away from the STI edge [8,9] (Figures 1.18 and 3.3). The result is that narrow devices will have longer

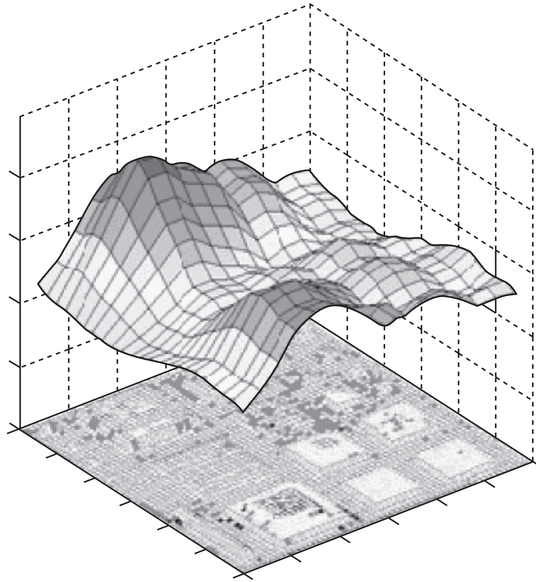

Figure 1.15 Design-dependent undulating chip surface. (Courtesy of Praesagus.)

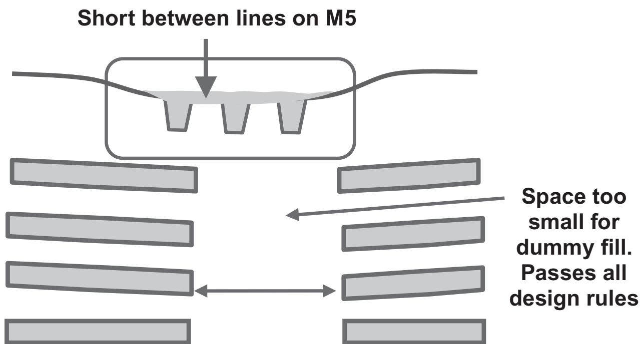

Figure 1.16 Catastrophic CMP failure.

channel lengths, hence lower drive current and higher V_t. Devices with different widths will exhibit varying V_t and drive. This can have a significant effect on memory bitcell designs where the pull-down and pass transistors are of different widths. If manufacturers cannot eliminate this effect, it must be reflected in the model, so that design engineers can design for it. As shown in Figure 1.18, the longer channel length persists even 100 nm from the STI edge, which means that a 200-nm device will be affected by this newly exacerbated phenomenon. Although all known techniques will be utilized to minimize this effect, it remains to be seen if all manufacturers can eliminate it. Therefore, when designing circuits that incorporate devices with narrow widths, as in

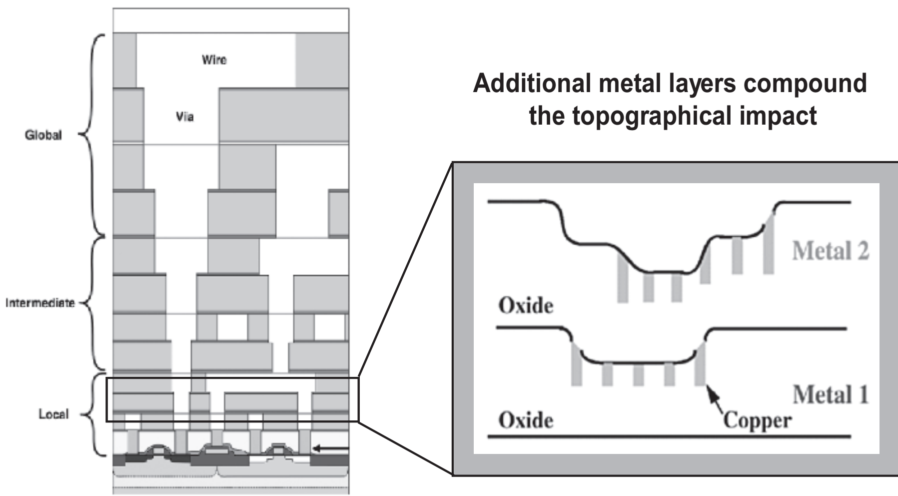

Figure 1.17 Growing problem with more metal layers. (Courtesy of Praesagus.)

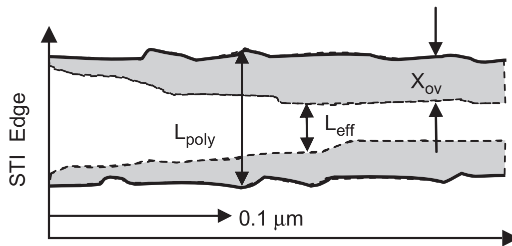

Figure 1.18 Narrow width effect on L_{eff}. X_{ov}, extension overlap; L_{eff}, effective channel length; solid lines, gate edge; dashed lines, junction edge.

memory bitcells, keeper devices, and other designs that use minimum geometry devices, it is important to keep this effect in mind.

1.8 WELL PROXIMITY EFFECT

This effect has been observed to affect circuit performance and functionality since the advent of the 130-nm node. Designers must exercise caution when placing critical circuit components, particularly transistors, in the scatterbands of a well (Figure 1.19). For about 2 µm from the edge of the well, we will see dopants scattered by the implant mask, resulting in higher well doping in the 2-µm band around the edge of the well. Inner corners are much worse than outer corners, as shown by the well scatter dose map in Figure 1.19. Designs that require matched transistors, self-time circuits, and hold-time delay

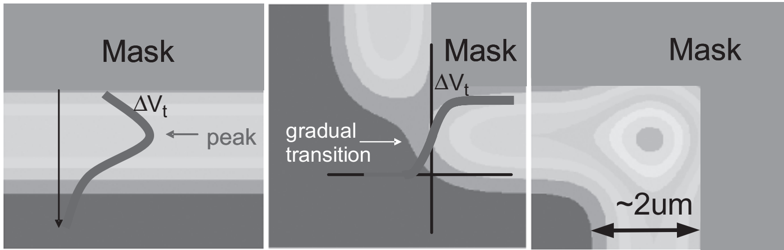

Figure 1.19 Well scatter dose map.

elements should preferably be located away from the well scatterbands. In later chapters we discuss design techniques to minimize the effect of well proximity effects on circuits that must be placed in the scatterbands as a result of space constraints, due to pitch-matching layouts.

1.9 NEED FOR MODEL-BASED DFM SOLUTIONS BEYOND 65 NM

Although the lithographic process is lagging behind aggressive dimensional scaling, we are seeing the circle of influence of proximity being enlarged proportionately. Furthermore, chip surface undulation modulated by the design effect on the CMP step adds another level of abstraction that affects the lithographic process, making it even more difficult to be described by rules alone. Problematic proximity interactions cannot be accomplished easily when using rules. Even if one could code the design rule checking (DRC) deck with the ability to find such interactions, it would be unmanageably complex, with a disproportionately large number of rules, to the extend that DRC runs would take so long that real-time data for the chip tapeout could not be provided. Even if the run-time problem could be surmounted using expensive supercomputers, the rules would be so extensive and complex that design productivity would be an issue. The misinterpretation of the rules and errors in the implementation would also cause delays in tapeouts and result in missing the product market window. Figure 1.3b is a classic case that can only be uncovered by a model-based solution.

The effects of lithographic distortions modulated by chip surfaces can be captured accurately only by using a physics-based model calibrated to the process [10]. The benefits include the ability to push rules to achieve a smaller chip area guided by the model. This will result in a smaller chip than in the case of blind adoption of the restricted design rules. Contour-based extraction also relies on a calibrated model which can be used to turn systematic variation into deterministic effects that can be simulated. Therefore, no extra margins need to be set aside for systematic variations, reducing overdesign.

The key to successful implementation of a DFM flow is the ability to provide early design feedback when the design is still fluid, to avoid major restructuring of the design, resulting in delayed tapeout. There is also a need for correct by construction DFM and timing-aware routing, to minimize reroutes and lead to a faster time to market. This requires a fast model-based analysis that guides the router. The first-generation guided router would be a hybrid of a rule-based guide and a fast model-based solution to deal with constructs that cannot be described by rules alone and to confirm the manufacturability of a particular construct. The model-based solution also guides the repair of a high-priority hotspot.

1.10 SUMMARY

Although we have not listed possible design process interactions exhaustively, the list in this chapter serves to illustrate the point that down the road of technology scaling we will see these effects becoming more predominant. We will also find new effects that we have not seen in earlier technology nodes. Process variability is also here to stay and is getting worse as dimensional scaling pushes toward atomic dimensions, where, for example, gate dielectric thickness has already approached a thickness of four atomic layers. Hence, we have to learn to deal with these effects and the increase in variability so that our circuits will be functional on the first silicon representation for cost and time-to-market economics. As a result, the need for model-based solutions will abound to help designers know when they can use minimum rules for a smaller chip area and how to avoid process-sensitive layout constructs for a predictable yield ramp for volume production. Model-based solutions also provide the basis to turn systematic variations into deterministic events, thus eliminating the need for additional design margins, which results in overdesigning. This book combines the expertise of circuit engineers, process technologists, lithography experts, computer-aided design developers, and academic researchers to provide practical solutions and silicon-proven methodology to deal with the problems described.

REFERENCES

1. M. Nowak, Bridging the ROI gap between design and manufacturing, SNUG 2006, Santa Clara, CA, 2006.
2. B. P. Wong et al., *Nano-CMOS Circuit and Physical Design*, Wiley, Hoboken, NJ, 2004.
3. A. Asenov, Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: a 3-D "atomistic" simulation study, *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec. 1998.

4. M. Pelgrom, Nanometer CMOS: an analog challenge, IEEE Distinguished Lectures, Fort Collins, CO, May 11, 2006.
5. B. P. Wong, Design in the nano-CMOS regime, DFM tutorial, Asia and South Pacific Design Automation Conference, Yokohama, Japan, Jan. 23, 2007.
6. B. P. Wong, Bridging the gap between dreams and nano-scale reality, DFM tutorial, Design Automation Conference, San Francisco, CA, July 28, 2006.
7. Y. Borodovsky, Marching to the beat of Moore's law, SPIE Microlithography Conference, San Jose, CA, 2006.
8. V. Moroz et al., Physical modeling of defects, dopant activation and diffusion in aggressively scaled bulk and SOI devices: atomistic and continuum approaches, Materials Research Society Conference, Boston, MA, Apr. 19, 2006.
9. H. Fukutome et al., Direct measurement of effects of shallow-trench isolation on carrier profiles in sub-50 nm N-MOSFETs, Symposium on VLSI Technology, Kyoto, Japan, 2005.
10. D. Perry et al., Model-based approach for design verification and co-optimization of catastrophic and parametric-related defects due to systematic manufacturing variations, SPIE Microlithography Conference, San Jose, CA, 2007.