Manufacturing-Aware Physical Design *

Puneet Gupta† and Andrew B. Kahng†‡
* Department of Electrical and Computer Engineering, UC San Diego, La Jolla, CA, USA
‡ Department of Computer Science and Engineering, UC San Diego, La Jolla, CA, USA
puneet@ucsd.edu, abk@ucsd.edu

Abstract

Ultra-deep submicron manufacturability impacts physical design (PD) through complex layout rules and large guardbands for process variability; this creates new requirements for new manufacturing-aware PD technologies. The first part of this tutorial reviews PD complications and methodology changes - notably in the detailed routing arena - that arise from subwavelength lithography and deep-submicron manufacturing (antennas, metal planarization and mask-wafer mismatch). Process variations and their sources are taxonomized for modeling and simulation. A framework of design for cost and value is described. The second part covers yield-constrained optimizations in PD, especially "beyond corners" approaches that escape today's pessimistic or even incorrect corner-based approaches. Statistical timing and noise analyses enable optimization of parametric yield and reliability. Yield-aware cell libraries and "analog" design rules (as opposed to "digital", 0/1 rules) can help designers explore yield-cost tradeoffs, especially for low-volume parts. We then examine performance impact-limited fill insertion which goes beyond mere capacitance rules. Modeling, objectives, and filling strategies are discussed. Finally, we discuss current and near-term prospects for the overall design-to-manufacturing PD methodology. Key aspects include better integrations with analysis and manufacturing interfaces, as well as cost-benefit tradeoffs for "regular" layout structures that are likely beyond 90nm, cost optimizations for low-volume production, and the role of robust and/or stochastic optimization in PD.

1 Introduction

Moore's law continues to drive higher performance with smaller circuit features. Aggressive technology scaling has introduced new variation sources and made process variation control more difficult. As a result, semiconductor manufacturing equipment will be strained to maintain constant process variation levels in future technology nodes. Despite the relaxation of some 3σ tolerances, there are no known solutions for a number of near-term variability control requirements (according to the ITRS [5]). Consistent improvements in the resolution of optical lithography techniques have been a key enabler of the continuation of Moore's law. However, as minimum feature sizes continue to shrink, the wavelength of light used in modern lithography systems is no longer several times larger than the minimum line dimensions to be printed, e.g., today's 130nm CMOS processes use 193nm exposure tools. As a result, modern CMOS processes are operating in a sub-wavelength lithography regime.

Increasingly complex design rules impact detailed routing, physical verification, resolution enhancement (RET) and mask data preparation (MDP). The loss in design tool quality as well as design productivity have resulted in increased project uncertainty and manufacturing NRE. Designer, EDA, and process communities must cooperate and co-evolve to maintain the cost (value) trajectory of Moore's law. At 90nm to 65nm transition, this is a matter of survival for the worldwide semiconductor industry. There needs to exist a bidirectional design-manufacturing data pipe with cost and value as the fundamental drivers. Limits of mask flow need to be passed on to design while functional intent needs to be fed into the mask flow.

The next section briefly discusses PD complications due to deep-submicron manufacturing issues, notably subwavelength lithography and process variation. Section 3 outlines various elements of yield-constrained design optimization. Section 4 reviews area fill insertion methods that minimize performance impact. Last, Section 5 describes prospects for closer links between mask and process technology and the physical design flow.

2 Modern Manufacturing and Its Impact on PD

This section of the paper reviews recent PD complications, and impact on methodology, that arise from subwavelength lithography and deep-submicron manufacturing. We also discuss the mask NRE component of design cost, and taxonomize process variations and their sources for modeling and simulation. This leads us to a framework of design for cost and value.

2.1 Subwavelength Lithography

Optical lithography is being pushed to new extremes, with 193nm lasers currently used to fabricate devices with dimensions of 90nm or less. The extension of optical lithography has been enabled by several developments such as chemically amplified photoresists and anti-reflective coatings. By predicting physical phenomena (especially diffraction and interference) behind optical systems and systematically compensating for them, the minimum feature and pitch that can be resolved are significantly extended. These Resolution Enhancement Techniques (RETs) are aimed at three major optical wave components, namely, direction, amplitude and phase.

Off Axis Illumination (OAI) techniques direct light at the photomask only at certain angles. The combination of the angle and the pitch of features in the mask can enhance

*We would like to thank our collaborators, including members of the UCSD VLSI CAD Laboratory and Prof. Dennis Sylvester at the University of Michigan.

ICCAD'03, November 11-13, 2003, San Jose, California, USA.

resolution of certain pitches, particularly dense pitches and small lines [9]. The design rules are complicated by the fact that certain large as well as certain fine pitches are well reproduced but some intermediate pitches may not print as well. This leads to some forbidden pitches and sizes.

Optical Proximity Correction (OPC) makes small alterations to the layout features to reduce linewidth variation. Sub Resolution Assist Features (SRAFs) or scatter bars are added to the mask to allow isolated features diffract light as dense features but themselves do not print [8, 9]. Adding SRAFs to intermediate pitches can be tough resulting in suboptimal printing performance for these features. A concern here is that one might be OPCing the OPC due to legacy design rules [8] which lead to unnecessary constraints for designers as well as lithographers.

The third category, phase, is controlled by Phase Shifting Masks (PSM). Parts of the mask are etched to create phase difference between regions and can enhance resolution for certain features by up to a factor of two [8]. Generating phase-compliant layouts is a major problem for future physical design [8, 17].

An example consequence is the impact on routing algorithms. As we move into the nanometer regime, some of the requirements such as spacing rules, reliability rules and process antenna rules impose severe constraints on routing algorithms [19]. Some of these restrictive rules are listed below.

- *"Antennas"* are formed by metal traces that accumulate static charge during manufacturing. Without a safe discharge path (through the reverse-biased diode at the output stage of a logic gate) any connected gate may be damaged due to electrostatic discharge. "Antenna rules" establish maximum allowable ratios of metal area to gate area in the absence of discharge path. The pure router-based solution is *bridging* (layer-hopping) to limit the amount of metal connected to a gate; this creates more wiring, vias and congestion. The combined router- and library-based solution is to drop reverse-biased diodes (source-drain contacts) close to the gate, i.e., (ECO substitution of) dioded cell variants, with negative area and power implications. Tightening of antenna ratios has lowered completion rates of detailed routers and led to more antenna waivers [12].

  Should liberal use of dioded cells be required, there will be high costs with respect to chip area and power metrics as well as non-trivial balancing of two sources of yield loss: increased die area versus antenna damage.

- *Via stacking and minimum area rules* arise because stacking of vias through multiple layers can cause minimum area violations with respect to stacking-dependent alignment tolerances. thought. Signal routing layers are often divided into local layers, intermediate layers, and global layers. layers within the same group have same pitches and parasitics. At the highest layer of a given group, the overhang of the "up-via" can be significantly larger than that of the "down-via" [19]. In addition, use of multiple-cut via cells to increase BEOL yield is complicated by dependencies on the layers and wire segment widths to be connected.

- *Width- and length-dependent spacing* rules make minimum spacing a function of both wire width and length of parallel adjacencies. This means that edge costs during heuristic search are dependent on path history. Especially pernicious are *influence rules* (stub rules, halo rules), where a wide wire will influence the spacing rule within its surroundings. This results in strange jogs and spreading when wires enter an influenced area, as well as complicated ECO effects. Another aspect of reliability which is gaining prominence is *resist pattern collapse* [12, 38]. Resist features collapse upon formation at high aspect ratios. Pattern collapse probability is length-dependent. This contributes to length-dependent spacing rules: longer parallel runs of wires require more spacing. Inserting jogs in the routing can avoid such effects.

2.2 Process Variation: Taxonomy and Simulation

Aggressive technology scaling has introduced new variation sources and made process control more difficult. As a result, future technology nodes are expected to see increased process variation and decreased predictability of nanometer-scale circuit performance [5, 13, 2]. Variability impact extends beyond just performance. For instance, leakage power has exponential dependence on gate length and hence its variation. Circuit variability arises from chip fabrication or circuit operation itself.

Based on inherent spatial scales, parametric variations are separated into two categories [4]:*inter-die* and *intra-die* variation.

- *Inter-die variation* is the difference in the value of a parameter across nominally identical die. These may be fabricated on the same wafer, on different wafers or in different lots. It is mostly design independent and is related to equipment properties, wafer placement, processing temperatures, etc [28].

- *Intra-die variation* is the deviation occurring spatially within any die. Such variation can arise from wafer-level trends as well as layout-pattern dependencies. Fluctuations in channel doping, gate oxide thickness, and ILD permittivity are primarily due to random variation. This type of variation is likely to have spatial correlation, making nearby devices more similar than ones that are across the die from one another.

Both kinds of variation can have systematic and random components. An example of systematic wafer-scale variation is the bowl-shaped or slanted-plane nature of some deposition processes. Similarly, on a die-scale variation due to layout patterns which arise from design is a systematic and largely predictable effect. When simulating the impact of variability using Monte-Carlo or other means, few cautions are as follows.

1. *Accounting for systematic variation correctly.* If the nature of systematic variation is known, the nature of distribution of delay variation due to the systematic source can be accurately modeled. For instance, using a symmetric Gaussian distribution for a variation for some source which is known to follow a bowl shaped structure across the wafer is a modeling error as there are more die at the periphery of the wafer than at the center. Filtering out systematic within-die variation is also important [13].

2. *Correct composition of variations.* Variation is decomposed into its various components (inter-die, intra-die) using analysis of variance techniques [1]. It is important to understand how the decomposition is done to recompose them in analysis or simulation. For instance, does decomposition assume perfect independence or perfect correlation between inter and intra-die variations or

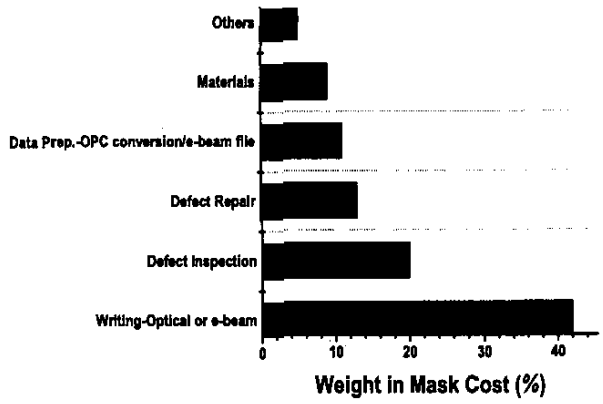

Figure 1: Relative contributions of various components of mask cost.

systematic and random components thereof is an important question to ask.

2.3 Mask Cost Model

With the growing complexity and expense of equipment, microlithography comprises over 30% of the cost of a new fabrication facility [30]. According to [32], the major contributors to mask cost are:

1. low mask yield (due to OPC and PSM as well as stringent CD requirements)
2. increased data preparation time
3. equipment cost
4. low equipment throughput

Figure 1 shows the major drivers of mask cost. The main drivers for increasing mask costs and turnaround time (TAT) include the increasing application of RETs and their higher write times. Variable-shaped electron beam mask writing combined with vector scanning (compared to traditional raster scanning, vector scanning allows features to be scaled up or down in size while maintaining sharpness but the write cost is proportional to feature complexity) is a widely used technique for high speed mask writing. In this method the input GDSII layout data is converted into the mask writer format (e.g. MEBES). After rule or model based OPC, the number of line edges is increased by 4-8X over a non-OPC layout, driving data volume up [31]. At the 90nm node, the data volume for a single mask layer of a design can approach 250GB, and mask writing times are strongly superlinear in data volume [34]. Mask writers are slowed by the software for e-beam data fracturing and transfer along with the extremely large file sizes involved. In [29] the authors show that, considering 500 wafer exposures per mask, the relative cost at the $0.13\mu m$ technology node is about 8.5X that o f the $0.35\mu m$ mask set. Total cost to produce low-volume parts such as most ASIC designs is dominated by mask costs [33]. Half of all masks produced are used on less than 570 wafers. At such low usages, high RET costs cannot be completely amortized and cost per die becomes very large.

2.4 Design for Value

Conventional design is based on the goal of performance optimization, or *design for performance* (DFP) [13]. As process variation inevitably leads to performance distributions, it implies the possibility of *design for value* (DFV) methodologies to maximize the yield. I Performance is measured by critical path delay T. It is a function of design variables x_i and process parameters y_i, i.e.

$$T = f(x_1, ..., x_m, y_1, ...y_n) \tag{1}$$

Design for performance seeks to find values of x_i to minimize T, given the nominal values of y_i, and ignoring process variations, i.e.

$$\text{Minimize } T \text{ s.t.} \quad y_i = y_{i\_nom} \tag{2}$$

Alternatively, worst-case values of y_i may be used. This is representative of a corner-based approach where all parameters are very pessimistically taken at their 3σ points, again within a deterministic framework.

We define *value* to be the total dollars earned from the chip that is sold on the open market. A value function $v(f)$ gives the market value of the chip for some performance measure f (e.g. speed, power). Thus, the total value of a given process is obtained as:

$$Value = \sum_f v(f) \times yield(f) \tag{3}$$

where $yield(f)$ has the usual meaning. Design for value seeks to find values of x_i to maximize yield of $T < T_m$, given the variability distributions of parameters y_i, where T_m is the target delay, i.e.

$$\text{Maximize } P_T(T_m) \text{ s.t.} \quad y_i = N(\mu_i, \sigma_i),\ \forall\ 1 \leq i \leq n \tag{4}$$

The two approaches of Equation 2 and Equation 4 may not be equivalent. This calls for research into such probabilistic optimizations, as well as efforts to quantify the potential value and costs associated with both manufacturing and design solutions to the process variability issue.

3 Yield Constrained Optimization in PD

This section of the paper covers yield-constrained optimizations in PD, especially "beyond corners" approaches that escape today's pessimistic or even incorrect corner-based approaches. Statistical timing and noise analyses enable optimization of parametric yield and reliability. Yield-aware cell libraries and "analog" design rules (as opposed to "digital", 0/1 rules) can help designers explore yield-cost tradeoffs, especially for low-volume parts.

A great deal of research effort has been spent on deterministic design optimization. Aggressive nominal optimization can harm the design yield or lead to drastic overdesign. One way to correct this is to model systematic and random effects and explicitly optimize for yield. Another way is to design variation-insensitive circuits. In current design methodologies, well-tuned circuits have a large number of equally critical paths. In the variation regime, the circuit delay is determined by maximum of path delay distributions. Circuit delay variance thus, increases with the number of equally critical paths. As a result the circuit yield at the desired selling point performance may be much less

than optimal even though the nominal optimum has been achieved. [20] present a sizing approach which penalizes having a large number of equally critical paths to avoid a high "wall" of critical paths. Two important components of a yield-based optimization are a statistical timer and usable models of variation in terms of libraries and design rules. We discuss these further in this section.

3.1 Statistical Timing Analysis

Traditionally, inter-die variation has been handled by case analysis and intra-chip variability is accounted for by heuristic derating factors which slow down data with respect to clock or vice-versa. This is implemented as linear combination of delays in IBM's EinsTimer and as an on-chip variation mode in Synopsys' PrimeTime [21]. There are a number of reasons due to which the conventional deterministic timing analysis paradigm is breaking down. With fast scaling critical dimensions, the variability in physical parameters is increasing. Temperature, IR drop and coupling noise induced variability, though systematic is difficult to analyze. Timing runs to incorporate all these effects are mandatory in current timing sign-off. Trying to worst-case all these variables gives rise to a formidable number of corners. Time to market constraints do not allow such exploding number of static timing runs for timing verification. Besides having feasibility issues, corner-based analysis suffers from pessimism but at the same time can be not "pessimistic enough" because it is impossible to exhaustively enumerate all possible cases. A simple but typically overlooked example is that of computing worst-case clock skew. If two clock paths are not *identical* and do not track perfectly, then the peak skew between them may not occur at any of the delay corners. This problem, which is the subject of some of our current research, is only further complicated by intra-chip variability.

The solution to these problems lies in statistical timing analysis, which if implemented correctly can reduce pessimism as well as improve timing verification turnaround time. Yield loss can be classified as *catastrophic* or functional (occuring due to dust particles and other random defects during manufacturing and rendering the chip nonfunctional) versus *parametric* (causing the chip to function with a range in performance figures of merit). Statistical timing predicts parametric yield. A statistical timer essentially propagates probability distributions. The probability distributions may be of sources of variation (such as L_{eff}, tox ,etc) or of the performance measure (such as delay) itself. It also has to take into account correlations between these distributions. The output is a circuit delay distribution or a distribution of slacks at each of the outputs. The major bottleneck in statistical timing analysis are efficiently calculating the maximum of two correlated probability distributions. Approaches in literature range from smart Monte-Carlo [24] to bounding distributions [25, 23].

An important improvement in deterministic as well as statistical timing analysis can be accounting for systematic variation. Example of such variation can be through pitch and through focus CD variation. Pattern dependent linewidth variation arising out of iso-dense bias is predictable after placement. A limited but representative set of environments for each cell can be simulated for printed wafer image. The results can be used to predict the CD of each gate instance and hence the delay of each timing arc in the design. An in-context timing analysis can then be performed to yield better accuracy as well as achieve tighter statistical distributions (as through pitch variation is the major component of CD variation). Preliminary results of work at IBM [27] indicate a 8-10% difference in timing result. Similar work targeting mainly the interconnect delay is given in [39]. Moreover, a library-based OPC approach as motivated in [26] can further this approach by accounting for RET before physical design. Preliminary work at IBM on the 90nm ASIC library suggests that the difference in printed CD of gates with library-based pre-layout OPC and post-layout extensive model-based OPC is minimal.

3.2 Yield-Aware Libraries and Design Rules

The usual medium of communication between design and process development is through a set of design rules. These rules in presence of RET and other manufacturing constraints are becoming overly restrictive. Given adequate models of MDP, RET and Litho flows, design tools can and should optimize parametric yield, $/wafer, profits. The prerequisites for such interesting optimizations are "analog" rules and yield-aware libraries. Current design rules are hard constraints which result into a nightmarish number of yes/no checks at the physical verification stage. Instead what is required are degrees of meeting various design rules with an indication of the corresponding yield penalty. Design rules are decided by the process community which is oblivious of the design requirements. This results in huge pessimism in coming up with these design rules. A cognizance of yield loss mechanisms (both parametric as well as catastrophic) is required in design so that penalty of *not* being 100% design-rule correct can be computed. An example of such a pessimistic rule is metal density constraint, which is the same all throughout the design. A metal density versus copper thickness tradeoff curve can be more useful for a design tool which can then use area fill more intelligently based on timing criticality of nets.

Yield-aware timing libraries are absolutely necessary for statistical timing analysis. These libraries can range from specifying simple (μ, σ) pairs for each timing arc (assuming Gaussian distribution of performance) to specifying complex distributions that decompose variability into its various components (systematic versus random, intra-die versus inter-die) and various correlations between them. The complexity of the yield library is a tradeoff between complexity of process characterization as well as implementation of the statistical timer and accuracy of the result. Care is required to pick the right point on this tradeoff curve so that after statistical timing, the result is not meaningless (i.e., an incorrect distribution instead of incorrect corners points).

Another aspect of designing future library design is making them variation aware. For instance, current industry practice is to design entire standard cell libraries with minimum width critical poly permissible by the technology. The advantage is better performance to area ratio. Here the important thing to note is that CD variation due to lithography is "absolute". Therefore, intentional increase in L_{eff} can lead to cells which are less susceptible to variation. Such cells would have poorer performance to area ratio but better leakage and predictability characteristics. A select number of variation-sensitive cells in the design can be replaced with these more reliable versions from the library. Moreover, certain transitions in a cell will tend to be better controlled. For example, a transition involving only PFET fingers of an inverter which are densely packed will tend to time close to nominal (since the lithography process is better controlled for tighter pitches). The relative sensitivities of various timing arcs within a cell can be used in the synthesis flow to yield less variation-sensitive designs. Finally, it is also of

interest to generate cell libraries which are conducive to application of post-tapeout RET [6].

4 Performance Impact Limited Fill Insertion

Chemical-mechanical planarization (CMP) and other manufacturing steps in nanometer-scale VLSI processes have varying effects on device and interconnect features, depending on local attributes of the layout. To improve manufacturability and performance predictability, foundry rules require that a layout be made uniform with respect to prescribed density criteria, through insertion of *area fill* ("dummy") geometries. *Chemical-mechanical planarization* (CMP) and other manufacturing steps in nanometer-scale VLSI processes have varying effects on device and interconnect features, depending on local attributes of the layout. To improve manufacturability and performance predictability, foundry rules require that a layout be made uniform with respect to prescribed density criteria, through insertion of *area fill* ("dummy") geometries.

Work at MIT Microsystems Technology Laboratories [16] proposes a rule-based area fill methodology. To minimize the added interconnect capacitance resulting from fill, a dummy fill design rule is found by modeling the effects on interconnect capacitance of different design rules (which are consistent with the fill pattern density requirement). Work at Motorola by Grobman et al. [7] points out that the main parameters to influence the change in interconnect capacitance due to fill insertion are feature ("block") sizes and proximity to interconnect lines. The larger the size of the block, the larger the consequent interaction between interconnect lines. Similarly, the closer blocks are to interconnect lines, the stronger their interaction will be. Lee et al. [15] describe the methodology used at Samsung for chip-level metal fill modeling. Their approach replaces the metal fill layer by an effective (i.e., equivalent) high-k dielectric. The increments of capacitance due to floating metal fill are dependent on the signal line width and spacing, inter-metal dielectric thickness and permittivity, density of metal fills, metal fill feature size, and metal layer thickness. RC extraction results in [15] show that the total interconnect capacitance increase can be up to 15% for some nets in an $0.18\mu m$ design. Thus, floating dummy metal fills should be included in chip-level RC extraction and timing analysis to avoid timing errors.

The first work aimed at true performance aware fill insertion was presented in [10]. The authors give the first formulations of the Performance Impact Limited Fill (PIL-Fill) problem with the objective of either minimizing total delay impact (MDFC) or maximizing the minimum slack of all nets (MSFC), subject to inserting a given prescribed amount of fill. Using Integer Linear Programming as well as efficient greedy techniques, [10] achieve up to 90% reduction in total delay overhead of dummy fill. In the MSFC formulation, to maximize minimum slack over all nets in the post-fill layout, we propose an Iterated Greedy approach based on iterations between the static timing analysis (STA) tool and the area fill synthesis. Capacitance impact due to fill feature insertion during area fill synthesis is written in Reduced Standard Parasitic Format (RSPF) as a file input to STA tool. The results given in [10] suggest that performance awareness in fill insertion can result in significant timing improvements without compromising layout density. This work ignores fill impact on fringing and overlap capacitances. A more accurate fill insertion method which has cognizance of multiple layers is still an open question for research.

A caveat here is that all fill insertion approaches currently rely on foundry-specified minimum and maximum layout densities. These rules are not aware of design as a result of which they tend to be pessimistic. Accurate CMP models as well as levels of density control can help in less pessimistic and timing correct fill insertion.

5 Manufacturing-Aware Physical Design Futures

In this section we describe futures for manufacturing-aware PD. This includes cost-benefit tradeoffs for "regular" layout structures that are likely beyond 90nm, cost optimizations for low-volume production, and the role of robust and/or stochastic optimization in PD. Future physical design needs to be tied more closely to manufacturing. Several novel objectives may need to be considered. For instance, "fracturing-aware design" may be beneficial, whereby OPC, phase-shifter, and functional feature shapes are chosen or perturbed for reduced shot count. Layouts can also be stretched (via insertion of submicron-scale "dead space") to help definition of major field boundaries (or, soft field boundaries) for mask writing. Our current research develops such optimizations. More complex extraction and characterization capabilities may also be required. For example, extraction and characterization of nonuniform poly CD can yield better estimates of timing and leakage.

5.1 Manufacturing Aware PD Design Flow

Traditionally, design, mask making and process engineering have depended on rule sets to isolate themselves from having to understand one another's technology. With number and complexity of these design rules exploding and ever decreasing yields, the traditional isolated deterministic design paradigm is breaking down. Close interaction between manufacturing, mask and design communities is inevitable. Figure 2 shows a near-term design to manufacturing flow.

5.2 Regular Layout Fabrics

As uncertainty increases and design guardbanding approaches ridiculous extents, it is increasingly important to ensure predictable printability. New solutions being explored at the design end include regularity. Full chip layouts may need to be assembled as a collection of regular *printable* patterns for technologies beyond 90nm. 65 nm has high likelihood for layouts to look like regular gratings: uniform pitch and width on metal as well as poly layers. Predictable layouts even in presence of focus and dose variations may be required.

Several regular layout fabrics have been proposed with varying degrees of performance overhead and flexibility. FPGAs are very flexible but suffer from huge area, power and performance overheads. Via programmable gate arrays [22] offer programmability of logic as well as interconnect using vias and contacts. These offer performance, power and area closer to ASICs as they do not have complex SRAM based programmable logic as FPGAs. Other examples of regular fabrics include Fishbone routing scheme and River PLAs [14].

Somewhat less restrictive regularity can be achieved by more manufacturable cell libraries with regular structures which will require supportive placement techniques. For instance, intricate whitespace management may be required to ensure that intended spatial regularity is achieved.

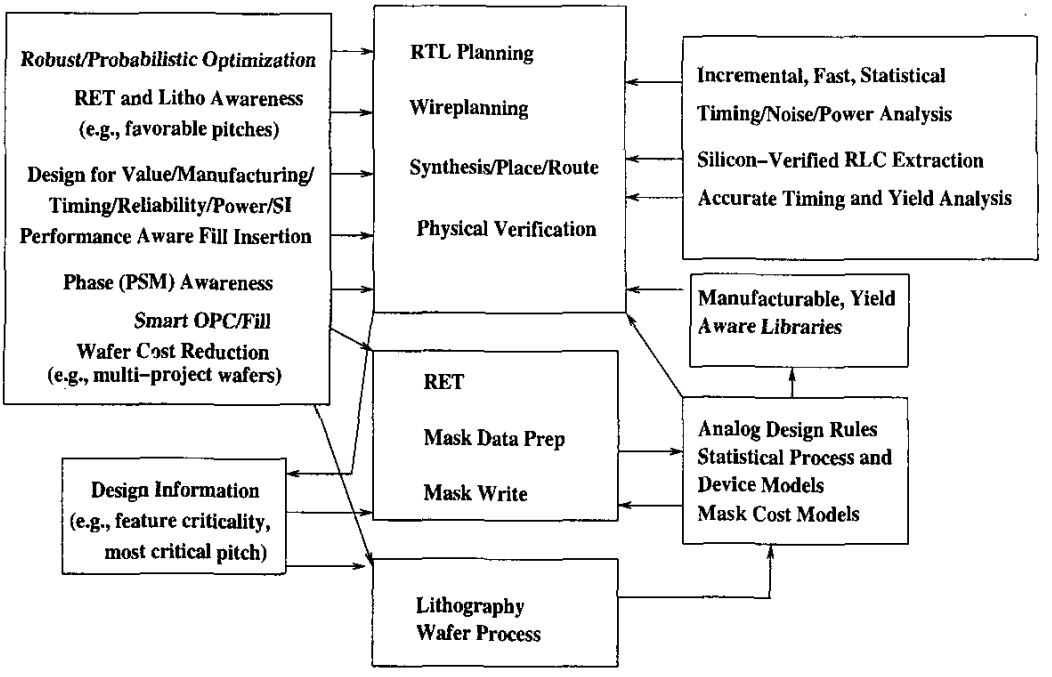

Figure 2: Near-Term Design-to-Manufacturing Flow

5.3 Driving Lithography with Physical Design

Traditionally, no concept of function is injected into the mask flow. Thus mask writers work equally hard in perfecting a dummy fill shape, a piece of the company logo, a gate in a critical path, and a gate in a non-critical path; errors in any of these shapes will trigger rejection of the mask in the inspection tool. The result is unduly low mask throughput and high mask costs. [11] is the first (and probably the only) work to explore the yield-cost tradeoffs for RET techniques. With their simple sizing techniques, [11] claims to achieve up to 70% reduction in cost of OPC. OPC in the work is classified into three "levels" of varying costs and accuracy and the work seems to be more applicable in library-based OPC context as in [26].

Current commercial OPC tools base their correction methodology on edge placement errors or EPEs rather than true linewidths. As mentioned in [26], better measures of "designer's intent" (for the polysilicon layer) are CD (width of poly over active region) and contact coverage (i.e. area of overlap between contact and poly). Given this context, a useful extension to the work in [11] is an algorithm to analyze a design and output EPE for every edge in the layout to meet given yield or guardbanding constraint. These instance-specific EPEs can then be used in the OPC flow for faster, less pessimistic OPC. Another related objective of physical design can be to maximize the minimum CD tolerance over the whole layout. This can lead to better process window for lithographers as the process window is predominantly determined by the CD tolerance specification. Typically, a process is tuned to print a particular pitch very well. Moreover, this tuned pitch may be changed (for example by changing the nominal exposure) on a design-to-design basis. Physical design tools can then help choose the most critical pitch in the design which needs the most predictable and accurate printability.

5.4 Novel Optimization Techniques

To handle uncertainties involved in design and manufacturing, new robust as well stochastic optimization techniques need to be explored. Robust as well as stochastic optimization has been studied in literature, especially in control theory [35]. Several formulations of the probabilistic optimization problems exist. Two of the most common ones are chance constrained problem and stochastic programming with recourse. The former typically looks as follows.

$$\begin{aligned} & min\ c^T x \\ & s.t. \\ & A_0 x = b_0 \\ & P(A_i x \geq h_i) \geq \beta_i \end{aligned} \tag{5}$$

Recourse formulations are multistage and work by penalizing constraints which are not met. [35] gives the stochastic formulation of the linear programming problem where the coefficient matrices are random variables with known statistics.

Another yet-to-be explored area in optimization in CAD is designing variation-insensitive solutions to design problems. An example to note here can be designing variation-insensitive power distribution networks. A power distribution network obeys $GV = I$ where G is the conductance matrix. Let $\| A \|_\infty$ denote the infinity norm of the matrix A . $\| A \|_\infty = max_i |a_i|$ if A is a vector. When A is a matrix $\| A \|_\infty = max_i \sum_{j=1}^{n} |a_{ij}|$. If V_{dd} is set to 0V, then the peak IR drop is given by $\| V \|_\infty$ where V is the solution to $GV = I$

For perturbation matrices E, e such that $(G + E)V' = I + e$ gives the perturbed solution to $GV = I$. An upper-

bound on $V' - V$ can be easily derived [36].

$$\frac{\| V' - V \|}{\| V \|} \leq \| G \| \| G^{-1} \| \left(\frac{\| E \|}{\| G \|} + \frac{\| e \|}{\| I \|} \right) \qquad (6)$$

Note that E has to maintain the structure of the conductance matrix. $\| G \| \| G^{-1} \|$ is referred to as the *condition number* of G. It is an indicator of robustness of the solution of the IR drop solution with respect to small variations in the conductance matrix as well as currents. Systematic perturbation of the power mesh conductance matrix G to yield better condition number can lead to more robust power distribution networks. Such probabilistic and robust optimization methods can be key to future CAD algorithms.

References

[1] B.E. Stine, D.S. Boning and J.E. Chung, "Analysis and Decomposition of Spatial Variation in Integrated Circuit Processes and Devices", *IEEE Transactions on Semiconductor Manufacturing*, 10(1), 1997, pp. 24-41.

[2] K.A. Bowman, S.G. Duvall and J.D. Meindl, "Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution for Gigascale Integration", *IEEE Journal of Solid-State Circuits*, 37(2), 2002, pp. 183-190.

[3] M. Orshansky, L. Milor, P. Chen, K. Keutzer and C. Hu, "Impact of Systematic Spatial Intra-Chip Gate Length Variability on Performance of High-Speed Digital Circuits", *Proc. IEEE/ACM ICCAD*, 2000, pp. 62-67.

[4] D. Boning and S. Nassif, *Design of High-Performance μP Circuits, Chapter 6: Models of Process Variation in Device and Interconnect*, 2001, pp. 98-116.

[5] *2001 International Technology Roadmap for Semiconductors (ITRS)*, http://public.itrs.net.

[6] J.A. Torres, D. Chow, P. de Dood and D.J. Albers, "RET Compliant Cell Generation for sub-130nm Processes", Mentor Graphics Whitepaper, June 2002.

[7] W. Grobman, M. Thompson, R. Wang, C. Yuan, R. Tian and E. Demircan, "Reticle Enhancement Technology: Implications and Challenges for Physical Design", *Proc. IEEE/ACM DAC*, 2001, pp. 73-78.

[8] F.M. Schellenberg, L. Capodieci and B. Socha, "Adoption of OPC and the Impact on Design and Layout", *Proc. IEEE/ACM DAC*, 2001, pp. 89-92.

[9] F.M. Schellenberg and L. Capodieci, "Impact of RET on Physical Layouts", *Proc. IEEE/ACM ISPD*, 2001, pp. 52-55.

[10] Y. Chen, P. Gupta and A.B. Kahng, "Performance-Impact Limited Area Fill Synthesis", *Proc. IEEE/ACM DAC*, June 2003, pp. 22-27.

[11] P. Gupta, A.B. Kahng, D. Sylvester and J. Yang, "A Cost-Driven Lithographic Correction Methodology Based on Off-the-Shelf Sizing Tools", *Proc. IEEE/ACM DAC*, June 2003, pp. 16-21.

[12] A.B. Kahng, "Research Directions for Coevolution of Rules and Routers" (invited paper), *Proc. IEEE/ACM ISPD*, April 2003, pp. 122-125.

[13] Y. Cao, P. Gupta, A.B. Kahng, D. Sylvester and J. Yang, "Design Sensitivities to Variability: Extrapolation and Assessments in Nanometer VLSI", submitted to *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems* 2003.

[14] F. Mo and R.K. Brayton, "Regular Fabrics in Deep Sub-Micron Integrated-Circuit Design" *International Workshop on Logic and Synthesis*, June 2002.

[15] W.S. Lee, K.H. Lee, J.K. Park, T.K. Kim, Y.K. Park, and J.T. Kong, "Investigation of the Capacitance Deviation Due to Metal-Fills and the Effective Interconnect Geometry Modeling", *Proc. IEEE International Symposium on Quality Electronic Design*, 2003, pp. 373-376.

[16] B.E. Stine, D.S. Boning et al., "The Physical and Electrical Effects of Metal Fill Patterning Practices for Oxide Chemical Mechanical Polishing Processes", *IEEE Trans. on Electron Devices* 45(3) (1998), pp. 665-679.

[17] A.B. Kahng, "Subwavelength Design: Lithography Effects and Challenges" (tutorial), *Proc. IEEE International Symposium on Quality Electronic Design*, 2000.

[18] L.W. Liebmann, "Layout Impact of Resolution Enhancement Techniques: Impediment or Opportunity?", *Proc. IEEE/ACM ISPD*, 2003, pp. 110-117.

[19] H.K.-S. Leung "Advanced Routing in Changing Technology Landscape", *Proc. IEEE/ACM ISPD*, 2003, pp. 118-121.

[20] X. Bai, C. Visweswariah, P. N. Strenski, and D. J. Hathaway, "Uncertainty-Aware Circuit Tuning", *Proc. IEEE/ACM DAC* 2002, pp. 58-63.

[21] C. Visweswariah, "Death, Taxes and Failing Chips", *Proc. IEEE/ACM DAC*, 2003.

[22] L. Pileggi, H. Schmit, A.J. Strojwas, P. Gopalakrishnan, V. Khetrapal, A. Koorapaty, C. Patel, V. Rovner and K.Y. Tong, "Exploring Regular Fabrics to Optimize the Performance-Cost Trade-Off", *Proc. IEEE/ACM DAC*, 2003.

[23] M. Orshansky and K. Keutzer, "A General Probabilistic Framework for Worst Case Timing Analysis", *Proc. IEEE/ACM DAC*, 2002.

[24] J.-J. Liou, A. Krstic, L.-C. Wang and K.-T. Cheng, "False-Path-Aware Statistical Timing Analysis and Efficient Path Selection for Delay Testing and Timing Validation", *Proc. IEEE/ACM DAC*, 2002, pp. 566-569.

[25] A.B. Agarwal, D. Blaauw, V. Zolotov and S. Vrudhula, "Statistical Timing Analysis using Bounds and Selective Enumeration", Proc. ACM/IEEE TAU, 2002, pp. 29-36.

[26] P. Gupta, F.-L. Heng and M. Lavin, "Merits of Cell-Based OPC", submitted to *SPIE Conf. on Design and Process Integration for Microelectronic Manufacturing.*

[27] F.-K. Heng, P. Gupta, R. Gordon, K. Lai and J. Lee, "Taming Pattern and Focus Dependent Variation", submitted to *SPIE Conf. on Design and Process Integration for Microelectronic Manufacturing.*

[28] K.A. Bowman, S.G. Duvall and J.D. Meindl, "Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution", *Proc. IEEE Int. Solid-State Circuits Conf.*, 2001, pp. 278-279.

[29] W. Carpenter, "International SEMATECH: A Focus on the Photomask Industry", http://www.kla-tencor.com/company_info/magazine/autumn00/Inter_SEMATECH_photomaskindustry_AutumnMag00-3.pdf.

[30] B. Bruggeman, et al., "Microlithography Cost Analysis", *Interface Symposium*, 1999.

[31] S. Murphy, "SEMATECH: Mask Supply Workshop", Dupont Photomask, *SEMATECH: Mask Supply Workshop*, 2001.

[32] SEMATECH: Mask Supply Workshop, 2001.

[33] M.L. Rieger, J.P. Mayhew, and S. Panchapakesan, "Layout Design Methodologies for Subwavelength Manufacturing", *Proc. IEEE/ACM DAC*, pp. 85-92, 2001.

[34] P. Buck, *ISMT Mask-EDA Workshop*, Dupont Photomasks, 2001.

[35] H.A. Hindi and S.P. Boyd, "Robust Solutions to l_1, l_2, and l_∞ Uncertain Linear Approximation Problems using Convex Optimization", *Proc. American Control Conference*, 6:3487-3491, 1998.

[36] G.W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press Inc., 1990.

[37] Y. Ermoliev, R.J.-B. Wets, *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, 1988.

[38] G. Czech, E.C. Richter and O. Wunnicke, "193nm Resists: A Status Report (Part One)", *Future Fab*, Volume 12, *http://www.future-fab.com*

[39] V. Mehrotra, S.L. Sam, D. Boning, A. Chandrakasan, R. Vallishayee and S. Nassif, "A Methodology for Modeling the Effects of Systematic Within-Die Interconnect and Device Variation on Circuit Performance", *Proc. IEEE/ACM DAC*, 2000, pp. 172-175.