Self-Compensating Design for Reduction of Timing and Leakage Sensitivity to Systematic Pattern-Dependent Variation

Puneet Gupta, *Student Member, IEEE*, Andrew B. Kahng, *Senior Member, IEEE*, Youngmin Kim, *Student Member, IEEE*, and Dennis Sylvester, *Senior Member, IEEE*

Abstract-Critical dimension (CD) variation caused by defocus is largely systematic with dense lines "smiling" through focus while isolated lines "frown." In this paper, we propose a new design methodology that allows explicit compensation of focus-dependent CD variation, in particular, either within a cell (self-compensated cells) or across cells in a critical path (self-compensated design). By creating iso and dense variants for each library cell, we can achieve designs that are more robust to focus variation. Optimization with a mixture of dense and iso cell variants is possible, both for area and leakage power in timing constraints (critical delay), with the latter an interesting complement to existing leakage-reduction techniques, such as dual-Vth. We implement both a heuristic and mixed-integer linear-programming (MILP) solution methods to address this optimization and experimentally compare their results. Results indicate that designing with a self-compensated cell library incurs 12% area penalty and 6% leakage increase over a baseline library while compensating for focus-dependent CD variation (i.e., the design meets timing constraints across a large range of focus variation). We observe 27% area penalty and 7% leakage increase at the worst case defocus condition using only single-pitch cells. The area penalty of circuits after using both the heuristic and MILP optimization approaches is reduced to 3% while maintaining timing. We also apply the optimization to leakage, which traditionally shows very large variability due to its exponential relationship with gate CD. We conclude that a mixed iso/dense library that is combined with a sensitivity-based optimization approach yields much better area/timing/leakage tradeoffs than using a self-compensated cell library alone. Selfcompensated designs show 25% less leakage power on average at the worst defocus condition compared to a design employing a conventional library for the benchmarks studied.

Index Terms—Across-chip linewidth variation (ACLV), design for manufacturability, focus, leakage, self-compensating, systematic variation.

I. INTRODUCTION AND PRIOR WORK

W ITHIN-DIE process variation has become one of the most important considerations in IC manufacturing, particularly as lithography moves into the deeply subwavelength regime [2], [3]. Variation can occur at the fabrication stage (intrinsic variation) or during circuit operation (dynamic

Digital Object Identifier 10.1109/TCAD.2007.895759

variation) [1]. There are two major components to intrinsic variation: random and systematic [1], [2], [5]. Because of the strong layout dependence of the systematic component, estimation of systematic variation is impossible until layout information is available. Due to numerous variation sources and their interactions, systematic variation is difficult to predict and often treated as random.

Effective channel-length (L_{eff}) variation is one of the biggest determinants of IC performance [3]. Prohibitive increases in the cost of process control necessitate relaxed control of L_{eff} from a manufacturing perspective, shifting the focus to a more proactive management of L_{eff} variation from a design perspective. Across-chip linewidth-variation (ACLV) control is critical to the timing and functionality of a design [4]. Various resolution enhancement techniques (RETs), such as subresolution assistant feature (SRAF), optical proximity correction (OPC), and phase-shifting mask are commonly used to achieve ACLV control and larger process window in current design-tomanufacturing flows [8], [15], [16].

One of the major sources of $L_{\rm eff}$ variation is focus. Focus variations can occur, for example, due to changes in wafer flatness or lens imperfections. Defocus is a key parameter that defines the process window along with exposure dose. For a fixed exposure dose, only a limited value of depth of focus is acceptable to print a CD within a certain value of variation tolerance. Traditional corner-case timing-analysis flows are very pessimistic in worst casing focus impact on critical dimensions. This is because layout pitch and focus have very systematic interactions, as shown by so-called Bossung plots.

There have been a number of papers studying patterndependent variability. In particular, Orshansky et al. [4], [24] examined the characterization and impact of systematic spatial gate-length variation on the performance of circuits. The authors claim that the systematic spatial intrachip CD variability, rather than pattern-dependent proximity effects, are the primary cause of circuit-delay variation and speed degradation. They classified all gates into 18 different categories based on the orientation (vertical or horizontal) and spacing between neighboring polylines (i.e., dense, denso, and isolated) to capture the interaction between the optical-lithography process and locallayout patterns. Dense is defined as minimum spacing, denso represents an intermediate distance, and all others are labeled as isolated. Lgate values are then measured from the testchips to build Lgate maps. The Lgate maps containing spatial information of each gate are fed into a tool to generate modified

Manuscript received July 17, 2006. This paper was presented in part at the Design and Automation Conference, Anaheim, CA, June 13–17, 2005. This paper was recommended by Associate Editor F. N. Najm.

P. Gupta is with Blaze DFM Inc., Sunnyvale, CA 94089 USA.

A. B. Kahng is with the University of California San Diego, La Jolla, CA 92093 USA.

Y. Kim and D. Sylvester are with the University of Michigan, Ann Arbor, MI 48109-1340 USA (e-mail: kimyz@umich.edu).



Fig. 1. Idealized Bossung plot representing the systematic and opposing behavior of dense and iso patterns through defocus.

netlists depending on the location of the gates. Results show about 17% of critical-path-delay variation and up to 25% of timing error and performance loss without proper consideration of systematic spatial gate-length variation components.

Gupta and Fook-Luen [6] also study systematic variation and propose a novel static-timing-analysis approach that exploits the effect of proximity effects and focus-dependent CD variation. The premise of their technique is that, since throughpitch and through-focus variation in a chip are systematic, they can be considered into the static-timing analysis. As observed in the so-called Bossung plots (Fig. 1), linewidth decreases under defocus conditions for larger spacings (audio frequencies can partially combat this, as will be discussed later) and linewidth increases for dense patterns. Results in the study in [6] indicate that there is about 10% variation in the target linewidth as the pitch decrease and the through-focus variation can take up 30% of total ACLV. The authors compare the new systematic-variation-aware timing considering both throughpitch and through-focus systematic variations to the standard corner-based timing and find reductions in timing uncertainty of up to 40%.

II. COMPENSATING FOCUS-DEPENDENT CD VARIATION

Systematic variation can be mitigated to some extent by performing OPC and inserting assist features but cannot completely be eliminated due to various reasons (modeling errors, algorithmic inaccuracies, process variations, etc.). The remaining linewidth variation due to layout is significant even after the use of complex RET techniques, with isolated and dense lines retaining opposite behavior under varying defocus [6]. Thus, there is a possibility of compensating for systematic variation in the design itself. This compensation can be achieved in two ways: 1) by ensuring that each standard cell is robust against focus variation and 2) by intelligently constructing a robust circuit out of inherently nonrobust building blocks or cells.

A. Self-Compensated Cell Layout

By self-compensated cell layout, we refer to a correct-byconstruction methodology that relies on within-cell compensation of CD variation caused by focus variation. For example, variation can be compensated in series-connected NMOS, if one device becomes shorter (thus, faster) under defocus and the other device becomes longer (thus, slower). This can be achieved by making one device "iso" and the other device "dense." The other way of generating self-compensated cells is to find spacing ranges in which the linewidth variation is negligible by focus variation. Each spacing between adjacent polylines should be one of these values. In this paper, we generate all the self-compensated cells by requiring polyspacing to be in the compensated spacing range (to be discussed further in the next section). We also explore the possibility of single pitched-cells where all polyspacings are set to one highly manufacturable value to eliminate the focus-dependent CD variation inside cells.

B. Self-Compensated Physical Design

This refers to compensation across cells (e.g., along a critical path). Consider two cells G1 and G2 that lie on the critical path G1 \longrightarrow G2. Focus variation, if not corrected by applying expensive RETs, can cause variation in critical-path delay and lead to potential timing failures or parametric yield loss. However, if G1 is explicitly made "iso" while G2 is made to act "dense," focus-dependent CD variation can be compensated. Assuming that iso and dense versions of library cells are available, designs that are robust to focus variation become possible.

In this paper, we compare and contrast the two approaches described above. For example, we seek to compare the area overheads of self-compensated libraries versus across-cell optimizations. We generate each cell variant based on lithography simulation, and the area overhead is then determined using place-and-route step. A sensitivity-based heuristic optimization approach for the self-compensating design for timing and area aspect was proposed in [23]. This heuristic algorithm can also be applied to reduce leakage while ensuring timing is met, which is explored in this paper. We also propose an integer linear-programming (ILP) formulation to ensure that timing is met across the expected focus range, and these results also allow us to determine the nonoptimality of the heuristic approach.

The remainder of this paper is organized as follows. Section III describes the construction of a cell library that consists of each version of cells, and Section IV describes self-compensating design techniques in area and leakage. We present experimental results in Section V, and Section VI provides conclusions.

III. LAYOUT GENERATION

The work in [23] is based on the lithographic simulation results after OPC and SRAF insertion using Calibre WorkBench (WB) [10]. Critical dimensions at every space and focus level are obtained from the five-line patterns. However, no layouts are actually generated for iso, dense, and self-compensated cell variants. In that case, the area of each cell and its parasitics were estimated based on the deviation from the original layout spacings to new (iso/dense/self-compensated) spacings. To obtain better estimates of delay and area after placement and routing, we generate each version of cells using an automated-layoutgeneration tool [21].

 TABLE I

 PARAMETERS USED IN CALIBRE WB [10]





Fig. 2. Linewidth variation with spacing (SBs are inserted at 420 and 660 nm).

A. Lithography Simulation

Lithography parameters used in Calibre WB [10] are shown in Table I. We use an optical lithography process with 248-nm wavelength and numerical aperture (NA) of 0.7. Optical models are generated at five different defocus levels (e.g., 0.0, 0.1, 0.2, 0.3, and 0.4 μ m) and a constant threshold resist model is used. We assume that the optical characteristics are symmetric in defocus (i.e., +0.1 μ m = -0.1 μ m defocus).

B. CD Measurement

To find a specific spacing range for iso, dense, selfcompensated, and single-pitched cells, we perform lithography simulation after OPC and SRAF insertion at two defocus values, namely, best and worst defocus values. The resulting printed linewidths are then measured. $L_{\rm eff}$ variations at the worst defocus value (0.4 μ m) are used to construct the criteria for spacing range of each variant of cells (i.e., iso, dense, selfcompensated, and single-pitched cells). The linewidth variation with spacing from 180 to 660 nm at 0.0- and 0.4- μ m defocus level is shown in Fig. 2. The 3-D graph with different left and right spacing from 180 to 630 nm with 50-nm step is shown in Fig. 3. As can be seen from these graphs, due to the use of scattering bars, the linewidth does not vary much at best focus even if the spacing between polylines increase. The tolerance of the self-compensated devices is set at 4 nm, since the 3σ for the gate CD control is 4 nm in 130-nm technology [11]. The first scattering bar is inserted at the spacing of 420 nm, and the second scattering bar is inserted when the spacing becomes 660 nm. Therefore, we define allowable spacings for dense



Fig. 3. Linewidth variation with asymmetric spacing for two defocus values, 0.0 and 0.4 μ m. The nearly flat surface represents 0.0- μ m defocus.

TABLE II Spacing Criteria for Cell Generation (SB = Scattering Bar)

Cell version	Spacing range (nm)
Dense	180(min.), 420(1sb), 660(2sb)
Iso	380 ~ 410, 600 ~ 650
Self-compensated	260 ~ 320, 460 ~ 480
Single-pitched	480

devices as 180 (minimum space), 420 (first scattering bar point), and 660 nm (second scattering bar point). We define 380–410 nm and 600–650 nm as the iso spacing range, and 260–320 nm and 460–480 nm as the self-compensated spacing regions. Finally, we select 480 nm as the spacing value for single-pitched cells from the self-compensated regions, because the minimum spacing for contact is 420 nm. Table II summarizes the spacing criteria for cell generation. We can set our intended spacing of polygates within technology files (ProTech [21]) to make each desired version of the library cells. A lumped-C model of capacitance is extracted and added into netlists to obtain more exact timing; to this end, we use a commercial parasitic-extraction tool (Calibre PEX [10]).

To analyze iso/dense/self-compensated behavior with defocus, we use a five-line pattern and sweep the spacing between the three center lines from 180 to 480 nm. Scattering bar insertion and OPC are performed on these patterns using Calibre [10]. The average linewidth of the center line is then measured for each pattern. Fig. 4 shows the variation in this critical dimension for different spacing values at nine different defocus values. The figure shows distinct space ranges where the patterns behave as iso, dense, or self-compensated.

Based on Figs. 3 and 4, we generate a lookup table (LUT) using the function CD = f(LS, RS, F), where LS is the left space, RS is the right space, and F is defocus. This allows us to obtain the exact degree to which specific patterns act isolated, dense, or self-compensated and also to predict CD-given defocus and spacings. The tolerance of the self-compensated devices is set at 4 nm. Thus, if linewidths are 4 nm larger than nominal at 0.4- μ m defocus, we assume those patterns are "dense"; similarly, if linewidths are 4 nm smaller than nominal, we classify the patterns as "iso." Finally, if the CD variation



Fig. 4. Linewidth variation with defocus level (nominal linewidth = 130 nm).

is less than 4 nm at 0.4- μ m defocus, we consider the pattern "self-compensated." The first scattering-bar insertion point is at a spacing of 420 nm, therefore, the "most iso" pattern has a spacing of roughly 400 nm. At 420-nm spacing and above, the pattern reverts to "dense" behavior as a result of scattering bar insertion. At the "most dense" spacing (180 nm on each side), the linewidth increases 13% from nominal and in the "most iso" case (i.e., 400 nm on each side), the linewidth decreases 11% from nominal at the 0.4- μ m defocus point.

The optimal scattering bar placement and width depend on numerous factors such as wavelength (λ), NA, illumination type, and others [12]. Reference [13] provides equations for optimal size and placement (defined as SRAF to main pattern spacing) of scattering bars, which are $(0.2 \sim 0.25) * (\lambda/NA)$ and $(0.55 \sim 0.75) * (\lambda/NA)$, respectively.

C. Edge Devices

Special consideration is required for the edge devices. Edge devices are those devices that are closest to the standard cell boundary. For example, since there is only one polyline for NMOS and PMOS in an INVX1 (minimum-sized inverter) layout, these are all edge devices. We identify two different types of edge devices: case 1 has no neighboring devices on either side (as in INVX1), while case 2 has no neighboring device on exactly one side (e.g., leftmost or rightmost devices in cells except INVX1, INVX2 which have no fingers). To investigate the edge effect in case 1, we first sweep the spacing from 180 nm to 1 μ m symmetrically on both sides. Fig. 5 shows linewidth versus spacing in case 1. As can be seen from the graph, linewidth is insensitive to focus after two SRAFs are inserted on each side of the polyline. For case 2 edge devices, we fix one side at 180 nm for dense and 380 nm for iso devices. The spacing on the other side is swept up to 2 μ m. Fig. 6 shows the case 2 edge effect of dense and iso cells, respectively. When two adjacent polylines are 1.2 μ m apart (i.e., two SRAFs are inserted at each side), the linewidth does not vary much even if the spacing becomes larger. Since the distance from edge devices to the cell boundary for all cells is over 600 nm in this technology (making the distance of two neighboring polylines more than 1.2 μ m), we assume that all edge devices of dense and iso cells



Fig. 5. Linewidth variation at 0.4- μ m defocus in case 1. Arrows indicate SB insertion points.



Fig. 6. Linewidth with spacing from 0.5 to 2 μ m at 0.0 and 0.4- μ m defocus in case 2 for dense and iso cells.

in case 2 follow the behavior seen in Fig. 6. For case 2 edge devices of self-compensated and single-pitched cells, we use the LUT linewidth value at one space from each layout and the other space at 660 nm [two scattering bars (SBs)].

D. Library Construction

The spacing between each polyline can be divided into three different ranges based on lithography-simulation results. Specific space values are used to generate each layout variant of the cells. A layout-synthesis tool is used to create the actual layouts in which all the spacings between polylines are fixed to the values of each category. From the range of self-compensated spacing, one spacing value for which $L_{\rm eff}$ variation is negligible is selected for single-pitched cells.

We consider 21 frequently used cells (INV: X1, X2, X6, X8, X12, NAND2(3), and NOR2(3): X1, X2, X4, X6). All five variants of each cell are generated using the same layout-synthesis tool, namely, original, dense, iso, self-compensated, and singlepitch. The original version is generated without any constraints in spacing, enabling the smallest area possible. The singlepitch version allows only one fixed spacing value between all polylines. This single spacing/pitch value is chosen based on its insensitivity to focus variation. The cell height is set to $4.2 \ \mu m$. Table III shows the average area overhead comparison

TABLE III Normalized Area Overhead of Each Cell Version

Cell version	Layout	Estimated [23]
Original	1.00	1.00
Dense	1.02	1.04
Iso	1.22	1.20
Self-compensated	1.10	1.10
Single-pitch	1.35	NA

found using both the actual cell layouts and the estimated areas taken from [23]. As can be seen from the table, the two approaches show similar values with self-compensated cell variants exhibiting $\sim 10\%$ area increase on average. In addition, single-pitch cells yield substantially higher area penalties than the other variants.

The LUT constructed from Fig. 4 gives CD of each variant of cells at two defocus values. Parasitic capacitances from the lumped-C model of parasitic-extraction tool are included in netlists, and timing- and power-characterization tool from Synopsys (Star-MTB) [26] is then extensively run to generate libraries of timing and power for each layout version of cells. The library (.lib) is standard Synopsys format which contains 7×7 table-based timing and dynamic power and leakages information.

IV. SELF-COMPENSATED DESIGN BASELINES

A. Self-Compensated Cell-Based Design

The most straightforward approach to creating a design that is insensitive to defocus in the lithographic process is to make each standard-cell-element self-compensating in isolation. Based on the previous section, we have created a selfcompensated library, and resulting circuit performance (area, delay, power) using this library will be used as a baseline in the results section of this paper.

B. Single-Pitch Cell Design

Designing circuits using a single pitch on the critical layer holds promise since a manufacturing process can be highly tuned to maximize manufacturability at a given pitch at the expense of printability of other pitches. We select one spacing value within the range of self-compensated spacings to generate single-pitch cells. Again, the resulting circuit performance using this library will be compared against when evaluating the optimization approaches introduced in Section V next.

V. Optimization (Self-Compensated Physical Design)

As can be seen from the previous section, a more robust design, with respect to focus variation, is possible by using either self-compensated or single-pitched cells. Another option is to generate optimized circuits using both dense and iso cells to meet timing at all focus points. Optimization with a mix of dense and iso cells is possible both in the timing (i.e., to meet critical delay) and power (i.e., to meet worst case leakage constraints) domains. In this section, we describe both heuristic and mixed ILP (MILP) solution methods to the selfcompensated physical-design problem.

A. Area-Driven Timing Optimization

The first optimization seeks to balance timing and area. We can generate new circuits that meet timing requirements through all defocus values by using both dense- and iso-cell variants; the goal will be to use as few iso variants as possible, thereby minimizing the area penalty.

1) Heuristic: Iso-dense self-compensating physical design can be viewed as a sizing problem. Since dense cells are slower (at worst case focus) and smaller while iso cells are faster and bigger, we start with the circuit initially synthesized with dense cells, then swap in iso versions to meet the timing at the worst case defocus level.

Initially, synthesis with the dense library results in the slowest timing at worst defocus conditions with minimal area. The optimization of delay versus area is implemented using a sensitivity-based approach to minimize area penalty while instantiating iso counterparts of dense cells in the circuit to meet timing constraints. In our experiments, the required time at the primary outputs is set to be the worst case delay with the original library at $0-\mu m$ defocus. The sensitivity of all gates with respect to a change from "dense" to "iso" variants can be defined as [17]

Sensitivity =
$$\frac{1}{\Delta A + K_1} \sum_{\text{arcs}} \frac{\Delta D}{\text{slack}_{\text{arc}} - S_{\min} + K_2}$$
 (1)

where ΔA is the change in area, and ΔD is the change in delay due to the swap. S_{\min} is the worst slack in the circuit when synthesized using the "dense" library, and the arcs consist of all rise and fall transitions from each input to output of the gate. The term slack_{arc} is the difference between arrival and required times of the timing arc, and K_1 and K_2 are small positive numbers to ensure numerical stability of the expression. Pseudocode for our optimization process is as follows:

Optimization {

Input: focus, Output: optimized circuits
While worst_slack is negative
Calculate sensitivities of all gates in the circuits
Sort Sensitivities in nonincreasing order
Swap the "dense" version with "iso" cell, based on order of sensitivities
Calculate new_delay of circuit
Update worst_slack

Post Processing {

If *worst_slack* at intermediate defocus value is not negative Finish optimization and exit

Else

}

}

Find the maximum-delay defocus point Perform *optimization* at the maximum-delay defocus point



Fig. 7. Illustration of optimization process (D denotes "dense" and I denotes "iso" cell; numbers are example sensitivities of gates to swapping to "iso" counterparts).

As the pseudocode indicates, we first sort sensitivities in nonincreasing order. The gate with maximum sensitivity is then swapped with its corresponding iso version. Incremental timing analysis updates the *worst_slack* value, and new sensitivities are then calculated if the timing is not met.

Fig. 7 shows an example of the swapping. The numbers in the illustration represent the sensitivity of each gate when changing from dense to iso counterparts. Since all gates are dense at first, the design may not meet timing at worst case defocus. Changing from dense to iso compensates for defocus along critical paths. The algorithm iterates until timing constraints are met. Table IV shows the (absolute) increases in area and leakage power when dense cells are swapped with their iso counterparts. Inverters show negligible impact on cell area when exchanged with iso counterparts, since there is often space in these cell layouts to make changes without impacting cell width. However, swapping more complex cells, such as NAND3 and NOR3, results in moderate area penalties. Most NAND2 or NOR gates, however, show relatively small leakage power increments compared to inverters (e.g., INVX12), since leakage in these cell types is inherently smaller. Absolute numbers are given since the algorithm directly operates on these and they will help shed light on decisions that the algorithm makes in the results section.

Even after the above optimization procedure (which ensures timing correctness at both best and worst focus conditions, assuming it is feasible), the circuit may not meet timing constraints at intermediate values of focus. This may occur since the optimization only uses information at perfect focus and worst case defocus in guiding decisions, leading to potential timing failures when focus is nonlinear or nonmonotonic. Thus, the timing constraint should be checked across defocus levels. We sweep the optimized circuits over all defocus values to find the maximum-delay focus condition. If the maximum-delay defocus point is out of the permissible focus range or the maximum delay is less than the required time, no further steps are needed. However, if the maximum-delay defocus point is within the permissible focus range, a postprocessing step, as described above, is required to globally meet the timing constraint. At the maximum-delay defocus point, we can apply the same sensitivity-based optimization process, which is shown above, to ensure that the optimized circuit meets timing throughout the expected defocus range. Delay at intermediate focus values (e.g., 0.11, 0.37 μ m, etc.) is calculated by interpolation from precharacterized cell delays at a small set of focus values (e.g., 0.1, 0.2, 0.3 μ m, etc.). In the interpolation, we assume CD to be a quadratic function of focus. In addition, we assume

TABLE IV Area and Leakage Power Change When Dense Cells Are Exchanged With Iso Counterparts

Cell	∆Area(um ²)	∆Leakage(nW)
invx1	0	0
invx2	0	0
invx6	0	37.09
invx8	0.84	64.65
invx12	0	118.46
nand2x1	0	4.95
nand2x2	1.47	10.66
nand2x4	2.35	59.54
nand2x6	3.53	116.49
nand3x1	0	17.50
nand3x2	3.07	37.12
nand3x4	7.06	100.98
nand3x6	8.23	197.70
nor2x1	0	3.31
nor2x2	2.65	33.42
nor2x4	2.35	50.44
nor2x6	6.89	93.60
nor3x1	0	17.05
nor3x2	4.70	39.50
nor3x4	6.59	88.90
nor3x6	7.77	118.13

that cell delay is a linear function of gate length for small perturbations of gate length.

2) Mixed ILP (MILP): Although the sensitivity-based heuristic optimization that uses dense and iso cells for compensating designs results in good solutions, postprocessing may be required to ensure the compensation is valid throughout the expected defocus range. Due to the nonlinearity of delay (or CD) with focus, an optimization approach should ideally guarantee that the resulting solution is valid at all defocus points.

To inherently consider the range of potential defocus conditions, we propose a new optimization approach based on ILP. For each gate *i*, let the area of component *i* as A_i , *P* be the set of all possible paths, and *n* be the number of gates in circuits. The problem of minimizing the total area subject to a maximum delay bound (required time) can be formulated as [19]

$$\begin{array}{ll} \text{Minimize} & \displaystyle\sum_{i=1}^{n} A_{i} \\ \text{Subject to} & \displaystyle\sum_{i \in p}^{i} D_{i} \leq D_{\max}; \quad \forall p \in P \\ & A_{i} \in A_{i}^{\text{dense}}, A_{i}^{\text{iso}}; \quad i = 1, \dots, n. \end{array}$$
(2)

The number of possible paths from primary inputs to primary outputs is exponential in n. Therefore, transforming the constraints on path delay into constraints on delay across components (e.g., arrival time) is widely accepted as a practical technique. a_i represents the arrival time at each node i while D_{max} is the maximum-delay bound (required time of the circuit). The above problem can be rewritten as

Minimize $\sum_{i=1}^{n} A_i$

Subject to

$$a_{j} \leq D_{\max}; \qquad j \in \text{outputs}$$

$$a_{j} + D_{i} \leq a_{i}; \quad i = 1, \dots, n \text{ and } \forall j \in \text{input}(i)$$

$$D_{i} \leq a_{i}; \qquad i = n + 1, \dots, n + s: \text{inputs}$$

$$A_{i} \in A_{i}^{\text{dense}}, A_{i}^{\text{iso}}; \qquad i = 1, \dots, n.$$
(3)

To include the delay variation due to defocus, we discretize the defocus into five levels (0.0, 0.1, 0.2, 0.3, and 0.4 μ m). The ILP problem can then be cast as

Minimize
$$\sum_{i=1}^{n} A_i$$

Subject to

$$\begin{aligned} a_{j,f} &\leq D_{\max}; \quad j \in \text{outputs} \\ a_{j,f} + D_{i,f} &\leq a_{i,f}; \quad i = 1, \dots, n \text{ and} \\ &\forall j \in \text{input}(i) \\ D_{i,f} &\leq a_{i,f}; \quad i = n+1, \dots, n+s: \text{inputs} \\ A_i &\in A_i^{\text{dense}}, A_i^{\text{iso}}; \quad i = 1, \dots, n \\ f &\in \{0.0, 0.1, 0.2, 0.3, 0.4\}; \quad \text{defocus} \quad (4) \end{aligned}$$

where $a_{i,f}$ is the arrival time of gate *i* at *f* defocus level, and $D_{i,f}$ represents the gate delay of the *i*th component at defocus level *f*.

Finally, given two choices (dense and iso) of gates, the problem can be transformed into an integer (binary) linearoptimization problem

Minimize
$$\sum_{i=1}^{n} \left[A_i^{\text{dense}}(1-x_i) + A_i^{\text{iso}}(x_i) \right]$$

Subject to

$$\begin{split} a_{j,f} &\leq D_{\max}; \quad j \in \text{outputs} \\ a_{j,f} + \left[D_{i,f}^{\text{dense}}(1-x_i) + D_{i,f}^{\text{iso}}(x_i) \right] \leq a_{i,f}; \\ i &= 1, \dots, n \text{ and } \forall j \in \text{input}(i) \\ D_{i,f}^{\text{dense}}(1-x_i) + D_{i,f}^{\text{iso}}(x_i) \leq a_{i,f}; \\ i &= n+1, \dots, n+s: \text{inputs} \\ x_i &\in 0, 1 \text{ (binary)}; \\ i &= 1, \dots, n(0 = \text{dense}, 1 = \text{iso}) \end{split}$$

 $f \in \{0.0, 0.1, 0.2, 0.3, 0.4\};$ defocus.

(5)

The integer (binary) linear problem can be efficiently solved using a commercial linear solver. In our case, we use the mixedinteger optimizer of CPLEX [20].

B. Leakage-Driven Timing Optimization

Leakage is highly sensitive to linewidth variations due to well-known short-channel effects in scaled MOSFETs. Therefore, we propose to perform optimization using dense and iso cells based on leakage characteristics rather than area. A new sensitivity metric that includes the leakage change when an iso cell replaces a dense cell can be formulated, as in (6). We ignore the area change that was considered in (1) and, instead, use Δ Leak in the denominator. As can be seen in the Bossung plot (Figs. 1 and 4), the linewidth of dense cells increases with defocus leading to less leakage. On the other hand, the linewidth of iso cells decreases with defocus, causing dramatically higher leakage in this case. The same heuristic algorithm is applied using this new sensitivity.

Sensitivity =
$$\frac{1}{\Delta \text{Leak}} \sum_{\text{arcs}} \frac{\Delta D}{\text{slack}_{\text{arc}} - S_{\min} + K_2}$$
 (6)

where Δ Leak is the leakage change in switching from dense cells to iso at the worst defocus condition.

VI. RESULTS

To quantify delay variation with defocus across the iso/ dense/self-compensated libraries and using our optimization approaches, timing libraries for three different variants of each cell are generated, as described in Section III. ISCAS85 benchmark circuits [18] are then synthesized with the "dense" library at minimum-timing constraints using Design Compiler [14].

Table V summarizes the normalized delay and area overhead using the various libraries for each benchmark circuit. The table shows that the original library incurs 13% slowdown at worst case defocus since cells in that library are inherently dense, while the delay decreases 16% on average when using the iso library alone. Both self-compensated and single-pitch cells lead to good robustness across defocus levels. The normalized leakage power information is shown on the right side of the Table V. As expected, original and dense libraries at the worst defocus value have much less (> 40%) leakage than the original library at perfect focus since linewidths systematically increase. On the other hand, leakage power with iso cells increase by more than $3 \times$ over the original cells at 0.4- μ m defocus. Leakage power overhead in both self-compensated and single-pitched cells is small (5%–6%).

The normalized area overheads incurred when using each cell variant (both uniformly and using the proposed optimization approaches) are shown in Table VII. The gate distribution and runtime of the optimization options are shown in the right side of the table. Heu1 refers to the heuristic optimization of timing and area (Section V-A) and heu2 represents the optimization of timing and leakage described in Section V-B. While self-compensated and single-pitched libraries lead to good timing behavior across focus, as already shown, they

	original		norm	alized o	lelay	normalized leakage					
benchmarks	0 um		0.4	um defo	cus	0.4 um defocus					
	defocus	original	iso	dense	self- comp.	single- pitch	original	iso	dense	self- comp.	single- pitch
c432	1.00	1.13	0.84	1.13	0.99	0.99 0.99 0.57 3.17		3.17	0.57	1.05	1.07
c499	1.00	1.12	0.83	1.13	1.00	0.99	0.58	3.17	0.57	1.06	1.07
c880	1.00	1.12	0.84	1.12	1.00	0.99	0.57	3.20	0.57	1.06	1.07
c1355	1.00	1.12	0.84	1.12	1.00	0.99	0.58	3.15	0.57	1.06	1.07
c1908	1.00	1.13	0.84	1.13	1.00	0 0.99 0.58 3.12		0.57	1.05	1.06	
c2670	1.00	1.14	0.83	1.13	0.99	0.99	0.58	3.12 0.58		1.05	1.06
c3540	1.00	1.12	0.84	1.13	0.99	0.99	0.57	3.18	0.57	1.05	1.07
c5315	1.00	1.13	0.83	1.14	0.99	0.99	0.58	3.11	0.57	1.05	1.06
c6288	1.00	1.13	0.83	1.13	1.00	0.99	0.58	3.14	0.57	1.05	1.06
c7552	1.00	1.12	0.83	1.13	1.00	0.99	0.58	3.09	0.57	1.05	1.06
Average	1.00	1.13	0.84	1.13	0.99	0.99	0.58	3.15	0.57	1.05	1.06

TABLE V Normalized Delay and Leakage Power for ISCAS85 Benchmark Circuits Synthesized in Each Library Type (Normalized to Original Cells at 0.0-µm Defocus Value)

also lead to relatively large area overheads of 11% and 27%, respectively. The ILP optimization provides an optimal solution and can be used to determine how well the heuristics are performing. The two sensitivity-based heuristics show 3%–4% area increases while meeting timing requirement throughout all defocus range. Note that the trend is toward smaller area penalties in the larger benchmarks, explainable by the fact that a smaller (relative) subset of gates are responsible for determining timing in these larger circuits. The first heuristic, in particular, achieves circuit areas very close to optimal, usually within 1%.

In Table VII, the heuristic optimization considering leakage power results in the use of fewer iso cells than the heuristic based on timing and area, since iso cells are being penalized more heavily by leakage than area due to the exponential dependence of the former. However, heu2 still shows a slightly larger area penalty since it is choosing to exchange gates that show small leakage penalties, which tend to be gates with stacked devices such as NAND2, NAND3, etc. [25]. These gates also are large and incur more severe area penalties when swapped from dense to iso variants. In contrast, heu1 selects very small gates such as inverters to convert to iso, since the change in area is being penalized in the sensitivity measure. Table VI provides details on the five most commonly swapped gates from dense to iso cells in optimizing the c5315 benchmark using the two different heuristics and the ILP. In line with the above discussion, we observe that there are substantial differences in both the total number of swapping and the type of swapped cells. Despite the fact that heu1 swaps over $10 \times$ more cells than heu2 and the ILP solutions, the area penalties are nearly identical for this circuit since most of the swapped cells in heu1 have little to no layout area penalties. As can be seen from the runtime of the various optimization approaches in Table VII, the heuristic techniques shows very reasonable efficiency with high-quality solutions relative to the ILP.

To further illustrate the differences between the heuristic and ILP optimizations, slack versus defocus is plotted for circuit c7552 in Fig. 8. The graph shows that, while the original circuit

TABLE VI TOP FIVE MOST SWAPPED GATES FOR CIRCUIT C5315 BY EACH APPROACH

heuri	stic 1	heur	istic 2	ILP				
(ar	ea)	(leal	kage)	ge) (area)				
cells	#of cells	cells	# of cells	cells	# of cells			
invx2	298	nand2x6	24	invx12	20			
invx8	198	nand2x4	8	invx6	11			
nand2x6	19	nor2x4	5	nand2x6	7			
nor2x4	7	invx2	4	nand2x4	6			
nand2x4	6	nand3x4	3	invx8	2			
totals	538		50		47			

(based on the original library) fails to meet the required time at defocus, both heuristic and ILP optimization solutions are able to meet the timing requirement throughout the defocus range. In the heuristic optimization, the timing requirement is met both at perfect focus and at the extreme defocus condition initially. However, timing failures occur at some intermediate defocus conditions due to the nonlinearity of delay and focus. The postprocessing step described in Section V-A can handle the problems and guarantee the positive slack in all defocus range. From the results of Table VII and Fig. 8, we observe that the sensitivity-based heuristics with postprocessing are very close to the ILP results. Therefore, we do not show the results of running the ILP formulation with the leakage objective instead of area.

A Monte Carlo simulation with 1000 trials is employed to investigate the impact of defocus variation on delay distribution. A normal distribution of focus with mean = 0.0 μ m and $3\sigma = 0.4 \mu$ m is assumed. Fig. 9 shows Monte Carlo simulation results for the c6288 benchmark. Self-compensated, singlepitch, and the two dense + iso optimization options meet the timing requirement at all 1000 randomly chosen defocus points.

Table VIII shows the change in leakage power at the worst defocus conditions compared to the original library at perfect focus using several self-compensating design options. As can be seen, both self-compensated cells and single-pitch cells

					norr	nalized	area			gate distribution						Runtime	
Tot benchmarks # c		oria		iaa	self-	single-	e- optimization			heuris (area	tic1 a)	heuris (leaka	tic2 ge)	illP		(sec)	
	gates	ong.	dense	ISU	comp.	pitch	heu1 (area)	heu2 (leakage)	ILP	dense	iso	dense	iso	dense	iso	heu1	ILP
c432	339	1.00	1.02	1.17	1.12	1.26	1.09	1.09	1.08	233	106	318	21	317	22	0.04	0.19
c499	682	1.00	1.00	1.17	1.11	1.27	1.00	1.02	1.00	581	101	569	113	584	98	0.09	1.70
c880	575	1.00	1.02	1.18	1.11	1.27	1.02	1.02	1.01	560	15	561	14	562	13	0.07	0.35
c1355	680	1.00	1.00	1.17	1.11	1.27	1.05	1.08	1.04	536	144	516	164	564	116	0.39	11.21
c1908	645	1.00	1.01	1.16	1.12	1.26	1.04	1.05	1.04	554	91	584	61	566	79	0.08	13.79
c2670	1040	1.00	1.01	1.15	1.11	1.25	1.05	1.05	1.04	1017	23	1020	20	1010	30	0.20	11.61
c3540	1313	1.00	1.01	1.17	1.10	1.27	1.01	1.01	1.01	1279	34	1287	26	1280	33	0.32	27.28
c5315	2028	1.00	1.00	1.16	1.11	1.27	1.01	1.01	1.00	1490	538	1978	50	1981	47	1.51	29.30
c6288	4102	1.00	1.00	1.16	1.11	1.26	1.06	1.07	1.05	3631	471	3820	282	3693	409	7.80	913.32
c7552	2700	1.00	1.00	1.15	1.11	1.25	1.01	1.01	1.00	2610	90	2658	42	2648	52	2.02	358.03
average		1.00	1.01	1.16	1.11	1.27	1.03	1.04	1.02								

 TABLE
 VII

 Normalized Area and Gate Distribution for Each Library and Optimization Approach



Fig. 8. Slack versus defocus for benchmark c7552 showing the effectiveness of various self-compensating design options. Note some defocus values (e.g., 0.1–0.18 μ m) at which the circuit fails to meet timing requirement under the heuristic optimization without postprocessing. The horizontal line at y = 0is added to highlight the timing constraint.

designs options shows modest ~7% leakage increases at worst case defocus. The area-driven dense + iso optimization shows 10% less leakage than the nominal case at 0.4- μ m defocus, although the results for this case vary widely. As expected, the leakage-driven optimization shows 25% less leakage than the original circuit and 15% less than heu1 since leakage is directly accounted for in this formulation.

VII. CONCLUSION

A novel design technique to compensate for lithographic focus-dependent CD variation is proposed in this paper. The general idea is to judiciously instantiate isolated and dense versions of library cells in a circuit to effectively negate the impact of expected focus variations. We present two heuristic approaches to self-compensated design for focus-dependent CD variation along with an ILP formulation. All three algorithms lead to circuits that can meet timing requirements across ex-



Fig. 9. Stacked histograms showing the delay distribution for c6288 (required time = 4.68 ns). Note that there is a break in the *y*-axis at 21.

pected defocus levels while incurring only very small area penalties. Specifically, we can achieve a focus compensated design with $\sim 3\%$ area overhead, compared to 11% and 27% in a self-compensated and single-pitch library-based design, respectively. In addition, we investigate the leakage impact of

LEAKAGE POWER CHANGE FOR SELF-COMPENSATING DESIGNS AND TWO HEURISTIC-BASED OPTIMIZATIONS AT 0.4-μm DEFOCUS COMPARED TO ORIGINAL LIBRARY AT 0.0-μm DEFOCUS

TABLE VIII

at 0.4 defocus	c432	c499	c880	c1355	c1908	c2670	c3540	c5315	c6288	c7552	Avg.
self-compensated	5.1%	5.7%	5.5%	5.6%	5.5%	5.0%	5.5%	5.5%	5.3%	5.4%	5.4%
single-pitched	6.6%	6.9%	7.0%	6.7%	6.2%	5.9%	6.8%	6.2%	6.4%	5.9%	6.5%
heu1 (area)	31.4%	-4.8%	-36.3%	11.3%	-6.5%	-36.5%	-34.5%	17.4%	-10.5%	-33.6%	-10.3%
heu2 (leakage)	-25.7%	0.6%	-36.9%	14.8%	-22.2%	-37.6%	-37.2%	-36.6%	-25.9%	-39.0%	-24.6%

defocus, and one of the heuristics seeks to minimize leakage while meeting timing requirements. Results using both iso and dense libraries together show 30% lower leakage compared to circuits designed using an inherently self-compensated library under worst case focus conditions.

REFERENCES

- Y. Cao et al., "Design sensitivities to variability: Extrapolations and assessments in nanometer VLSI," in Proc. ASIC/SOC, 2002, pp. 411–415.
- [2] S. R. Nassif, "Design for variability in DSM technologies," in *Proc. ISQED*, 2000, pp. 451–454.
- [3] S. R. Nassif, "Within-chip variability analysis," in *IEDM Tech. Dig.*, 1998, pp. 283–286.
- [4] 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," in *Proc. Int. Conf. Comput.-Aided Des.*, 2000, pp. 62–67.
- [5] P. Gupta and A. B. Kahng, "Manufacturing-aware physical design," in Proc. Int. Conf. Comput.-Aided Des., 1998, pp. 681–687.
- [6] P. Gupta and H. Fook-Luen, "Toward a systematic-variation aware timing methodology," in *Proc. Des. Autom. Conf.*, 2004, pp. 321–326.
- [7] L. Capodieci, P. Gupta, A. B. Kahng, D. Sylvester, and J. Yang, "Toward a methodology for manufacturability driven design rule exploration," in *Proc. Des. Autom. Conf.*, 2004, pp. 311–316.
- [8] L. Liebmann, G. Northrop, J. Culp, L. Sigal, A. Barish, and C. Fonseca, "Layout optimization at the pinnacle of optical lithography," *Proc. SPIE*, vol. 5042, pp. 1–14, 2003.
- [9] L. Liebmann, D. Maynard, K. McCullen, N. Seong, E. Buturla, M. Lavin, and J. Hibbeler, "Integrating DfM components into a cohesive design-tosilicon solution," *Proc. SPIE*, vol. 5756, pp. 1–12, 2005.
- [10] Calibre version 2005.3_6.10. [Online]. Available: http://www.mentor.com
- [11] International Technology Roadmap for Semiconductors 2003. [Online]. Available: http://public.itrs.net/Files/2003ITRS/Home2003.htm
- [12] T. Yorick *et al.*, "ArF imaging with off-axis illumination and subresolution assist bars: A compromise between mask constraints and lithographic process constraints," *Proc. SPIE*, vol. 4691, pp. 1522–1529, 2002.
- [13] A. J. Lori, T. R. Michael, D. Jason, and J. Christiane, "Effect of scattering bar assist features in 193-nm lithography," *Proc. SPIE*, vol. 4691, pp. 861– 870, 2002.
- [14] Design Compiler version V-2003.12. [Online]. Available: http://www. synopsys.com
- [15] A. B. Kahng and Y. C. Pati, "Subwavelength lithography and its potential impact on design and EDA," in *Proc. Des. Autom. Conf.*, 1999, pp. 799–804.
- [16] L. W. Liebmann, S. M. Mansfield, A. K. Wong, M. A. Lavin, W. C. Leipold, and T. G. Dunham, "TCAD development for lithography resolution enhancement," *IBM J. Res. Develop.*, vol. 45, no. 5, pp. 651– 665, 2001.
- [17] S. Sirichotiyakul *et al.*, "Stand-by power minimization through simultaneous threshold voltage selection and circuit sizing," in *Proc. Des. Autom. Conf.*, 1999, pp. 436–441.
- [18] F. Brglez and H. Fujiwara, "A neutral netlist of 10 combinational benchmark circuits and a target translator in Fortran," in *Proc. ISCAS*, May 1989, pp. 695–698.
- [19] C.-P. Chen, C. C. N. Chu, and D. F. Wong, "Fast and exact simultaneous gate and wire sizing by Lagrangian relaxation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 18, no. 7, pp. 1014–1025, Jul. 1999.

- [20] CPLEX 9.0 Reference Manual. [Online]. Available: http://www.ilog.com
- [21] Prolific, Inc. [Online]. Available: http://www.prolificinc.com/
- [22] Cadence Design Systems. [Online]. Available: http://www.cadence.com/
- [23] P. Gupta, Y. Kim, A. B. Kahng, and D. Sylvester, "Self-compensating design for focus variation," in *Proc. Des. Autom. Conf.*, 2005, pp. 365–368.
- [24] M. Orshansky, L. Milor, and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial masklevel correction," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 1, pp. 2–11, Feb. 2004.
- [25] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *Proc. IEEE Symp. VLSI Circuits*, 1998, pp. 40–41.
- [26] Synopsys Corp. [Online]. Available: http://www.synopsys.com/



Puneet Gupta (S'01) received the Ph.D. degree in computer engineering from University of California (UC) San Diego, La Jolla.

Since 2004, he has been at Blaze DFM Inc. as Cofounder and Product Architect. He has authored over 40 papers and 12 pending patents related to physical design and manufacturing.

Mr. Gupta was the recipient of the IBM Ph.D. fellowship and has given tutorial talks at International Conference on Computer-Aided Design, WesCon, Chemical-Mechanical Polish (CMP)-Multilevel In-

terconnection (MIC), as well as UC Santa Cruz and UC San Diego.



Andrew B. Kahng (A'89–M'03–SM'05) received the A.B. degree in applied mathematics from Harvard College, Cambridge, MA, and the M.S. and Ph.D. degrees in computer science from the University of California (UC), San Diego.

He is currently a Professor in computer science engineering and electrical and computer engineering at UC San Diego. Since 1997, his research in very large-scale integration (VLSI) design for manufacturing has pioneered methods for automated phaseshift mask layout, CMP fill synthesis, and parametric

yield-driven, cost-driven, and variation-aware optimizations. From October 2004 to September 2006, he served as Chairman and Chief Technology Officer (CTO) of Blaze DFM, a company that provides new cost and yield optimizations in the VLSI design-to-manufacturing interface.

Prof. Kahng was the recipient of the 1992 NSF Young Investigator Award and has approximately 300 publications in the VLSI CAD area, including five Best Paper Awards and six other Best Paper Nominations. He was the founding General Chair of the International Symposium on Physical Design and Technical Program Cochair of the 2004 and 2005 Design Automation Conferences. From 2000 to 2003, he chaired both the U.S. and international working groups for Design Technology for the International Technology Roadmap for Semiconductors. He has also served as executive committee member and Theme Leader for the MARCO Gigascale Systems Research Center from 1998 to 2006.



Youngmin Kim (S'02) received the B.S. degree in electrical engineering from the Yonsei University, Seoul, Korea, in 1999, and the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, in 2003, where he is currently working toward the Ph.D. degree.

His research interests are variability-aware design methodologies, design for manufacturability, and interconnect modeling and analysis.



Dennis Sylvester (S'95–M'00–SM'04) received the B.S. degree (*summa cum laude*) in electrical engineering from the University of Michigan, Ann Arbor, in 1995, and the M.S. and Ph.D. degrees in electrical engineering from University of California (UC), Berkeley, in 1997 and 1999, respectively. His dissertation research was recognized with the 2000 David J. Sakrison Memorial Prize as the most outstanding research in the UC-Berkeley EECS Department.

He has held research staff positions in the Advanced Technology Group of Synopsys, Mountain

View, CA, and at Hewlett-Packard Laboratories, Palo Alto, CA. He is currently an Associate Professor in electrical engineering and computer science at the University of Michigan. He was also a Visiting Associate Professor in electrical and computer engineering at the National University of Singapore, Singapore, during the 2006 to 2007 academic year. He has published numerous articles along with one book and several book chapters in his field of research, which includes low-power circuit design and design-automation techniques, designfor-manufacturability, and onchip interconnect modeling. He also serves as a Consultant and a technical advisory board member for several electronic design-automation firms in these areas.

Dr. Sylvester was the recipient of the NSF CAREER award, the 2000 Beatrice Winner Award at International Solid State Circuits Conference (ISSCC), an IBM Faculty Award, an SRC Inventor Recognition Award, and several best paper awards and nominations. He was the recipient of the Association for Computing Machinery (ACM) Special Interest Group on Design Automation Outstanding New Faculty Award, the 1938 E Award for teaching and mentoring, the Vulcans Education Excellence Award from the College of Engineering, and the University of Michigan Henry Russel Award. He has served on the technical program committee of numerous design automation and circuit design conferences and was general chair of the 2003 ACM/IEEE System-Level Interconnect Prediction Workshop and 2005 ACM/IEEE Workshop on Timing Issues in the Synthesis and Specification of Digital Systems. He is currently an Associate Editor for IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN and served as Associate Editor for IEEE TRANSACTIONS ON VLSI SYSTEMS from 2003 to 2006. He helped define the circuit and physical design roadmap as a member of the International Technology Roadmap for Semiconductors U.S. Design Technology Working Group from 2001 to 2003. He is a member of ACM, the American Society of Engineering Education, and Eta Kappa Nu.