

Impact of Systematic Spatial Intra-Chip Gate Length Variability on Performance of High-Speed Digital Circuits

Michael Orshansky, Linda Milor*, Pinhong Chen, Kurt Keutzer, and Chenming Hu

University of California, Berkeley, CA, 94720

*eSilicon Corporation, Berkeley, CA, 94710

Abstract

Using data collected from an actual state-of-the-art fabrication facility, we conducted a comprehensive characterization of an advanced 0.18 μm CMOS process. The measured data revealed significant systematic, rather than random, spatial intra-chip variability of MOS gate length, leading to large circuit path delay variation. The critical path value of a combinational logic block varies by as much as 17%, and the global skew is increased by 8%. Thus, a significant timing error (~25%) and performance loss takes place if variability is not properly addressed. We derive a model, which allows estimating performance degradation for the given circuit and process parameters. Analysis shows that the *spatial*, rather than proximity-dependent, systematic Lgate variability is the main cause of large circuit speed degradation. The degradation is worse for the circuits with a larger number of critical paths and shorter average logic depth. We propose a location-dependent timing analysis methodology that allows to mitigate the detrimental effects of Lgate variability, and developed a tool linking the layout-dependent spatial information to circuit analysis. We discuss the details of the practical implementation of the methodology, and provide the guidelines for managing the design complexity.

1 Introduction

The increasing complexity of semiconductor processes makes the interaction between manufacturing and design more severe. New models and methods are needed to quantify this interaction [1,2,3]. In CMOS digital technologies, the single most important processing parameter affecting the circuit performance is the gate length (Lgate) of the MOS transistor. Control and accurate modeling of Lgate is thus of utmost importance for accurate modeling and design of ICs.

Deep sub-micron technologies, however, exhibit a new variability pattern, which is not addressed by the previously developed models and methods: the *systematic spatial intra-chip Lgate variability*. As a result, the printed transistors display a distinct spatial Lgate map, making their characteristics dependent on the location within the chip. This variation is mainly caused by the stepper-induced illumination and imaging non-uniformity due to lens aberrations, which are worst near the optical resolution limit [4,5]. Because the continuing scaling of semiconductor processes, following Moore's Law, forces us to operate closer to the optical resolution limit of stepper systems, the intra-chip Lgate variability will only increase.

In this work we collected data from a state-of-the-art fabrication facility to study the complex interactions of design and manufacturing. We found a significant spatial variation of the circuit timing properties that lead to degradation of the overall circuit speed, if not properly addressed at the design stage. We provide a novel analytical framework that allows estimation of performance degradation for the given circuit and process parameters. The systematic nature of intra-chip variability makes previously used approaches to statistical circuit analysis, such as worst-case analysis, insufficient and inaccurate. Instead, we propose a new approach that makes the device characteristics dependent on their location within the chip. By accurately predicting the spatial dependence of circuit characteristics,

the detrimental effect of intra-chip variability can be substantially reduced.

The paper is organized as follows. In section 2, we describe the experimental procedure necessary for accurate characterization of systematic Lgate variability. In section 3 we present a set of analytical models for evaluating the impact of Lgate variation on circuit performance. Section 4 discusses the methodology for location-dependent timing analysis. Section 5 discusses the relation between location-dependent timing and traditional worst-case analysis. Section 6 summarizes the methodology, and outlines its future extensions.

2 Experimental Investigation of Systematic Intra-Chip Lgate Variability

We performed a comprehensive, silicon-based experimental characterization of an advanced production 0.18 μm logic process technology with the goal of capturing all the relevant variability patterns. One of the most important aspects of the characterization was to address the possible interaction between the global lens aberration, and the local layout pattern-dependent non-uniformities due to the optical proximity effect. Toward this end, we classified all the gates into 18 categories depending on their orientation in the layout (vertical or horizontal) and the spacing to the nearest neighboring gate (Figure 1). To capture a particular lens aberration, the coma effect, we also

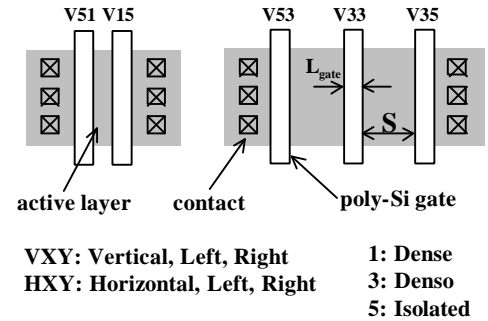


Figure 1. Spatial profiles depend on proximity effects. We categorize all gates according to their local layout patterns.

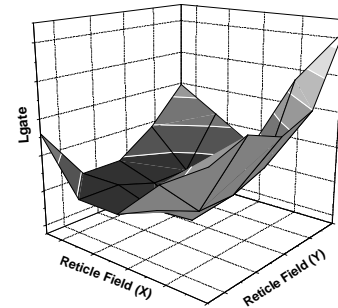


Figure 2. Lgate map for category V53, most frequent category in the design.

distinguished the relative position of the surrounding gates, i.e. the neighbor being on the left vs. the neighbor being on the right. In order to characterize the spatial Lgate profile, a test-chip contained a 5x5 grid of test-modules. Finally, Lgate values were measured on 18 test-chips coming from 3 distinct wafers.

The spatial Lgate maps were measured separately for each gate category (Figure 2). The range of variation of the Lgate surface ($L_{max} - L_{min}$) / L_{ave} is 8-12% depending on the category, with the mean of 10.2%. The Lgate maps for different categories have quite distinct spatial behaviors, due to the interaction between the global lens aberration and the pattern-dependent optical proximity effect. Thus, at least for some gate categories, the distinct spatial maps have to be used in the course of timing analysis.

Statistical *F*-tests verified that the generated topological maps of Lgate variation over the chip are statistically significant, i.e. that *the level of systematic variation is large in comparison with the random noise*. Any accurate modeling approach must consider the variation of the mean Lgate, not limiting itself to the assumption of purely random Lgate variation.

3 Simulation of the Impact of Systematic Intra-Chip Lgate Variability on Circuit Performance

Presence of systematic spatial Lgate variation significantly impacts the timing and, even functional, properties of integrated circuits. In this section we describe a tool, capable of incorporating the spatial information into the verification flow. We use it to study the impact of systematic intra-chip Lgate variation on design, and discuss the implementation details.

Incorporation of systematic Lgate information into timing verification requires making the device properties depend on the device's location within the chip. For this we developed *SpaceTimer*, a tool with the following functionality (Figure 3). A netlist is first extracted from the original circuit layout. The layout and the netlist are then passed to the *SpaceTimer* that classifies each gate as belonging to a particular category, and determines the spatial location of each gate within the layout (chip). Using this information together with the set of Lgate maps produced at the stage of characterization, *SpaceTimer* generates a modified netlist in which each gate has a proper location-dependent Lgate value, and simulates it using a circuit simulator. The simulator can be either dynamic (SPICE), or a static timing simulator.

The most direct impact of spatial Lgate variation is the resulting variation of a CMOS gate delay. In fact, because of the non-linear delay vs. Lgate relationship ($delay \sim L_{gate}^{1.5}$), variation of circuit speed is larger than Lgate variation. We evaluate speed variation by analyzing a 151-stage NAND ring oscillator (RO), often used as a predictor of the chip performance. To achieve highest accuracy, the SPICE simulator is used in this case to generate a spatial RO frequency map [6]. Results showed that the ring oscillator frequency

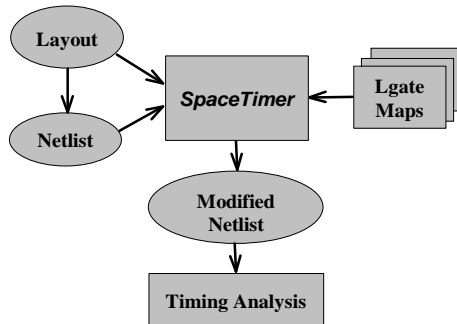


Figure 3. *SpaceTimer* employs original layout, netlist, and Lgate spatial maps to produce a location-dependent timing report.

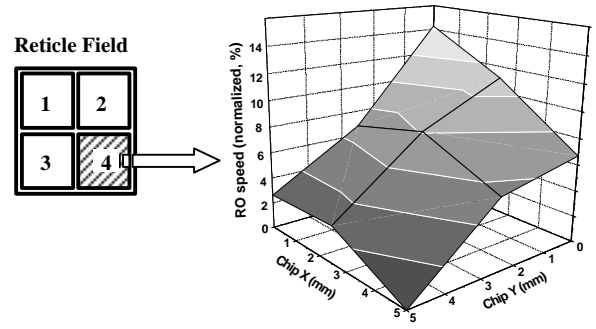


Figure 4. As a result of spatial Lgate variability, the circuit timing properties significantly vary across the chip. (Chip 4 is shown).

map is consistent with the spatial patterns of Lgate variation (Figure 2): the frequency is highest in the center of the chip, where Lgate is minimal. A comparison with the measurements was made for 4 ring oscillators available for test within each reticle field, and a good agreement was observed confirming the accuracy of the simulation result. The range of variation in RO speed across the chip is 14.5% (Figure 4).

Such a large variation in device performance also strongly affects the timing behavior of critical paths in the design. We simulated timing behavior of a benchmark combinational circuit from ISCAS'85 [7], containing 1764 CMOS devices, using a static timing simulator PathMill from Synopsys [8]. Analysis was done for 9 spatial locations on the reticle field in a uniform 3x3 grid. For the chip 4 located in the lower-right quadrant, the delay of the same path placed at different corners (Fast and Slow) of the chip varied from $t_{crit}^{fast} = 1188ps$ to $t_{crit}^{slow} = 1376ps$ – a 16% difference. The variance of path delay distribution is also different: $s_D^{slow} / s_D^{fast} = 1.6$ (Figure 5). Thus, circuit paths with identical designed-for delays will, in reality, have considerably different delay distributions, depending on the physical location of the path within the chip. As a result the overall critical path delay distribution is broadened around the designed-for delay, with some slower and some faster paths. (We discuss the consequences of this effect in section 4.)

Importantly, the order of critical paths also changes depending on location of the combinational block within the chip. Let us consider the extreme case, and compare the top 20 critical paths associated with the spatial points giving the fastest {F} and the slowest {S} path

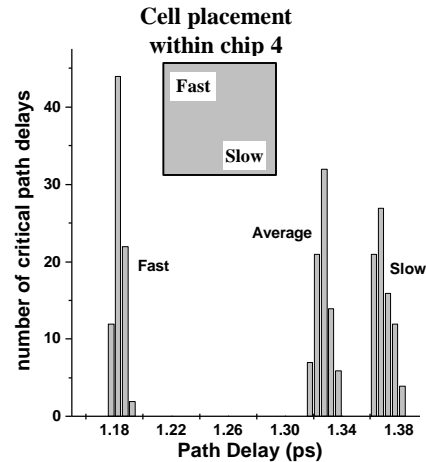


Figure 5. Distribution of critical paths of a combinational logic cell changes significantly if placed at different locations within chip 4.

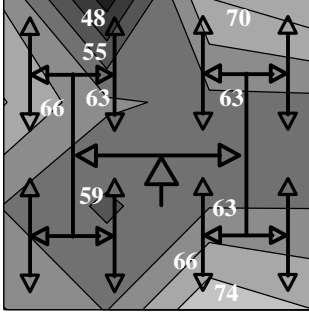


Figure 6. Global skew map for H-tree clock (ps) for chip 4. Chip size is 5x5mm². Maximum skew is 74ps. This is 8% of clock cycle.

delays: the locations with the smallest and largest Lgate. The comparison shows that only 6 of the paths found in {F} can also be found in {S}. In particular, the paths S_1, S_2, S_3 become the paths F_{15}, F_7, F_{13} . Such re-grouping significantly complicates the use of pre-designed circuit blocks physically localized within the chip, such as hard intellectual property (IP) blocks since their pre-characterized timing behavior is likely to correspond to a different location within the chip.

The systematic across-chip Lgate variation affects also the global circuit properties, such as clock skew in clock distribution networks containing buffers for driving and restoring the signal. Control of clock skew is critical, since in determining a conservative clock cycle time, a percentage delay due to clock skew is additive to the set-up times and hold times of the circuitry. We considered clock skew of the global clock network, distributed using the popular H-tree scheme. The basic intent of such a clock network is to equalize the arrival time of all the clock signals to the output loads; thus, skew then is defined as the *maximum* difference between any of the clock arrival times. Let D_i be the delay of the clock from the central buffer to the output node i , and the skew be defined as $S_i = \max\{D_{\max} - D_i, D_i - D_{\min}\}$, where D_{\min} and D_{\max} are the minimum and maximum delay values for the 16 output nodes. For one of the chips (the upper-left quadrant of the field), the maximum systematic skew is 75ps (Figure 6). This is 8% of the total clock cycle! And since the clock skew is additive to the set up time in determining the clock cycle, the total (combined) timing error of location-independent timing may be as high as 25%.

4 Analytical Model of the Circuit Speed Degradation Due to Lgate Variation

We now develop a theoretical framework that allows explicit study of the impact of intra-chip gate Lgate variation on circuit speed degradation of *complex VLSI chips in a mass production*. We show that the intra-chip Lgate variation has a significant detrimental impact on the overall circuit performance, shifting the whole chip speed distribution towards slower speeds. In contrast, the inter-chip Lgate variation, traditionally considered in statistical circuit analysis, leads to the variation of chip speed around the average speed value.

4.1 Path delay variation due to Intra-Chip Lgate Variation

In order to assess the impact of intra-chip Lgate variability on circuit speed, we decompose the overall intra-chip Lgate variation into three separate components: spatial, proximity-dependent, and the random residual. The decomposition is necessary because the spatial and proximity-dependent components affect speed variation differently:

$$L^{intra} = L^{prox} + L^{spat} + \mathbf{e} \quad (1)$$

All the terms are difference-terms, with the mean equal to zero. L^{prox} is a discrete random variable, whose distribution is circuit-specific, and can be generated by the empirical analysis of the layout, and which depends on the frequency of each gate category in the layout. Even though the spatial variation component is predominantly systematic, i.e. better described by the contour map than by variance, for simplicity of analysis, we may approximate it by a normal distribution $L^{spat} \sim N(0, \mathbf{S}_{spat}^2)$. The random residual component is distributed normally, $\mathbf{e} \sim N(0, \mathbf{S}^2)$.

In order to model the impact of intra-chip Lgate variation on circuit speed we consider variability of *path* delays. Indeed, the frequency (f), at which a circuit can be operated, is determined by the slowest path delay. We use a compact gate delay model to link the Lgate variation with the variation of gate delays [4]:

$$d = \frac{C_L V_{dd}}{n} (I_{dn}^{-1} + I_{dp}^{-1}) \quad (2)$$

where I_{dn} and I_{dp} are the drain currents of NMOS and PMOS, V_{dd} is supply voltage, $n=3.7$, and C_L is capacitance load. For deep sub-micron MOS devices, the saturation current may be described by the universal empirical equation [4]:

$$I_{dsat} \sim L_{eff}^{-0.5} T_{ox}^{-0.8} (V_{dd} - V_t) \quad (3)$$

We can simplify analysis by assuming that the parasitic junction capacitance is small and $L_{eff} \approx L_{gate}$, so that $C_L \equiv L_{gate} \cdot W \cdot C_{ox}$.

Combining the above equations, we get $d = k \cdot L_{gate}^{1.5}$, where k is a lumped process-specific constant. Because we are interested in the analysis of how the delay is affected by the pattern-dependent Lgate variability, we need to consider non-inverter logic gates comprised of devices of different categories. The delay of such complex gates, however, can be accurately described by the inverter-based delay equation (Eq. 2) if the value of L_{eff} properly takes into account the parallel and series connection of transistors within the gate.

Delay of a gate path is $D = \sum_1^m d_i = k \sum_1^m L_i^{0.5} L_{i+1}$, where d_i is a single-stage delay, and m is the number of gates in the path. The path delay variance is: $Var\{D\} = Var\{\sum_1^m d_i\}$. We can find variance of the path delay using the delta method. The expression for delay of a single gate is expanded around L_o , distinguishing delay responses to Lgate variation of the driver and of the capacitive load:

$$d = d_o + \frac{\partial d}{\partial L_i} L_i + \frac{\partial d}{\partial L_{i+1}} L_{i+1} = d_o + \frac{0.5d_o}{L_o} L_i + \frac{d_o}{L_o} L_{i+1} \quad (4)$$

We assume that the critical paths are spatially confined to a relatively small region of the chip significantly smaller than the range of variation of the spatial Lgate component. The random residual component \mathbf{e} is spatially un-correlated, i.e. it affects each poly-silicon line comprising the complex logic gate in an un-correlated manner. Since within the complex gates its effect is averaged, we can simplify the modeling if we assume that each complex gate is affected by the averaged residual term, dependent on the fan-in of the gate. Our analysis shows that across multiple layouts the average fan-in is close to 2. Then, the effective residual is $\mathbf{e}' = (\mathbf{e}_1 + \mathbf{e}_2)/2$ and $\mathbf{e}' \sim N(0, \mathbf{S}^2/2)$. Treating separately the pull-up and pull-down networks of a load stage, we can write from (3):

$$d = d_o + \frac{0.5d_o}{L_o} \left[(L_i^{prox} + L^{spat} + \mathbf{e}_i') + 2(L_{i+1}^{prox} + L^{spat} + \mathbf{e}_{i+1}') + (\mathbf{e}_{i+1}' + \mathbf{e}_{i+1}')/2 \right] \quad (5)$$

The variance of the path delay is given by the sum of the three terms. For large m and using $D_o = md_o$, the path delay variance is:

$$s_D = \sqrt{\text{Var}\{D\}} = \frac{1.5D_o}{L_o} \left(\frac{s_{Lcat}^2 + 0.28s^2}{m} + s_{Lspat}^2 \right)^{1/2} \quad (6)$$

4.2 Clock-cycle degradation due to path delay variation

High-speed digital products, such as microprocessors, are usually manufactured in such a way that the clock speed is the highest possible for each given chip. Let Π be the set of all circuit paths and let D_{ij} denote the delay of the path $j \in \Pi$ in the manufactured chip i . Then, the clock period for chip i is:

$$T_{clock}^i = \max\{D_{ij} \mid j \in \Pi\} \quad (7)$$

This equation is the key to understanding the difference in the impact of intra- and inter-chip variability. Complex high-performance silicon chips are designed in such a way that there is a large number of paths with delay close to the maximum designed-for delay D_o . *Inter-chip* Lgate variation affects each path, *shifting* the entire path delay distribution. In contrast to that, *intra-chip* spatial and proximity-dependent Lgate variability leads to the path delay variation *around* D_o , by slowing down some paths and speeding-up the others (Figure 7). As a result, on average, the maximum path delay is greater than D_o , and the average clock period is increased (Figure 8).

We now derive a set of analytical models that allow predicting the clock period degradation due to intra-chip Lgate variability. Let N be the number of paths with delay close to D_o and let D_o^i be the maximum delay for chip i , if there were no intra-chip variation, i.e. if $s_D = 0$. The path delays D_{ij} are random variables, and for tractability of analysis, we approximate their distribution as multivariate normal with the diagonal covariance matrix (Monte-Carlo analysis, described below, confirms the validity of this approximation):

$$D_{ij} \sim N(D_o^i, s_D^2) \quad (8)$$

Instead of finding T_{clock}^i analytically, we estimate its expected value as the maximum path delay value, which we will 'see' on average if N normally distributed random variables (path delays) are considered. The number of trials required on average for an event of probability p_o to happen is $N = 1/p_o$, then:

Theorem 1. For a deviation factor h , let $p_o = \text{Prob}\{(T_{clock}^i - D_o^i)/s_D \geq h\}$. The expected clock period T_{clock}^i for chip i is:

$$E\{T_{clock}^i\} = D_o^i + hs_D \quad (9)$$

In other words, intra-chip path delay variation causes the clock period T_{clock}^i for chip i to deviate (on average) by hs_D from the chip's

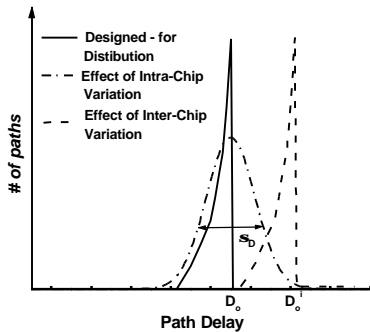


Figure 7. Intra-field Lgate variation leads to broadening of the distribution of path delays around the designed-for critical path delay value D_o .

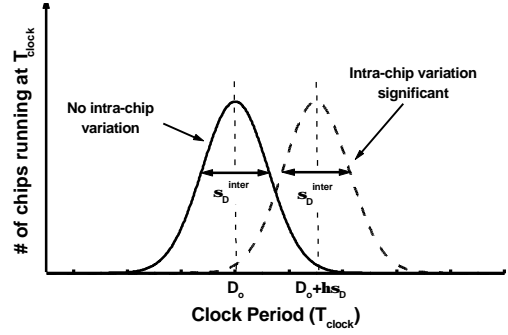


Figure 8. In contrast to between-field Lgate variability intra-field variation component degrades average delay, shifting the whole distribution.

designed-for maximum. We can find the expected value, $E\{T_{clock}\}$, of the clock period *across all the chips* determined by the inter-chip Lgate variability.

Theorem 2. For n chips, let $D_o = \frac{1}{n} \sum D_o^i$:

$$E\{T_{clock}\} = E\{D_o^i + hs_D\} = D_o + hs_D \quad (10)$$

In other words, across all the chips, intra-chip path delay variation causes the clock period T_{clock} to deviate (on average) by hs_D from the designed-for critical path delay D_o . For example, for $N=1000$, $h=3.2$ and $E\{T_{clock}\} = D_o + 3.2s_D$. Table III gives the values of h for several different N , and other values can be found from the table of normal distribution.

Table III. Values of the deviation factor for various N

Number of Paths	100	500	1000	5000
Deviation factor, h	2.3	2.9	3.2	3.5

Theorems 1 and 2 clearly show that the inter-chip variation component leads to variation of a chip's critical path delay around the designed-for critical path delay D_o , while the presence of the intra-chip Lgate variation *degrades the average circuit delay!* We can also compare the impact of both Lgate variation components on the clock period.

Theorem 3. The overall deviation of the actual critical path delay T_{clock}^i from the designed-for value D_o is:

$$E\{(T_{clock}^i - D_o)^2\} = \text{Var}\{T_{clock}^i\} + h^2 s_D^2 \quad (11)$$

The first term is the variance of T_{clock}^i due to inter-chip Lgate variation, and by analogy with Eq. 6 can be shown to be $s_D^{inter} = \frac{1.5D_o}{L_o} s_{Linter}$. The second term in Eq. 11 is the shift of the average T_{clock}^i , and is given by Theorem 2. Then,

$$E\{(T_{clock}^i - D_o)^2\} = \left(\frac{1.5D_o}{L_o}\right)^2 \left[s_{Linter}^2 + h^2 \left(\frac{s_{Lcat}^2 + 0.28s^2}{m} + s_{Lspat}^2 \right) \right] \quad (12)$$

This expression allows estimating the relative magnitude of the degradation of the average circuit speed compared to the variation around the average value. For example, if $N=1000$, so that $h=3.2$, the squared deviation of the average speed from the designed-for speed D_o is 1.7 greater than the random variation around D_o . Clearly,

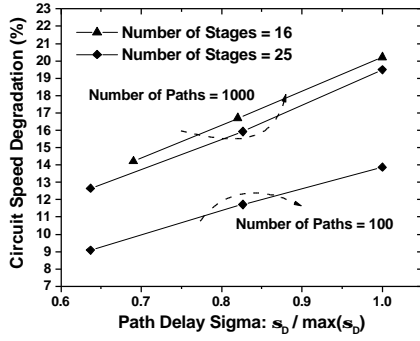


Figure 9. Circuit speed is reduced by up to 21% compared to designed-for speed. Degradation is worse for more complex circuits (larger N) and smaller average path length (smaller m).

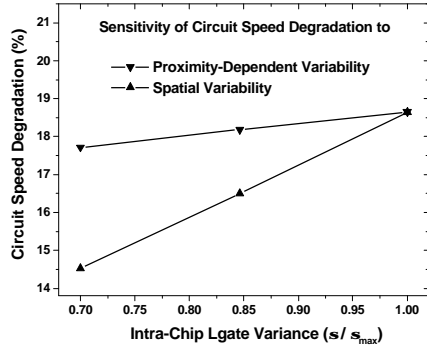


Figure 10. *Spatial* Lgate variability degrades circuit speed much more significantly than *proximity-dependent* variability.

the effect of intra-chip Lgate variability on circuit performance is very significant.

A Monte Carlo simulation was performed for model verification by generating a number of random Lgate vectors following the specified distribution of L^{prox} and L^{spat} . We calculated the path delay for each of the vectors, and compared the resulting delay variances with the model predictions. The results show that very good accuracy is achieved between the model-based predictions of path delay variance (Eq. 6) and the variance given by Monte Carlo simulation. The average error of prediction is just 1.7%. The model of T_{clock}^i (Eq. 9) also appears to be very accurate: the average error of prediction is only 1.2%.

We evaluated the impact of Lgate variation on circuit performance using the measured characteristics of the production 0.18 μ m CMOS process (section 2) for different values of the model variables (Figure 9). Both the Monte Carlo simulation and the model predict an up to 20% degradation of the average circuit speed as a result of intra-chip Lgate variation. Speed degradation is worse for more complex chips, since they contain more critical paths (larger N) and for shorter paths (smaller m). Also, important is that spatial intra-chip variation has a much stronger effect on degradation of circuit speed than proximity-dependent Lgate variation. This is because the averaging of Lgate of the gate stages within the path reduces the delay variation (Figure 10).

5 Practical Implementation of Location-Dependent Timing Analysis

The analysis of the previous section showed that systematic intra-chip Lgate variability has a large detrimental effect on the overall (average) circuit speed. These negative effects may be reduced by a location-dependent circuit analysis approach that takes the systematic Lgate variation into account. While it may bring much benefit, the

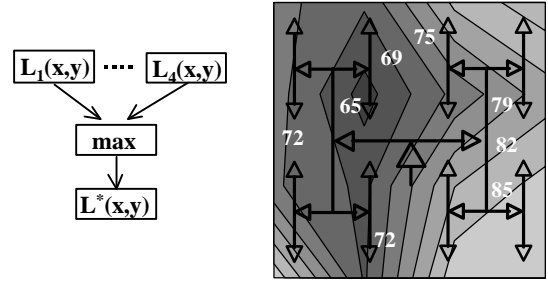


Figure 11. Clock-skew map (ps) generated by ‘collapsing’ the 4 chips of the reticle field. Maximum skew is now 87ps.

practical implementation of location-dependent circuit analysis faces several difficulties. One important complication is that the proper unit of *systematic and repeatable* Lgate spatial profile is the reticle field of a photolithographic stepper machine. Because the number of chips per reticle field is usually more than one, we, in reality, are faced with several different layouts and designs. For example, the National Technology Roadmap for Semiconductors 1997 projects that for microprocessor products, the number of chips per reticle field will be 2-4 [11]. For n chips per reticle field, we may have to keep track of n distinct designs and optimize them individually. The resulting design would achieve the highest performance possible, but this is the most expensive approach and can be justified only for high-end designs.

An alternative approach is to give up optimality in exchange for simplicity of working with a single design. This may be achieved by using a location-dependent timing analysis based on the combined Lgate map. If n is the number of chips per reticle and $L_{(x,y)}^i$ is the Lgate map for chip i , the combined Lgate is:

$$L_{(x,y)}^* = \max\{L_{(x,y)}^i \mid i = 1..n\} \quad (13)$$

This guarantees that the timing analysis based on $L_{(x,y)}^*$ is properly conservative, e.g. that the predicted clock period $T^* = \max\{T_i\}$, where T_i is the critical path delay on chip i . Note that such ‘collapsing’ is more accurate than ‘standard’ timing analysis, which does not consider location-dependent timing properties. Indeed, unless there is no Lgate variation, there must exist a point (x_o, y_o) for which $L(x_o, y_o) > L_{ave}$, and $T(L(x_o, y_o)) > T(L_{ave})$. The standard analysis has to assume that $T_{clock} = T(L_{ave})$, and is in error since $T(L_{ave})$ clearly does not represent the maximum path delay, and thus can not be an accurate clock cycle estimate.

Combining the multiple Lgate maps into one through a transformation of Eq. 13 will in certain situations result in sub-optimal circuit performance, e.g. when the clock cycle time has to be set by a rigidly designed clock generator, based on a fixed estimation of circuit critical path. In this case any clock-cycle time in excess of the actual (chip-specific) critical path is a direct performance loss. Specifically, we can define the performance loss for chip i as $PL_i = (T^* - T_i)/T_i$ and $PL_{ave} = \sum PL_i / n$. We evaluated these quantities for the benchmark combinational circuit c499. For 4 chips/reticle, the critical path delays $\{T\}$ are (in ps) $\{1190, 1330, 1280, 1380\}$. Collapsing $\{T\} \rightarrow T^*$ gives $T^* = 1380$ ps and $PL_{ave} \sim 6.5\%$. The maximum performance loss, however, is $PL_{max} = 16\%$! In the clock-skew example, combining the four skew maps into the overall skew map (Figure 11) raises the maximum skew to 88ps, and for the 1st chip (skew of 75ps), the performance loss is 17%. Clearly, the performance loss due to ‘collapsing’ may be quite significant. In each individual case, depending on the performance loss and potential gains, a suitable

approach may be chosen to balance the trade-off between higher performance (multiple designs) and simplicity ('collapsing').

6 Location-Dependent and Worst-Case Timing Analysis

It is important to clarify the relation between the location-dependent delay variability analysis, which we considered so far, and the traditional worst-case timing analysis, since both deal with deviations from the 'nominal' case. In particular, we show that the systematic Lgate variation cannot be accurately modeled by traditional statistical methods, and itself is not a substitute for the statistical methods. The goal of statistical circuit timing analysis is to determine the probability density function (pdf) of the circuit delay, or, equivalently $(\mathbf{m}, \mathbf{S}^2)$. In most cases, however, a tacit assumption is made that the mean is known a priori, and one is concerned only with finding the delay variance. This is the approach taken by such a widely used statistical method as worst-case analysis [9,10]. In contrast, the location-dependent timing analysis is concerned with *variation of the mean* timing properties of the circuit as a function of position.

Despite the formal differences between the two approaches, one could argue that it is possible to get an accurate prediction of the statistical circuit behavior using the traditional worst-case analysis if only the value of sigma properly included the spatial and proximity-dependent variability. This is so because systematic variability can always be absorbed in the random variation component. Simulations show, however, that a significant prediction error is likely to occur.

The traditional statistical worst-case analysis assumes the following statistical model $L = L_{ave} + N(0, \mathbf{S}^2)$, where L_{ave} is the overall mean and \mathbf{S} absorbs all of the random Lgate variation. Then, the worst-case Lgate value is $L^{wc} = L_{ave} + 3\mathbf{S}$. For location-dependent timing analysis, both the mean and variance are proximity-dependent, and the mean is location-dependent: $L = L_{ave}^{prox}(x, y) + N(0, \mathbf{S}_{prox}^2)$, where $L_{ave}^{prox}(x, y)$ is the spatial proximity-dependent Lgate map. In this case, the worst-case Lgate value is $L^{wc}(x, y) = L_{ave}^{prox}(x, y) + 3\mathbf{S}_{prox}$.

We compared the accuracy of the above modeling approaches through the statistical worst-case simulations of the benchmark combinational circuit. The circuit was simulated using worst-case Lgate values, assuming placement at two locations within the chip: a center point and a corner. The results (Table IV) suggest that such worst-case analysis is overly pessimistic, at least for certain spatial locations. Thus, in predicting the worst-case behavior of a circuit when it is located in the center of the chip, the traditional worst-case analysis gives an error of 11%. This is a significant error for designs with tight timing constraints.

TABLE IV. Statistical worst-case timing simulations

Location Within Chip	Traditional Worst-Case	Location-Dependent Worst-Case
Center	1833ps	1652ps
Corner		1847ps

7 Conclusions

Deep sub-micron processes necessitate increased interaction between the manufacturing and design of integrated circuits. The

detailed knowledge about process characteristics, available to process engineers, typically finds its way into the CAD tools only by means of simplified worst-case models. This paper shows that for advanced processes this will increasingly result in inaccurate modeling and lost performance.

We demonstrated, using experimental evidence gathered from state-of-the-art 0.18 μm fabrication facilities, the presence of significant systematic intra-chip Lgate variability. This variability causes an error of up to 15% in timing analysis of critical paths, resulting in a corresponding performance loss. The variability also leads to increased global skew of about 8%, which can be additive to the set-up time error. We developed a theoretical framework allowing explicit analysis of circuit speed degradation due to intra-chip Lgate variability. This analysis is easily extendable to other sources of process variability. We propose a location-dependent timing analysis methodology as a way to deal with the increasing systematic intra-chip variability of Lgate. We show that the proposed methodology cannot be subsumed by a statistical, e.g. worst-case, timing analysis. In a situation of multiple chips per reticle field, one can either treat the problem as a multi-design problem (high-performance), or 'collapse' timing information into a single set of timing characteristics (simplicity). Finally, given the potential for significant performance degradation demonstrated in this paper, the most direct solution to spatial Lgate variability may be to apply correction at the masks, in addition to optical proximity correction and phase-shift masking.

8 Acknowledgements

The authors wish to acknowledge Advanced Micro Devices for providing experimental data used in this work and for financial support through the AMD/SRC Fellowship.

9 References

- [1] A. Kahng, Y. Pati, "Subwavelength optical lithography: challenges and impact on physical design," *Proceedings of ISPD*, p.112, 1999
- [2] S. Nassif, "Within-chip variability analysis," *IEDM Technical Digest*, p.283, 1998.
- [3] V.Mehrotra, S.Nassif, D.Boning, J.Chung, "Modeling the effects of manufacturing variation on high-speed microprocessor interconnect performance," *IEDM Technical Digest*, p.767, 1998.
- [4] C. Yu et al, "Use of short-loop electrical measurements for yield improvement," *IEEE Trans. on Semiconductor Manufacturing*, vol. 8, no. 2, May 1995.
- [5] B. Stine, D. S. Boning, J. E. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Trans. On Semiconductor Manufacturing*, No.1, pp. 24-41, Feb. 1997.
- [6] *HSPICE User Manual*, Avant!, 1999.
- [7] Benchmark Combinational Circuits, *ISCAS*, 1985.
- [8] *PathMill User Guide*, Synopsys, 1999.
- [9] S. Nassif, "Statistical worst-case analysis for integrated circuits," *Statistical Approaches to VLSI*, Elsevier Science, 1994.
- [10] P. Yang et al, "An integrated and efficient approach for MOS VLSI statistical circuit design," *IEEE Trans. on CAD*, No 1, Jan. 1986.
- [11] Semiconductor Industry Association, *National Technology Roadmap for Semiconductors*, 1997.
- [12] D. Sylvester, K.Keutzer, "Getting to the bottom of deep-submicron," *Proc. IC-CAD 1998*.