

Impact of *Within-Die* Parameter Fluctuations on Future Maximum Clock Frequency Distributions

Keith A. Bowman and James D. Meindl

Georgia Institute of Technology, Atlanta, GA

Abstract

The impact of parameter fluctuations on future circuit performance is evaluated by employing rigorously derived device and circuit models to calculate the critical path delay distributions resulting from die-to-die and within-die fluctuations. Utilizing these distributions with a recently derived FMAX distribution model validated by measured data, the effect of within-die fluctuations on the FMAX mean is forecast for the 180, 130, 100, 70 and 50 nm technology generations. Systematic within-die fluctuations impose the largest performance degradation resulting from parameter fluctuations. Assuming a 3σ channel length deviation of 20%, projections for the 50 nm technology generation indicate that essentially a generation of performance gain can be lost due to systematic within-die fluctuations. This analysis should encourage efforts toward tightening within-die process controls and developing circuit design methodologies that suppress the impact of within-die parameter fluctuations on circuit performance.

I. Introduction

A processor's maximum clock frequency (FMAX) distribution is significantly influenced by the magnitude of critical path delay deviations resulting from both die-to-die (inter-die) and within-die (intra-die) fluctuations [1]. FMAX is a measurement performed at wafer sort in which each functional die is tested for its maximum operating clock frequency. Die-to-die fluctuations resulting from lot-to-lot, wafer-to-wafer and some sources of the within-wafer variations affect every element on a chip equally. Conversely, within-die fluctuations consisting of both random and systematic components produce a non-uniformity of electrical characteristics across the chip [2].

Integrated circuits have always been vulnerable to inherent die-to-die and within-die fluctuations in the manufacturing process. Traditionally, die-to-die fluctuations have been the main concern for CMOS digital circuit designs, and the within-die fluctuations were neglected [2, 3]. As polysilicon gate lengths have decreased below the wavelength of light used in the optical lithography process, however, the systematic and random within-die fluctuations of channel length have exceeded the die-to-die fluctuations [2]. Moreover, as MOSFET's continue to scale, the fluctuations in the location of dopant atoms in the device active region induce drain current fluctuations. This is an intrinsic effect since it cannot be eliminated by external

control of conventional manufacturing processes [4]. Thus, within-die fluctuations are a growing threat to the performance and functionality of future gigascale integration (GSI) circuits.

The importance of accurately estimating the impact of parameter fluctuations on circuit performance is directly related to a company's overall *revenue*. An overestimation increases the design complexity, possibly leading to an increase in design time, an increase in die size, rejection of otherwise good designs and even missed market windows [2]. Conversely, an underestimation can compromise the product's performance and overall yield as well as increase the silicon debug time [2]. In summary, overestimating fluctuations impacts the design effort, and underestimating fluctuations impacts the manufacturing effort.

Previous research modeled the FMAX distribution to accurately describe the impact of the individual contributions resulting from die-to-die and within-die parameter fluctuations [1]. This work reveals that within-die fluctuations directly impact the FMAX mean and die-to-die fluctuations determine the majority of the FMAX variance. Employing the recently derived FMAX distribution model [1] and a generic critical path model described by physical device and circuit analyses [5, 6], a new circuit-level methodology enables projections of the impact of die-to-die and within-die fluctuations on future circuit performance. *The research objective of this paper is to evaluate the limitations imposed by parameter fluctuations to facilitate opportunities for further advancement of GSI systems.*

II. FMAX Distribution Model

The FMAX distribution model [1] is based upon statistical circuit simulations of die-to-die (D2D) and within-die (WID) parameter fluctuations on several critical paths for a recent 0.25 μm microprocessor. The D2D and WID circuit simulations determine the nominal critical path delay and the delay standard deviations resulting from D2D and WID fluctuations. The impact of D2D and WID fluctuations on *one* critical path is modeled as a normal distribution. Since D2D fluctuations have an equal effect on each element across the chip, only *one* distribution is necessary to model its impact. WID fluctuations affect the elements on a chip non-uniformly such that each critical path must be statistically combined to determine the WID maximum delay distribution for the entire chip. Assuming a number (N_{cp}) of independent critical paths for the entire chip, the WID maximum critical path delay distribution is derived. It is important to note that

for small values ($1-10^2$) of N_{cp} the dependency of the WID distribution on N_{cp} is significant; however, as N_{cp} increases to larger values (10^2-10^4), the dependency of the WID distribution on N_{cp} decreases substantially. As the number of transistors per chip increases and the number of average gate delays per critical path is reduced [7], N_{cp} is expected to increase for each technology generation, thus diminishing the relative sensitivity of the FMAX predictions to N_{cp} .

Figure 1 compares the FMAX distribution model [1] for both D2D and WID fluctuations with $N_{cp}=100$ against FMAX measured data obtained at wafer sort for the 0.25 μm microprocessor. The predicted FMAX distribution agrees closely with the distribution of measured data in *mean*, *variance* and *shape*. Figure 1 also plots the distributions resulting from only D2D and only WID fluctuations to illustrate the individual effects on the FMAX distribution. These results clearly reveal that *within-die fluctuations directly impact the FMAX mean; and die-to-die fluctuations the FMAX variance* [1].

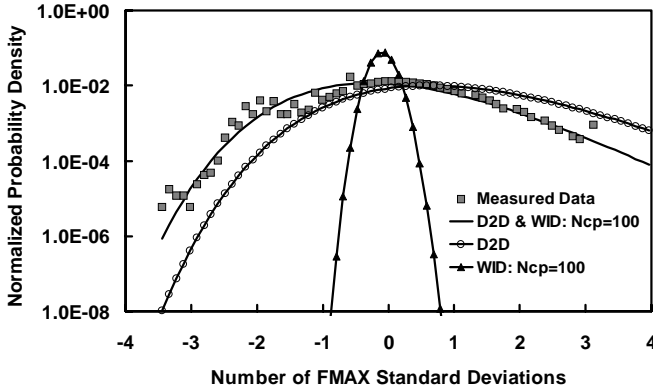


Figure 1. Comparison of the FMAX model projections with measured data. The individual contributions of die-to-die (D2D) and within-die (WID) fluctuations to the FMAX distribution are also provided.

III. Generic Critical Path (GCP) Model

The critical path delay distributions resulting from die-to-die and within-die fluctuations are calculated from D2D and WID statistical simulators, which both use a SPICE-equivalent circuit simulator [8] with a 0.25 μm process file and a netlist of speed-limiting paths from the microprocessor. Many of the device parameters provided in the process file are empirically calculated to fit measured $I-V$ data. In projecting the impact of parameter fluctuations on future circuit performance, it is unclear how these empirical parameters might scale with technology. Therefore, a generic speed-limiting path model is developed through physically based device and circuit analyses [5, 6] to evaluate the critical path delay distributions resulting from D2D and WID fluctuations.

As illustrated in Figure 2, the generic critical path (GCP) is modeled by a number (n_{cp}) of identical two-input

static CMOS NAND gates with a fan-out of three, where each gate drives an average wiring capacitance. The static CMOS logic gate is chosen for its low standby power drain, large operating margins, scalability and flexibility of logic functions [9]. The average propagation delay through a two-input NAND gate is modeled by averaging the delay through two series-connected NFET's and the delay through one PFET given as

$$T_{nand} = \frac{f_{ineff} T_{pd,n} + T_{pd,p}}{2}, \quad (1)$$

where f_{ineff} is the effective fan-in factor [10] for series-connected MOSFET's and $T_{pd,n}$ and $T_{pd,p}$ are the NFET and PFET CMOS propagation delays, respectively, as derived from the *physical* alpha-power law model [5]. The critical path delay is then calculated as

$$T_{cp} = n_{cp} T_{nand}. \quad (2)$$

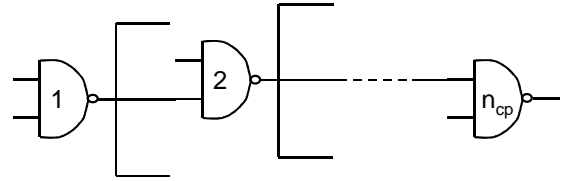


Figure 2. Generic critical path (GCP) model.

In calculating the critical path delay distributions that are used by the FMAX distribution model for the analysis in Figure 1, a rigorously developed WID fluctuation model is employed. This model is empirically derived through an analysis of manufacturing data specific to the 0.25 μm technology generation. The WID fluctuation model represents systematic within-die variations by expressing the device-to-device correlation as a function of the distance between the devices. This correlation function, however, is significantly influenced by specific manufacturing capabilities. Currently, there is little insight into understanding how this distance correlation might scale for future technology nodes. Therefore, the GCP model analyzes two separate WID fluctuation cases: 1) completely dependent gates (*gate correlation=1*) and 2) completely independent gates (*gate correlation=0*), which may be viewed as extreme conditions of systematic and random fluctuations, respectively.

Using only the D2D and WID device parameter standard deviations for the 0.25 μm technology, critical path delay distributions are calculated through the GCP model for a gate correlation of one and zero. The results of these critical path delay distributions are inputs into the FMAX distribution model [1]. Figure 3 compares the cumulative distributions of FMAX projected by the GCP model with a gate correlation of one and zero to measured and simulated distributions. The circuit simulation clearly provides much better agreement with measured data than either of the GCP projections due to the accuracy of the systematic WID correlation model. *The GCP model, however, enables a key insight into the projections by establishing boundaries of the*

actual FMAX distribution with the two extreme cases of completely systemic and completely random within-die fluctuations. Moreover, Figure 3 illustrates that systematic within-die fluctuations (gate correlation=1) decrease the FMAX mean more severely than random within-die fluctuations (gate correlation=0).

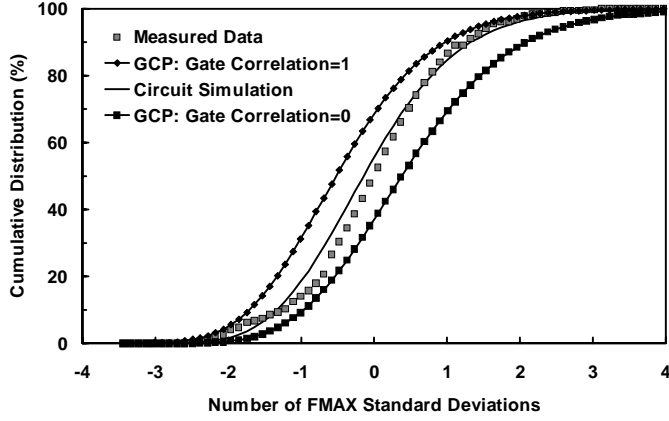


Figure 3. Comparison of the FMAX projections using both the GCP model and the circuit simulation with measured data.

This result can be explained physically as follows. In the completely systematic case, the variations have the same impact on every element in a critical path so that

$$\frac{\sigma_{T_{cp}}}{T_{cp}} = \frac{n_{cp} \sigma_{T_{nand}}}{n_{cp} T_{nand}} = \frac{\sigma_{T_{nand}}}{T_{nand}}, \quad (3)$$

where $\sigma_{T_{cp}}$ and $\sigma_{T_{nand}}$ are the standard deviations of the critical path delay distribution and the NAND gate delay distribution, respectively. For completely systematic WID fluctuations, the ratio of the standard deviation to mean is equal for the critical path delay distribution and the gate delay distribution. In the case of completely random fluctuations, however, the fluctuations in the critical path delay are expected to have an averaging effect over the number of gates in the path [3] such that

$$\frac{\sigma_{T_{cp}}}{T_{cp}} = \frac{\sqrt{n_{cp}} \sigma_{T_{nand}}}{n_{cp} T_{nand}} = \frac{1}{\sqrt{n_{cp}}} \frac{\sigma_{T_{nand}}}{T_{nand}}. \quad (4)$$

For completely random WID fluctuations, the ratio of standard deviation to mean for the critical path delay distribution is inversely proportional to the square root of n_{cp} [3]. Thus, for n_{cp} greater than one, systematic WID fluctuations induce a larger amount of performance degradation than random WID fluctuations.

IV. Projecting the Impact of Within-Die Fluctuations on Future FMAX Distributions

In projecting the impact of parameter fluctuations on future circuit performance, the nominal values in Figure 4 are selected judiciously by using the International Technology Roadmap for Semiconductors (ITRS) [11] as a guideline. As discussed earlier, the WID parameter fluctuations directly

impact the FMAX mean. Using the GCP and the FMAX distribution models [1], Figure 5 projects the impact of within-die fluctuations on the FMAX mean for the 180, 130, 100, 70 and 50 nm technology generations. The GCP model assumes the 3σ effective channel length deviation is 20% of the nominal gate length [11]. Figure 5 provides a range of percentages (50 and 100%) for the ratio of the WID channel length variance to the total channel length variance (WID and D2D). These ranges are plotted for the GCP model using a gate correlation of one (completely systematic WID fluctuations) and a gate correlation of zero (completely random WID fluctuations).

Technology Generation (nm)	180	130	100	70	50
L_{Gate} (nm)	140	85	65	45	32
t_{OX} (nm)	2.5	1.9	1.5	1.5	1.5
V_{DD} (V)	1.52	1.40	1.04	1.00	0.70
V_{TL} (V)	0.30	0.30	0.28	0.38	0.36
N_A ($\times 10^{18} \text{ cm}^{-3}$)	0.85	1.37	1.88	3.00	2.74
n_{cp} (#gates/CP)	10	9	8	7	6
N_{cp} (#CP/chip)	10^2	10^3	10^3	10^4	10^4
I_{OFF} (nA/ μm)	5	10	20	40	80
$F_{CLK,nom}$ (GHz)	1.25	2.10	3.50	6.00	10.00

Figure 4. Nominal values used in the projection analysis.

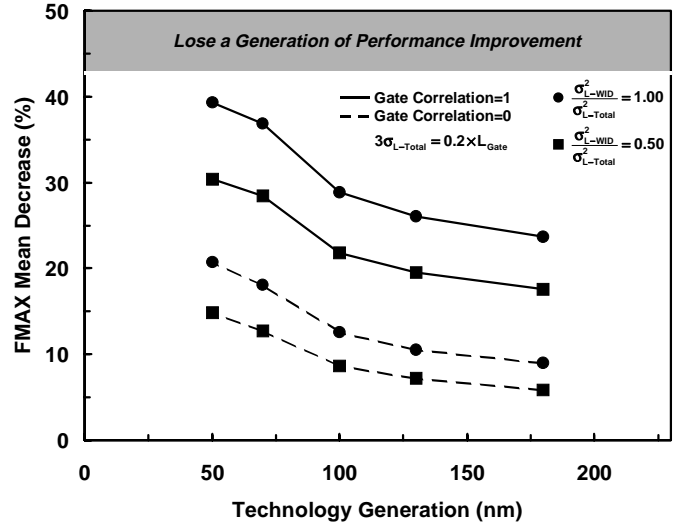


Figure 5. Reduction in FMAX mean resulting from within-die parameter fluctuations versus technology generation.

Figure 5 includes a shaded region to indicate the *limit* at which within-die parameter fluctuations degrade the FMAX mean such that the performance gained from a generation of transistor scaling is completely lost. Typical technology scaling decreases the gate delay (T_{Gate}) by 30% [12]. The improvement in clock frequency resulting strictly from technology scaling is

$$\%F_{CLK} \text{ increase} = \frac{\frac{1}{0.7T_{Gate}} - \frac{1}{T_{Gate}}}{\frac{1}{T_{Gate}}} \approx 43\%. \quad (5)$$

The clock frequency improvement from one technology generation to another is also aided by architecture advances, such as reducing the number of gate delays in a critical path. The *limit* at 43% provides a criterion for evaluating the impact of the WID fluctuations on circuit performance.

In analyzing Figure 5 for the 50 nm technology generation, the GCP model using the gate correlation of one projects a degradation in the FMAX mean of 30 and 39% corresponding to ratios of WID channel length variance to total channel length variance of 50 and 100%, respectively. For the same technology node and ratios of channel length variance, the GCP model using a gate correlation of zero predicts a decrease in performance of 15 and 21%. Figure 5 indicates that the performance degradation resulting from systematic WID fluctuations is much worse than the performance loss resulting from the random WID fluctuations. Since WID fluctuations directly impact the FMAX mean, *the systematic within-die fluctuations are the most significant performance limiter resulting from parameter fluctuations*. This result is of concern since early characterizations of a 0.18 μm manufacturing process indicate a more systematic than random WID fluctuation [13]. Figure 5 projects that essentially *a generation of performance gain can be lost due to systematic within-die fluctuations* at the 50 nm technology node. These results should motivate efforts towards narrowing WID process tolerances and improving circuit design methodologies that suppress the impact of WID parameter fluctuations on future circuit performance.

V. Conclusion

The impact of parameter fluctuations on future circuit performance is analyzed by using a physically based generic critical path model to determine the critical path delay distributions resulting from die-to-die and within-die fluctuations. Utilizing the results of these distributions with the recent FMAX distribution model, projections are made for the 180, 130, 100, 70 and 50 nm technology generations. *Results indicate that systematic within-die fluctuations are the most significant performance limiter resulting from parameter fluctuations*. Assuming a 3σ channel length deviation of 20%, projections for the 50 nm technology generation indicate that approximately *a generation of performance improvement can be lost due to systematic within-die fluctuations*. As device fluctuations increase with decreasing dimensions and the probability of longer critical path delay deviations increases with growing transistor

density, *within-die fluctuations may become the key obstacle to achieving GSI*.

Acknowledgements

The authors gratefully acknowledge the support of SRC, DARPA and the Intel[®] TCAD Group. The authors are especially indebted to Steve Duvall of Intel[®] for his essential contribution to this work. The authors would also like to express their sincere appreciation to George Sery, Adam Brand, Nagib Hakim, Russell Gee, Sam Hu, Mary Wesela, Evan Cohn of Intel[®] and Jim Joyner of Georgia Tech for their helpful advice and kind support.

References

- [1] 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," to be presented at the *IEEE ISSCC*, Feb. 2001.
- [2] S. G. Duvall, "Statistical Circuit Modeling and Optimization," *5th International Workshop on Statistical Metrology*, June 2000, pp. 56-63.
- [3] M. Eisele, J. Berthold, D. Schmitt-Landsiedel and R. Mahnkopf, "The Impact of Intra-Die Device Parameter Variations on Path Delays and on the Design for Yield of Low Voltage Digital Circuits," *Proc. of the 1996 ISLPED*, Aug. 1996, pp. 237-242.
- [4] K. A. Bowman, X. Tang, J. C. Eble and J. D. Meindl, "Impact of Extrinsic and Intrinsic Parameter Fluctuations on CMOS Circuit Performance," *IEEE J. Solid-State Circuits*, pp. 1186-1193, Aug. 2000.
- [5] K. A. Bowman, et al., "A Physical Alpha-Power Law MOSFET Model," *IEEE J. Solid-State Circuits*, pp. 1410-1414, Oct. 1999.
- [6] B. Agrawal, V. K. De and James D. Meindl, "Opportunities for Scaling FET's for Gigascale Integration (GSI)," *Proc. of the 23rd ESSDERC*, Sept. 1993, pp. 919-926.
- [7] P. E. Gronowski, et al., "High-Performance Microprocessor Design," *IEEE J. Solid-State Circuits*, pp. 676-686, May 1998.
- [8] Meta-Software Inc., *HSPICE User's Manual*, Mar. 1995.
- [9] J. D. Meindl, "Low Power Microelectronics: Retrospect and Prospect," *Proc. IEEE*, pp. 619-635, Apr. 1995.
- [10] T. Sakurai and R. Newton, "Delay Analysis of Series-Connected MOSFET Circuits," *IEEE J. Solid-State Circuits*, pp.122-131, Feb. 1991.
- [11] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors (ITRS)*, 1999.
- [12] V. De and S. Borkar, "Technology and Design Challenges for Low Power and High Performance," *Proc. of the 1999 ISLPED*, Aug. 1999, pp. 163-168.
- [13] M. Orshansky, et al., "Intra-Field Gate CD Variability and Its Impact On Circuit Performance," *IEEE IEDM Tech. Digest*, Dec. 1999, pp. 479-482.