RC23303 (W0408-104) August 24, 2004
Electrical Engineering

IBM Research Report

Model to Hardware Closure for nm Generation Technologies

Sani R. Nassif
IBM Research Division
Austin Research Laboratory
11501 Burnet Road
Austin, TX 78758

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Model to Hardware Closure for nm Generation Technologies

Sani R. Nassif,
IBM Austin Research Laboratory,
11501 Burnet Rd., Austin, TX 78758
(nassif@us.ibm.com)

Categories and Subject Descriptors: B.8.2 [Performance and Reliability]: Performance Analysis and Design Aids

General Terms: Design.

Keywords: Variability, Uncertainty, Integrated Circuit.

ABSTRACT

The integrated circuit manufacturing process has inevitable imperfections and fluctuations that result in ever-growing systematic and random variations in the electrical parameters of active and passive devices fabricated. The impact of such variations on various aspects of chip performance has been the subject of numerous recent papers, and techniques for analyzing and dealing with such variability -broadly labeled *design for manufacturability* are emerging as the next hot topic in this area [18].

The focus of much of the current work in this area has been on timing, but it is also well known that modern integrated circuits are very heavily power limited and that static (leakage) and dynamic power have emerged as first class design objectives. This means that the same level of investment is going to be needed in order to characterize, analyze and optimize for power and its variability as has been the case for timing.

In the timing area, there has been a major recent push for *Statistical Static Timing* which promises to combine the convenience and efficient of static timing analysis with accurate models of physical and environmental variability that account for both temporal and spatial sources of correlation in the parameters that control timing. The outcome of this effort is to be a tool that allows the designer to make (timing) yield vs. performance trade-offs and design optimization for robustness.

Critical to achieving the goal above is the knowledge of the sources, magnitudes, as well as spatial and temporal sources of correlation in the various parameters that model the sources of physical and environmental variability. This is a problem that is far from solved, and is the subject of this work.

In this paper, we review the various sources of variability that impact circuit performance -with a special emphasis on timing and on power. We then propose the notion of Model to Hardware correlation, defined as the set of activities that are implemented to characterize, model and simulate the behavior of a design in order to insure predictability of manufacturing results.

1. PROCESS VARIABILITY

Recently, statistical static timing analysis (STA) has been proposed as probably the only viable methodology for handling process variability of high performance large digital integrated circuits [2, 3, 4, 5, 6]. In Statistical STA, models of the timing variability of all gates and wires in a digital integrated circuit are used to predict the variability in the resulting design performance. These variability models must typically account for various sources and distributional assumptions of parametric variability. Statistical STA research has focused on the propagation and correlation of variability through these models and ultimately to the overall cycle time and timing margin of a design.

1.1 Physical vs. Environmental Variability

As we examine this phenomena further, we find that the sources of timing or delay variability can be categorized into (a) physical and (b) environmental types:

- Physical sources include lithography-induced systematic and random variations in critical device dimensions such as transistor length and width [7], as well as wire and via width. They also include random phenomena like the impact of discrete doping placement on MOSFET threshold voltage [8], and systematic phenomena like inter-layer dielectric thickness variations with layout density [9]. By and large, these sources of variability are constant with respect to time.

- Environmental sources of variation include power supply noise [14, 24], temperature variation [22, 23] across the die due to uneven power dissipation, and line-to-line coupling between neighboring wires[15]. These sources are time dependent and have a large range of temporal time constants that range from the pico-second for line-to-line coupling to the milli-second for temperature effects.

Both the environmental and physical sources of variation are function of the details of the design implementation. For example, device dimension variations due to lithography are

a strong function of local layout detail, while power supply variations are clearly a function of the placement and power grid design. This means that the assessment of the impact of these sources is difficult to achieve without knowledge of the specifics of the design. This has profound implications on the applicability of statistical STA in the context of a realistic design flow.

1.2 Spatial and Temporal Variability Distribution

In addition to the physical/environmental breakdown, parametric variability also exhibits spatial and temporal distribution properties that are important from a modeling and analysis point of view [10, 11]. Spatial variations are often caused by wafer-level phenomena that cause the value of some parameter to depend on the location of the chip within the wafer (we will examine this aspect of variability again in the next section).

Spatial variations can be caused by environmental sources of variation. For example, varied levels of power density across the chip will cause a variation in temperature across the die, change the electrical parameters of MOSFETs and wires, and hence impact the delay and other important performances such as leakage [19]. Note that this type of spatial variation is naturally a within-die phenomena.

Spatial variations can also be caused by physical phenomena that occur within-die (as opposed to the within-wafer example above). An excellent example of such a phenomena is inter-layer dielectric thickness variation with layout shape density due to chemical-mechanical polishing [9].

It has been observed [12, 13] that these within-die variations are significantly more difficult to deal with than the more tradition die-to-die variations which are typically accommodated using worst-case corners.

The temporal distribution of variations has a variety of time constants ranging from the design's clock cycle -for phenomena like line-to-line coupling, to months or even years of time for some reliability-related phenomena like electromigration. The challenge of characterizing and predicting the effect of such temporal variations lies in the need to delineate the various phenomena in such a manner that they can be analyzed effectively and separately. For example, if one wanted to analyze the impact of line-to-line coupling at the same time as the impact of L di/dt package induced noise, one would have to carefully examine the temporal relationship between the system clock and the overall resonance between the chip and package in order to find the correct worst-case alignment to insure that these particular sources of variability are correctly handled.

1.3 Random vs. Systematic Variability

We further differentiate variability into random and systematic components in relation to the manner in which it occurs. A component of variability referred to as random is not dependent on design details, although its impact may very well be. Particulate contamination, and the resulting defects it causes, is an example of a random source of variability [17]. Another example is threshold voltage variation due to discrete doping effects [8].

Systematic variability refers to randomly distributed phenomena the magnitude of which is highly correlated with specific design details. One example might be the methodical difference between the deviation in channel length of nested and isolated PolySilicon structures. Another example would be the variation in power supply voltage across the die due to changes in circuit type and activity across the integrated circuit. A third example is the variation in certain physical or electrical parameters across the wafer, often causing specific sites on the wafer to have a higher probability of failure than others.

In older technologies, systematic variability was dominated by sources of variability related to wafer gradients and environmental operation factors. In more recent technologies, within-die physical variability like variations in MOSFET channel length and in wiring dimensions have become more important [21].

Note the link between the systematic and spatial characterizations of variability, in the sense that one usually implies the other.

1.4 Variability and Uncertainty

Thus far we have considered the sources (physical vs. environmental), distributions (spatial and temporal) as well as type (random vs. systematic) of variability. The assumption we made is that we have an essentially infinite amount of resources and knowledge about each phenomena, and that we are able to create a model for each source and type of variability and to analyze such a model in order to understand its impact on design performance.

In reality, however, it is often the case that we do *not* have sufficient knowledge about all phenomena. This can occur because the phenomena is difficult to model, as is the case -for example- with changes in MOSFET channel dimensions with respect to layout practice and various resolution enhancement techniques commonly applies in sub-100nm technology nodes.

This can also occur, however, not because of the intrinsic difficulty of generating a model for variability, but because of staffing or other resource limitations. In such a situation, a source of systematic variability may end up being treated as if it were random, and we would refer to it instead as an *uncertainty*. Broadly speaking, systematic sources would be referred to as variability, while random sources would be referred to as uncertainty.

An example of such a situation would be power supply variations. If (a) the design of the chip and its power deliver system (power grid, decoupling capacitors and package) is available, and (b) if there exist sufficient engineering and analysis resources in order to simulate the behavior of the chip and the power delivery system; then one can confidently predict the value of the power supply voltage at every point in the design. Once the power supply voltage is known, it can be taken into account when predicting -say- the timing performance of the design, and suitable action -for example resizing certain gates or wires- can be taken to insure that the performance meets the requirements.

If -on the other hand- insufficient engineering resources exist, *or* perhaps earlier in the design cycle, insufficient knowledge about the details of the design exist, then it is not possible to predict the power supply voltage. In such a situation, the designer would have to revert to a worst case analysis methodology where the desire is to simple *bound* the effect of the relevant source of variability. The result might be that the power supply noise is assumed to be in some range -typically within 10% of the nominal value- and the whole design is corrected for the worst-case condition.

Obviously this results in *over-design* and can adversely impact the competitiveness of a particular chip.

The message is that whenever a source of systematic variability is ignored, worst-cased, or under-analyzed, it is equivalent to treating it as a source of random variability, thus uncertainty. Uncertainty leads to over-design, while variability leads to just-enough design.

2. CHARACTERIZATION

Design for manufacturing in general, and Statistical Static Timing Analysis in particular have an Achilles heal: Variability Characterization. Without robust accurate models of parameter variations within die, across a wafer, and from wafer to wafer, a statistical static timer is not capable of correctly predicting the distribution of design performance, and -in such cases- becomes little better than the more traditional corner-based simulation methodologies in wide current industrial use. Furthermore, because of the relative novelty of this field in the digital design area, relatively little research is ongoing in support of future efforts.

The characterization of variability is inherently more difficult than that of the nominal or extreme behavior of the manufacturing process. This is because of:

- The large amounts of data required to make estimates of distributions and correlations amongst parameters. Gathering such data requires significant investment in design, silicon, and test/characterization resources. Without an existing EDA (Electronic Design Automation) tool infrastructure to make sure of such characterization, the return on such an investment may not be sufficiently high to justify making it instead of investing in -say- improved process control.

- It is a fact that these variability distributions need to be broken down across (a) the physical implementation hierarchy: facility, lot, wafer, die, within die, and (b) temporally in order to determine and correct trends, to justify investment in design or process control policy changes. This means that variability characterization cannot be done once and once only, but needs to be something that is done continuously and monitored regularly.

- While corner or worst case analysis can be done substantially independently from the details of any one design, many of the relevant phenomena for variability require a deep understanding of design style and thus become difficult in situations when -for example- the design and fabrication groups are organizationally separated, as would be the case for a typical fabless design company using a silicon foundry.

2.1 Device Characterization

We commonly consider characterization to include the generation of circuit simulation (e.g. SPICE [20]) models of active and passive devices. This is an established area of research and development which acts as the major interface between the design and manufacturing area. There exist numerous papers on extracting transistor model parameters from measured current vs. voltage and capacitance vs. voltage curves [25, 26]. With technology scaling and the corresponding rapid increase in the complexity of MOSFET models [27], this task has been getting increasingly difficult and in many cases the *physical nature* of model parameters has had to be compromised in order to achieve better fits to ever more complex device behavior [28].

The decreasing physical significance of device model parameters is very important in the context of modeling variability. If we consider that a model of the statistical behavior of a device is to be create by generating the appropriate probability density function description of the parameters that define the model's behavior, we would be led to the fact that a physically significant statistical model would by its nature result in a set of correlated model parameters [29]. The complexity of current models and the lack of clear physical interpretation or uniqueness of fitting parameters, makes this task currently difficult and -unfortunately- getting even harder as we scale technology further and have to account for an ever increasing level of non-ideal behavior in devices.

It is the author's opinion that what is required here is a careful analysis of the nominal vs. statistical errors associated with a device model, and that the result of such an analysis will be a radically simpler device model which may lack accuracy for the nominal case, but exhibits the correct statistical behavior and is therefore predictive from the point of view of variability. The creation, characterization and deployment of such a model is a difficult task because it also involves the process characterization and circuit simulation EDA community and -unless the model becomes rapidly accepted and standardized- will have difficulty replacing the existing foundry / fabless design house interface.

2.2 Cell Level Characterization

In the early years of digital system design, it was sufficient to model the delay of elements, so digital simulation tools used step functions to represent signals. As technology scaled and switching speeds increased, however, it became increasingly necessary to take into account the transition time of the signal. The simplest way of modeling the transition time is to change the model from a step to a ramp, and this has been the state of digital design since the early Eighties. This has resulted in a number of problems, an example of which is that of threshold selection, which -when not done properly- can lead to to negative delays [32]. Furthermore, many of the deep-submicron phenomena (e.g. inductive interconnect, coupled noise, power supply current) are difficult or impossible to model accurately with the simple linear ramp model [30, 31].

The simplicity of the ramp approximation has several advantages: (1) it makes the task of building models for timing analysis easy, (2) it is conceptually easy to grasp and to translate from the model parameters to a pictorial representation or to a Spice input specification, (3) it is information dense in that two real numbers (delay, slope) and a boolean (rising/falling) completely encapsulate the behavior of the waveform in question. Characterizing these models simply requires performing a number of circuit level simulations of the cell in question varying the relevant variables; usually these are the input waveform slope and the capacitive load on the output. The resulting output waveform from each of the simulations is *processed* in order to estimate it by a ramp, and a table (which we will refer to as the timing table) is thus generated that includes:

- The input waveform parameters (one or more slopes).

- The output loading conditions (capacitance).

- The output waveform parameters (delay and slope).

From that point on, one can use either a table interpolation technique to use the data directly in a timing analyzer, or one can *fit* the observed behavior using analytical or regression formulas.

Nevertheless, extending this simple model to improve its accuracy and in order to handle statistical variations is not an easy task [33]. This is difficult because timing models (a) have a fairly large amount of error in them that can be up to 10% in certain cases, and (b) are typically not parameterized with respect to physical and environmental sources of variability.

Consider the error in timing models first. Such an error can occur if the table used to represent the behavior of the circuit in question is not sufficiently *dense* to make interpolation effective. It can also occur if the analytical or regression formula used to fit the table is not capable of exactly modeling the circuit's behavior. In either case, the error is likely to be *systematic* in that it will depend on the region of operation. If the input to the timing model has variability, and if the error is systematic, then it is not possible to distinguish variability in the output due to the input from that due to the error.

Consider now the problem of parameterizing a timing model with respect to various sources of physical and environmental variability. Each time that a new source is added, three things happen:

1. The number of simulations required to generate the timing table increases. Depending on the type of *sampling* used to generate the table, the increase may even be exponential in nature!

2. The size of the resulting timing table increases. This happens obviously because we are adding columns to the table, but also because of the increase in the number of rows in the table due to the previous point. This increase results in an increase in the computational resources required to perform a timing analysis.

3. The potential for inaccuracy in the timing model increases since we are attempting to model more dependencies simultaneously.

All of these factors will make it more difficult for the industry in general, and for the EDA timing tools in particular, to adopt and support the use of variation-aware models.

It is the author's opinion that we need to pay as much attention to timing models as we do to the algorithms that are currently being developed to perform statistical static timing analysis. We must do the theoretical and practical work needed to generate timing models with sufficient (a) fidelity to allow distinguishing real variability from model error, and (b) flexibility to allow modeling an ever larger number of sources of potential variability; Only then will we achieve the full potential of statistical STA.

3. IMPEDIMENTS AND CONCLUSIONS

Some of the major issues in variability modeling and characterization are as follows:

1. The availability of within-die variability models is directly linked to the availability of characterization structures that span a full die or even a significant part of a die. By its nature, such a structure would need to (a) span a significant amount of area, (b) include a large number of measurable devices, and therefore (c) requires significant silicon and test resources in order to measure and characterize. Designing such structures such that they mimic real circuit is crucial due to the many source of variations that depend on detailed implementation practices like device layout, interconnect layer usage assignment, and power grid interaction [16].

2. Tracking of manufacturing facility over time. Because of the cost and complexity of variability characterization, this activity often happens once only, perhaps during the bring-up phase of the technology. This means that these characterization often get *etched in stone* and do not track the process as improvements are made. The outcome is a gradual decrease in the accuracy of the variability models with respect to hardware. It is unfortunate that such drift typically happens exactly during the period in which process learning is providing improvements in overall variability.

3. Environmental sources of variability like power supply noise and across-die temperature variations are impossible to predict early in the design cycle because they depend on global placement information that is available relatively late in the design cycle. Because of this fact, these phenomena are often dealt with using traditional worst-case corner methodologies. The ability to abstract early design information to predict the variability in such environmental quantities will become important for future design, especially since power dissipation continues to increase and thus causes ever larger within-die variability in these quantities.

Another major issue in variability modeling is the limited design, silicon, and characterization resources that can be deployed to attack the problem. This economic reality means that we must do a careful analysis of all these sources of variability and attempt to find the correct balance of effort and required accuracy. Thus it is important to be able to quickly get broad bounds on each of the sources in order to insure that the appropriate level of modeling and analysis investment is made in order to bound or worst-case each component without undue pessimism. For example, attempting to model -say- temperature variability across the die may prove too costly for the amount of *pessimism removal* that it provides. In such a case, a simple worst-case bound based on the cooling capacity of the package and heat-sink might suffice.

Finally, it is also important to have a first order understanding of the technology *trends* in each of these sources of variability. This will allow the designer and CAD tool developer to anticipate future problem areas and plan new test structures, develop models, or possible create work arounds as needed.

4. REFERENCES

[1] S. Nassif "Modeling and Analysis of Manufacturing Variations," *Proceedings of CICC*, 2001.

[2] C. Viswesvariah "Statistical Timing of Digital Integrated Circuits," *Proceedings of ISSCC*, 2004.

[3] H. Chang and S. Sapatnekar "Statistical Timing Analysis Considering Spatial Correlations Using a Single PERT-like Traversal," *Proceedings of ICCAD*, 2003.

[4] A. Devgan and C. Kashyap "Block-Based Static Timing Analysis with Uncertainty," *Proceedings of ICCAD*, 2003.

[5] A. Agarwal and V. Zolotov and D. Blaauw "Statistical Timing Analysis Using Bounds and Selective Enumeration," *IEEE Trans. CAD*, Sep. 2003.

[6] A. Gattiker and S. Nassif and R. Dinakar and C. Long "Timing Yield Estimation from Static Timing Analysis," *Proceedings of ISQED*, 2001.

[7] L. Liebmann "Resolution Enhancement Techniques in Optical Lithography, It's not just a Mask Problem," *Proceedings of ISPD*, 2003.

[8] A. Bhavnagarwala and B. Austin and A. Kapoor and J. Meindl, "CMOS system-on-a-chip voltage scaling beyond 50nm," *Proceedings of GLVLSI*, 2000.

[9] V. Mehrotra and S. Nassif and D. Boning and J. Chung, "Modeling the Effects of Manufacturing Variations on High-Speed Microprocessor Interconnect Performance," *Proceedings of IEDM*, 1998.

[10] B. Stine and D. Boning and J. Chung "Analysis and Decomposition of Spatial Variation in Integrated Circuit Processes and Devices," *IEEE Trans. Semiconductor Manufacturing*, Feb 1997.

[11] M. Orshansky and C. Spanos and C. Hu "Circuit performance variability decomposition," *Proceedings of IWSM*, 1999.

[12] K. Bowman and T. Xinghai and J. Eble and J. Meindl "Impact of extrinsic and intrinsic parameter fluctuations on CMOS circuit," *IEEE Journal of Solid State Circuit*, Aug 2000.

[13] M. Orshansky and L. Milor and P. Chen and K. Keutzer and C. Hu "Impact of Systematic Spatial Intra-Chip Gate Length Variability on Performance of High-Speed Digital Circuits," *Proceedings of ICCAD*, 2000.

[14] R. Ahmadi and F. Najm "Timing Analysis in Presence of Power Supply and Ground Voltage Variations," *Proceedings of ICCAD*, 2004.

[15] R. Arunachalam and K. Rajagopal and L. Pileggi "TACO: Timing Analysis with Coupling," *Proceedings of DAC*, 2000.

[16] D. Boning and J. Panganiban and K. Gonzalez-Valentin and S. Nassif and C. McDowell and A. Gattiker and F. Liu "Test structures for delay variability," *Proceedings of TAU*, 2002.

[17] J. Khare and W. Maly "From Contamination to Detect Fault and Yield Loss," Kluwer Academic Publisher, 1990.

[18] W. Maly "The Future of IC Design, Testing and Manufacturing," *IEEE Design And Test of Computers*, Vol. 13 No. 4, 1996

[19] H. Su and F. Liu and A. Devgan and E. Acar and S. Nassif "Full Chip Leakage Estimation Considering Power Supply and Temperature Variations," *Proceedings of ISLPED*, 2003.

[20] L. Nagel "SPICE2: A Computer Program to Simulate Semiconductor Circuits," *PhD Thesis, University of California, Berkeley*, 1975.

[21] M. Eisele and J. Berthold and D. Schmitt-Landseidel 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," *IEEE Trans. VLSI*, Dec 1997

[22] G. Digele and S. Lindenkreuz and E. Kasper "Fully Coupled Dynamic Electro-Thermal Simulation," *IEEE Trans. VLSI*, Sep 1997

[23] V. Szekely and A. Poppe and A. Pahi and A. Csendes and G. Hajas and M. Rencz "Electro-Thermal and Logi-Thermal Simulation of VLSI Designs," *IEEE Trans. VLSI*, Sep 1997

[24] H. Chen and D. Ling, "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design," *Proceedings of DAC*, 1997

[25] M. Sharma and N. Arora, "OPTIMA: A Nonlinear Model Parameter Extraction Program with Statistical Confidence Region Algorithms," *IEEE Trans. CAD*, Feb 1993

[26] P. Antognetti and G. Massobrio, "Semiconductor Device Modeling with SPICE," *McGraw-Hill*, 1988

[27] Y. Cheng and M. Chan and K. hui and M. Jeng and Z. Liu and J. Huang and K. Chen and J. Chen and R. Tu and P. Kp and C. Hu, "BSIM3v3 Manual," *EECS Dept, Berkeley*, 1996

[28] M. Chan and X. Xi and J. He and K. Cao and M. Dunga and A. Niknejad and P. Ko and C. Hu "Practical Compact Modeling Approaches and Options for Sub-0.1 micro-meter CMOS Technologies," *Microelectronics Reliability*, 2003

[29] K. Krishna, "Statistical Parameter Extraction," *PhD Thesis, Carnegie-Mellon University*, 1995

[30] A. Agarwal and F. Dartu and D. Blaauw "Statistical Gate Delay Model Considering Multiple Input Switching," *Proceedings of DAC*, 2003

[31] V. Chandramouli and K. Sakallah "Modeling the Effects of Temporal Proximity of Input Transitions on Gate Propagation Delay and Transition Time," *Proceedings of DAC*, 1996

[32] F. Najm and J. Abraham "Accounting for Very Deep Sub-Micron Effects in Silicon Models," *EE-Design*, Jan 2001

[33] C. Amin and F. Dartu and Y. Ismail "Weibull Based Analytical Waveform Model," *Proceedings of ICCAD*, 2003