Statistical Characterization of Standard Cells for Variation Aware Delay Modeling

Minakshi Das
Department of Electrical Engineering,
University of California Los Angeles,
Los Angeles, CA 90095
mdas@ucla.edu

Abstract - With the advance in CMOS VLSI circuit design through technology scaling, process variability has significantly increased uncertainties in the response of sub-45nm CMOS circuits. To address this challenge and obtain reliable fabricated chips, statistical timing analysis and library characterization tools have become significantly important. CPU-time intensive Monte Carlo (MC) simulations embedded on circuit simulators become the most accurate approach for statistical characterization of standard cell libraries. However, such approaches require a prohibitive number of electrical simulations for a full library characterization. The challenge is thus to achieve near MC accuracy with a limited set of simulations. This papers looks at some interesting approaches proposed in literature. It discusses the state-of-the-art in statistical library characterization as well as surveys more advanced techniques such as Statistical-Design of Experiments for reducing the number of Monte Carlo simulations.

I. INTRODUCTION

Aggressive scaling in CMOS technology in sub 90nm nodes has created huge challenges. The existence of manufacturing variations (or process variations) in both devices and interconnect networks have become an important feature of today’s CMOS technologies. Process variations are increasing significantly due to fundamental physical limits as well as process control limitations. At present, intra-die process variations (also known as local random) mainly contribute to statistical circuit responses such propagation delay, transition time and power variability. This increase in variability at every technology node has imposed new challenges to VLSI design and characterization.

Device parameter variations are broadly classified into two basic types -- global which are same on all devices on the chip and local which differs from one device to another on the same chip. Increased local variations within a chip have increased the importance of statistical static timing analysis (SSTA). SSTA is performed by modeling gate delay as a probability distribution function of process parameters and propagating these distribution functions to compute the distribution of circuit delay. In order to achieve accurate SSTA results, accurate and efficient methods of statistical standard cell characterization are essential. The delay variations of each cell due to global as well as local intra-cell mismatch variations need to be characterized and incorporated into the statistical characterization methodology. A straightforward approach to computing cell delay variability is to assign random variables for each device in the cell. However, such a model becomes infeasible when the number of devices in gate and mismatch variables become very large. With the number of process/environmental parameters (variation sources) potentially reaching tens, the characterization run time and size of variation aware extensions of the standard-cell libraries grows significantly. A compact variation-aware standard cell model, as well as an efficient and accurate variation-aware library character methodology remains to be a challenge.

This paper discusses few such statistical cell characterization techniques. The paper is organized as follows. Section II discusses the state-of-the-art in variation aware STA and statistical library characterization which is the sensitivity based analysis (SA) approach. Section III explains the basic concepts of sensitivity based delay modeling approach. Section IV presents improvements to the SA technique to reduce number of simulations significantly. Section V summarizes work that I have done for modeling delay due to local parameter variations using Monte Carlo simulations. Section V describes the statistical Design of Experiment method that determines a small set of synthetically generated samples that represent the original statistical sample space thereby reducing the number of Monte Carlo simulations but retaining high accuracy level. Section VI concludes this paper and points out future work to be done in the project.
II. STATE-OF-THE-ART VARIATION AWARE STA

Today variation-aware static timing analysis (STA) includes either multi-corner timing analysis or statistical STA [1]. Both require specially constructed libraries providing way of computing, delays for different corners, and sensitivity of delay and slew with respect of process and environmental parameters. Multi-corner STA requires multiple standard cell timing libraries, generated using different process, voltage and temperature (PVT) parameters in the transistor level simulations. Although the computational effort is proportional to the number of required libraries, the characterization time can be significantly shortened by using parallel computing. Still, there remains several open issues: first, the input vector sensitization step is common for different corners. Second, the selection of characterization corners, can be different from that desired by variation aware STA users. This often requires multi-dimension interpolation of the models, which is prone to numerical errors.

In the case of SSTA flow, the required cell model should provide sensitivity of delay and slew to a number of process and environmental parameters. The sensitivity data can be generated in two ways. First way is similar to a multi-corner library, where a full set of simulations for characterizing a library is performed under a small change of the concerned process parameter. The resulting measurement is used to compute the sensitivities through finite differencing, which often incurs significant numerical errors. This process is repeated for each process parameter. In the special case of random process parameter each transistor within a cell has to be varied separately to obtain its net effect on the cell delay. For a cell with N transistors and r random parameters, it takes N·r simulations, bringing number of simulations for the whole library to tens of millions and storage up to the gigabyte range.

III. SENSITIVITY CHARACTERIZATION

Each cell in a library is characterized for ‘m’ global parameters and ‘n’ local parameters. If ‘p’ is the number of devices in a cell, each global variable ΔX impacts all devices identically and hence represents a single statistic. However, each local variable ΔR represents a separate random variable (statistic) for each device in the cell. The basic idea is then to extract the mean and variance of the performance metric (e.g., delay, output slew, etc.) and use them to represent the variation-aware delay. The delay of a timing arc can be represented as follows [4]:

\[ D = D_0 + \sum_{i=1}^{m} d_i \Delta X_i + \sum_{j=1}^{n} \sum_{k=1}^{p} \sigma_{jk} \Delta R_{jk} \]

where \( D_0 \) is the nominal delay value which is characterized by setting variations \( \Delta X_i \) and \( \Delta R_{jk} \) to zero. The quantities \( d_i \) and \( \sigma_{jk} \) are direct sensitivities of cell delay with respect to the global variations and local variations respectively. These are deterministic quantities obtained from characterization results. The aim of statistical characterization is to determine these \( d_i \) and \( \sigma_{jk} \) quantities, which are delay sensitivities to the global and local mismatch parameter variations. Thus, characterization of cell-delay variation due to global variations is performed by varying a given parameter such as channel length and threshold voltage for all devices in a cell. To characterize for global variations, all devices are set to a single (correlated) random global variable. The delay variation for each timing arc is determined with respect to this single parameter. However, to characterize intra-cell mismatch variations, the variations of each device in the cell impact the delay variation of timing arc. In order to characterize for the local variations, each device in the cell is assigned a separate random variable and the effective delay variation due to all such intra-cell mismatch variables need to be determined. Typically, during statistical timing analysis, for each physical variable \( \Delta R_{jk} \) the mismatch components are assumed to be independent uncorrelated variables. This enables the \( \Delta R_{jk} \) variables for different devices to be combined into a single variable and rewrite the above equation as:

\[ D = D_0 + \sum_{i=1}^{m} d_i \Delta X_i + \sum_{j=1}^{n} \sigma_j \Delta R_j \]

where \( \sigma_j \) is the equivalent delay sensitivity for local variations.

IV. MONTE CARLO SIMULATION

The most accurate method of finding delay variations due to local device parameter variations is to perform Monte Carlo simulations. It often serves as a reference for comparison against other methods. For each cell delay characterization, all NMOS and PMOS device variation parameters are varied randomly and the resultant delay variation is measured. However, Monte Carlo simulation will be prohibitively expensive especially for large number of devices and parameters. Monte Carlo simulation results are obtained to serve as a baseline for accuracy comparisons. As part of this project, I performed Monte Carlo simulations to compute the intra-cell delay and leakage power variations of standard cells from a 45nm cell library. Random samples for threshold voltage following normal distribution were generated using MATLAB. For 4 transistor gates such as NAND, NOR, Vt of each
device was assigned a random variable. Monte Carlo simulation was then performed to determine the delay probability distribution function of the gate. A database of the delay PDFs is thus created for each standard cell in the library. Figure 1 shows the delay PDF obtained from Monte Carlo (4000 points) simulation of a NOR gate.

![PDF of NOR Delay](image)

Figure 1: PDF of NOR Delay (A→Y rise) ($\mu=1.787\times10^{-11}$, $\sigma=2.625\times10^{-12}$)

V. METHODS TO REDUCE SA SIMULATIONS

Due to the huge simulation runtime and storage requirement involved with Monte Carlo simulations with large number of devices and parameters, several methods to reduce the total number of simulations have been proposed in literature.

In [3] a statistical gate-delay variation using response surface method is proposed. The model calculates intra-cell variability through sensitivity constants which are computed by considering the devices only on the transition path (charging/discharging path). Even though the intra-cell delay variance is represented finally using a single statistic, computing the sensitivity constants requires an additional ‘$p$’ characterizations where is $p$ the number of devices in the cell resulting in characterization complexity of the order of ‘$n.p$’ where $n$ is number of parameters.

In [4], authors propose a technique to reduce number of SA simulations by using a clustering approach to model intra-cell mismatch variations. Through experiments authors proved that the variations in switching devices both on the transition path and on the non-transition path are significant contributors to intra-cell delay variations. Impact of variations in non-switching devices is small and can be negligible. Taking advantage of these two properties a new approach called the clustering is proposed to characterize the cell delay variance. The basic idea of the approach is to group all devices on the NMOS and PMOS stack separately, resulting in two clusters for each cell. Then fluctuations or random variables are assigned to the cluster instead of each device. This is equivalent to mapping any combinational cell to an inverter-like structure. The runtime depends only on the number of intra-die parameters and is independent of the number of devices within the cell. Experiments indicate that the proposed approach models the delay variations due to intra-cell mismatch within 12% accuracy of Monte Carlo simulations.

The main drawback of the sensitivity characterization is that it assumes linear dependencies between the process parameter variations and the timing related metrics of the cell. It ignores the underlying correlation among device variation parameters, thus leading to inaccurate response prediction. As a consequence, these SA techniques have not been yet widely adopted by the design industry because of their lack of accuracy [5].

VI. STATISTICAL DESIGN OF EXPERIMENTS

In [5], the authors present a Variability Aware Design of Experiments (DoE) approach, called Statistical DoE Using S-DoE, the authors claim to have achieved about two orders of magnitude of better accuracy in tail response as compared to SA. S-DoE achieves simulation effort linear to the cell complexity, hence comparable to SA and requires about one to two orders of magnitude less simulations than Monte Carlo (MC).

In a standard MC based cell characterization flow, the accuracy of the delay variation is limited by the number of simulations $N$. In the proposed S-DoE flow the input parameter domain is separated into statistical ($\Delta V_t$ and $\Delta B/\beta$) and deterministic (load, slew) domains. For the statistical domain, S-DoE is used as a pre-processing step to determine a small set of ($N_{doe}$) synthetically generated points that represent the original statistical population of ($N$) $\Delta V_t$ and $\Delta B/\beta$ variants. The speedup of the flow relies on the fact that $N_{doe}<<N$. Thus, the number of simulations is much smaller than with MC. This way, first the circuit response is obtained for the $N_{doe}$ selected points through simulations. Then, an optimal response surface method (RSM) is built using a non-linear regression model that relates inputs to outputs and predicts the outcome of simulations for any other non-simulated point of the statistical domain.

[5] compares existing DoE alternatives and shows how SA is a very simple but powerful form of DoE where points are positioned along the parameters axes (Figure 2). The drawback is that the selected points ignore the existing correlations between variation parameters. Therefore it becomes impossible to build
response models predicting cross term interactions between them.

On the other hand, S-DoE detects the statistical correlations present in the parameter domain and places the DoE points along their main correlation axes (Figure 3) so that all selected points have an equal weight regarding the way statistical population is represented.

The achievement of S-DoE is that it considers the statistical nature of the process variation parameters and their correlations. By using a minimum of 2n + 1 points (n is the number of variation parameters of the circuit), it guarantees a response model with cross-terms enabling better accuracy than any of the earlier approaches. However, as compared to SA which requires ‘n’ simulations, S-DoE requires 2n+1 simulations. Hence, S-DoE has twice overhead compared to SA, still 10x less than MC for complex cells, and 100x less for simpler ones.

VII. CONCLUSION AND FUTURE WORK

A study on statistical characterization of standard cells has been performed. Both state-of-the-art and improved techniques of reducing the number of MC simulations have been discussed. The sensitivity analysis based approaches have been found to be effective as well as simple to implement. However, they ignore the underlying correlation between device parameters. The S-DoE based approach has been proved to be accurate at the same time as efficient as SA approach. In the first phase of the project, I have modeled device local parameter variations as random variables and obtained delay PDFs through MC simulations. These can be treated as golden or reference delay distribution to compare against alternate methods.

The future work for this project would be implemented some of the methods discussed in this paper particularly the S-DoE approach. Comparison can be made between SA and S-DoE approaches in terms of reduced number of simulations and accuracy of results. I shall continue to search for alternate methods and will try to implement if alternate methods are found to be attractive.

REFERENCES