

Circuit Analysis and Optimization under Manufacturing Variations and Impact on Dynamic Voltage Scaled Systems

Sachin Idgunji[†] and Ayhan Mutlu[§]

[†] ARM Incorporated, 150 Rose Orchard Way San Jose, CA 95134.

[§]Extreme DA Corporation, 3211 Scott Blvd., Santa Clara, CA 95054.

Abstract—In this paper we present a flow to efficiently determine process variation effects on the timing of circuits. The methodology involves modeling circuit performance using quadratic response surfaces generated by transistor level simulations. We employ techniques significantly reducing the model generation cost, making quadratic response surface model generation feasible in practical applications. Using these models, an iterative optimization procedure is employed to optimize the circuit for its desired yield. We also study the impact of voltage scaling in designs that use dynamic voltage and frequency scaling (DVFS) on timing variability, exploring the limits due to this variability on the gains in power when operating at lower voltages.

I. INTRODUCTION

With scaling of semiconductor technologies, process variation has emerged as a key contributor to circuit variability [1], [2]. At lower technology nodes the impact of within die variability has been increasing such that it is insufficient to analyze the performance of the design at the corners. Historically, designers adopt corner analysis, assuming that a circuit that performs adequately at the extremes should perform properly at nominal conditions [3]. These traditional corner methods do not provide adequate information about the parametric yield and robustness of the design [4]–[6]. These issues with corner analyses led to the development of statistical techniques, which rely on more accurate representation of uncertainty and its impact on circuit functionality and performance.

Response surface models are one of the commonly used techniques to capture circuit performance variability caused by manufacturing variations. These models approximate circuit performance as a polynomial function of varying process parameters. Historically many of these models have been using linear approximations [7]. If the variations are sufficiently small, the circuit performance f can be approximated in a form of linear response surface, as shown below.

$$f(X) = B^T X + C \quad (1)$$

where $X = [x_1, x_2, x_3, \dots, x_n]^T$ is an array containing design, process, and environment parameters, $B \in R^n$, and $C \in R$ stand for the model coefficients, and n is the total number of parameters. However, these linear approximations are not accurate enough for modeling large-scale process

variations present in the current technologies, specially when one employs dynamic voltage scaling where the impact of process variability is exacerbated. Therefore, it is necessary to utilize high-order response surface models to guarantee high approximation accuracy. A quadratic response surface model can be written as,

$$f(X) = X^T A X + B^T X + C \quad (2)$$

where $A \in R^{n \times n}$ denotes the quadratic coefficients, $B \in R^n$ represents the linear coefficients, and $C \in R$ is the constant term. One challenge in higher response surface models is the number of samples required to generate the model. Especially, in today's technologies where intra-die variations are becoming increasingly important. These variations need to be modeled using many additional random variables proportional to the number of the transistors in the circuit. This makes model fitting more expensive. For example, for quadratic response surface fitting, at least $O(n^2)$ sampling points are required. The key disadvantage of quadratic response surface models is the need to compute all the elements of matrix A . In many practical problems, this matrix is sparse and rank deficient. Therefore, instead of finding the full matrix A , we approximated A with another low rank matrix A_L [8]. A_L is being computed by minimizing Frobenius norm, which is square root of the sum of the squares of all the elements of $A - A_L$. For any symmetric matrix $A \in R^{n \times n}$, the optimal rank- p approximation with minimum Frobenius norm is,

$$A_L = \sum_{i=1}^p \lambda_i P_i P_i^T \quad (3)$$

where λ_i is the i -th dominant eigenvalue and $P_i \in R^n$ is the i -th dominant eigenvector. The main advantage of rank- p projection is that, to approximate matrix $A \in R^{n \times n}$ only λ_i and P_i , where $(i = 1, 2, \dots, p)$ need to be determined. This reduces the number of unknowns to $O(pn)$ with a minimal compromise in accuracy of the model. Figure 1 gives the percent error between the low-rank model and Monte Carlo in the delay variability of a d-flop using 45nm process for a different number of samples. As seen from the figure, low-rank quadratic model gives the accuracy of about 4000 Monte

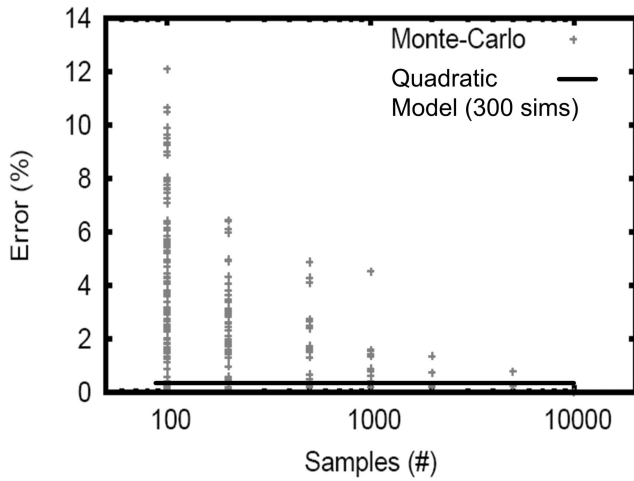


Fig. 1. Percent error in clk-to-q delay variation.

Carlo samples with sample size less than 10% of the number samples required by Monte Carlo.

The flow presented here allows users to improve the circuit performance while simultaneously ensuring a manufacturable design. This capability has been achieved through the use of advanced, application-specific numerical modeling and optimization methods.

II. RESULTS

Figure 2 gives the statistical analysis and optimization flow utilized in this work. The flow starts with a desired design performance specifications and an initial working design. Then, we generate quadratic response surface models for the performance specifications as a function of design, environment and process variables.

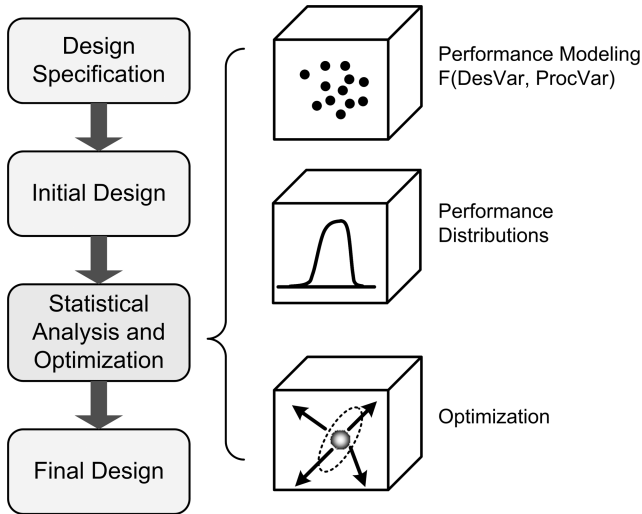


Fig. 2. Statistical analysis and optimization flow.

In our studies we used a d-flop manufactured at 45nm process technology. There were 211 local, global process

variables, and design parameters in the system. First we performed screening to filter out the parameters that do not have significant effect on the measured responses. This filtering process has brought the number of parameters down to 25. Then, in the statistical analysis phase the response surface model has been constructed. Table I gives normalized model accuracy comparison results to 5000 Monte Carlo runs. As depicted from the figure, with 200 simulations the model gives an excellent accuracy at mean, sigma, and worst performance corner (99.7%) prediction. One important feature of the

TABLE I
MODEL VERSUS MONTE CARLO COMPARISON

Performance	Model			Monte Carlo		
	Mean	Sigma	Worst	Mean	Sigma	Worst
Delay (Clk-to-Q)	0.997	0.916	0.998	1.0	1.0	1.0
Energy	1.0	0.9375	0.997	1.0	1.0	1.0
Number of samples	200			5000		

quadratic model is that it captures interaction between process, design, environment variables, and measured performances. By changing one, users can observe the effect of others on the performance variations. Figure 3 gives a delay distribution change when we change the width of a selected transistor. As depicted from the figure, increasing the width would result in a tighter delay distribution. These studies were all performed on

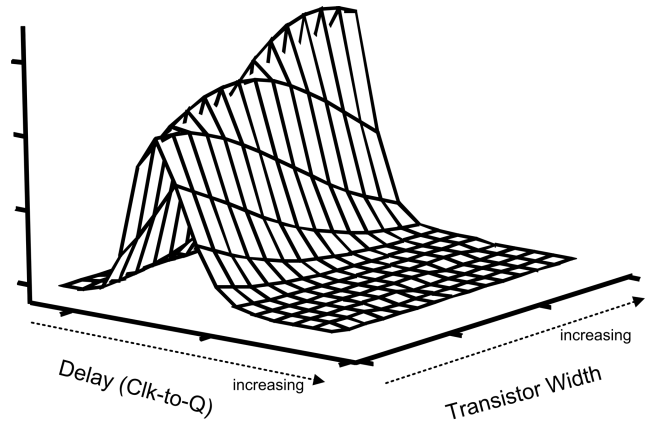


Fig. 3. Change in Clk-to-Q delay variation with respect to change in a transistor width.

the quadratic model generated with 200 simulations. In order to be able to perform such analyses with Monte Carlo requires, say N , samples for each width change. For 5 different widths the sample size would be $5N$, which could take significant time depending on the value of N being used.

In the optimization phase, we tried to minimize the delta of the setup time distribution while keeping clk-to-q delay, setup time, and energy dissipation under certain values. Delta is defined as the difference between $+3\sigma$ and -3σ of the setup time distribution. In each optimization iteration, we generated an response surface model, and moved the cost function and

constraints towards the desired direction. After performing 8 iteration into the optimization, we were able to reduce the delta setup time by 5.16% and $+3\sigma$ corner value of the setup time by 18.41%.

Finally, using the model, we studied the relationship of energy with delay for the circuit at different supply voltages. Figure 4 shows the spread of the energy and delay at the different operating voltages.

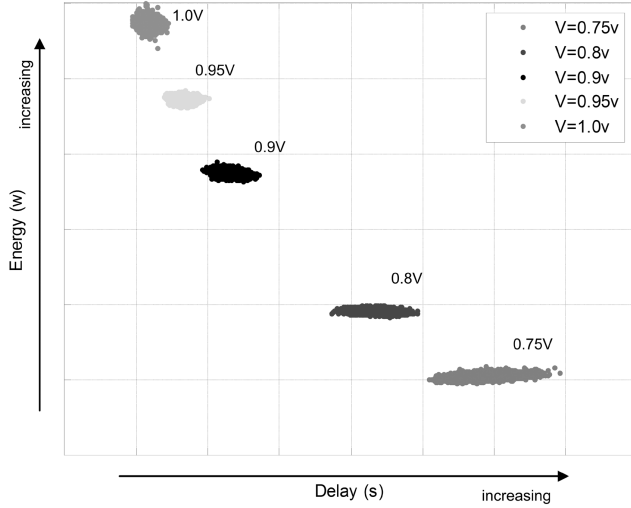


Fig. 4. Energy and delay variability with the scaled voltage.

The impact of increased variability at lower voltages and the widening distributions for different samples considering both global and local variations, shows that the change in relative variation (sigma/mean) has a point of operation (supply voltage) below which the relative variation increases significantly, as shown by the wider spreads in delay at the lower voltages.

An aspect of the design that is most drastically impacted by the increased variability is correct functionality of the logic at a lower voltage. Although, performance could be limited, appropriate management of the operating frequency by the power management unit (DVFS scheduler) will allow the system to run at a lower frequency than what would have expected thereby adhering to the setup constraint requirements of the design.

III. CONCLUSIONS

In this paper, we have demonstrated a statistical analysis and optimization flow on a range of applications. The low-rank quadratic response surface model generation technique has shown very good correlation with Monte Carlo both on linear and non-linear distributions. The most important aspect of the technique was the number of the simulations required to fit a model. Using quadratic response surface models enabled us to model interaction and higher order effects on the design, process, and environment variables. The overall flow is an extremely useful tool to both digital and analog design community trying to design more robust circuits in the presence of variations.

REFERENCES

- [1] S. R. Nassif, "Design for variability in DSM technologies," *ISQED Proceedings*, pp. 451-454, 2000.
- [2] A. Chandrakasan, W. J. Bowhill, F. Fox., *Design of High-Performance Microprocessor Circuits*. New York: Wiley, 2001.
- [3] A. Nardi, A. Neviani, E. Zanoni, M. Quarantelli, and C. Guardani, "Impact of unrealistic worst case modeling on the performance of VLSI circuits on deep submicron CMOS technologies," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 12, No. 4, pp. 396-402, Nov. 1999.
- [4] S. G. Duvall, "Statistical circuit modeling and optimization," *5th International Workshop on Statistical Metrology*, pp. 56-63, June 2000.
- [5] K. Singhal and V. Visvanathan, "Statistical device models from worst case files and electrical test data," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 12, No. 4, pp. 470-484, Nov. 1999.
- [6] J. A. Power, B. Donellan, A. Mathewson, and W. A. Lane, "Relating statistical mosfet model parameter variabilities to IC manufacturing process fluctuations enabling realistic worst case design," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 7, pp. 306-318, Aug. 1994.
- [7] S. R. Nassif, "Modeling and analysis of manufacturing variations," *IEEE CICC*, pp. 223-228, 2001.
- [8] X. Le, J. Le, L. T. Pileggi, A. Strojwas, "Projection-based performance modeling for inter/intra-die variations," *ICCAD Proceedings*, pp. 721-727, 2005.