# Accuracy Consideration of a Non-Gaussian Interconnect Delay Model for Submicron CMOS Statistical Static Timing Analysis

Amir Zjajo, Qin Tang, Michel Berkelaar, Nick van der Meijs

Circuits and Systems Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
amir.zjajo@ieee.org

**Abstract**

In submicron CMOS technology, due to the nonlinearity of the mapping from variation sources to the gate/wire delay, the distribution of the delay is no longer Gaussian. As the widening of process variability calls for accurate non-Gaussian timing models, their deployment requires well-controlled characterization techniques to cope with the complexity and scalability. In this paper, we present a corresponding analysis for the underlying interconnect timing model characterization infrastructure of statistical timing analysis. As the experimental results indicate, the non-Gaussian quadratic interconnect timing model is accurate within 1% error of the corresponding Monte Carlo simulation.

## Introduction

Gate delay and power dissipation are critical issues in present day low power VLSI circuit design. As we are moving towards nanometer technology, variations in process, voltage, and temperature are increasing, causing significant uncertainty in the delay estimation [1] and greatly impacting the yield [2]. As a consequence, various statistical static timing analysis (SSTA) algorithms [3-5] have been proposed to compute the statistical variations of timing performance due to the underlying process parameters. Deriving an efficient characterization methodology [6] and model order reduction techniques that can provide parameterized interconnects and facilitate efficient logic stage delay calculation is one of the critical tasks. Hence, in this paper[§], we propose a methodology for characterization of the quadratic timing model [3] that can capture large range non-Gaussian process variations, and based upon the adjusted parameter dimension reduction technique, we propose a timing accuracy verification flow for nonlinear logic gates and logic stage delays.

## Accuracy of Interconnect Delay Model

The most efficient way of quadratic model characterization is to adopt suitable parameter dimension reduction techniques that provide smart guidance for data sampling. The several model reduction methods for large sparse problems are related to Padé approximations of the underlying transfer function, or (directly) based on Krylov subspace techniques. In this paper, we adjust the dominant subspaces projection model reduction (DSPMR) [7], which is a numerically advantageous version of the balanced truncation technique [8]. A quadratic timing delay $\chi$ in reduced parameter space is expressed as

$$\chi(Z) = \chi_0 + \beta^T Z + Z^T H Z \tag{1}$$

where $Z$ is the reduced parameter set, and $\beta$ and $H$ are first and second order parameter dependencies, respectively. The resulting model in the full parameter space is given by

$$\chi(p) = \chi_0 + (\beta^T B)\Delta p + \Delta p^T (B^T H B)\Delta p \triangleq \chi_0 + \Delta\chi \tag{2}$$

where $\Delta\chi$ is the resulting circuit delay variation with $\Delta\chi=\chi(p)-\chi_0$, and $\Delta p$ is the parameter variation. Unlike in [7], to reduce the numerical cost of obtaining the projection matrix $B$, we approximate (up to machine precision) low rank Cholesky factors $C$. The diagonal matrix containing the singular values of reduced order $r$ has the same dimensions as the factored matrix

$$B = U_{(1:r)} \quad U\Sigma V^* = C \quad S = CU\Sigma^{-1/2} \tag{3}$$

In contrast to linear interconnects, where transfer function moments can be computed efficiently, the characterization of a nonlinear logic stage is more complex. In this paper, we propose an accuracy verification flow based upon the adjusted least squares approximation method. Denoting

$$\chi^k_{(i)}(p) = \chi_0 + \Delta\chi^k_{(i)} \tag{3}$$

the estimated delay vector for the $k^{th}$ gate at the $i^{th}$ iteration can be found by finding the solution for the transformation $\chi^k_{(i+1)}=F_i(\chi^k_{(i)})$ subject to

$$\left\|\chi^k_{(i+1)} - \chi^k_{(*)}\right\| < \left\|\chi^k_{(i)} - \chi^k_{(*)}\right\| \tag{4}$$

where

$$\chi^k_{(*)} = \arg\left\{\min_{\chi^k \in R^n} \varepsilon(\chi^k)\right\} \tag{5}$$

is the ideal solution of the delay for the error $\varepsilon$. We select an error mapping $F_i$ in the form of

$$\chi^k_{(i+1)}(p) = \chi^k_{(i)}(p) + d_i \Delta\chi^k_{(i)} \tag{6}$$

where $d_i$ is called error function and needs to be constructed. A quadratic function is selected in this paper to approximate the error function

$$d_t = \sum_{j=1}^{n} \gamma_j \Delta p_j + \sum_{j=1}^{n}\sum_{l=1}^{n} \gamma_{jl} \Delta p_j \Delta p_l, \quad t = 1, 2, ..., n \tag{7}$$

where $d=[d_1,d_2,\ldots,d_n]^T$, $\Delta p=[\Delta p_1, \Delta p_2,\ldots, \Delta p_n]^T$, $\gamma_j$ and $\gamma_{jl}$ are the coefficients of the error function at the $i^{th}$ iteration. The coefficients are determined by fitting the equation to the data set under the least square criterion. Once the error function is established, the performance function is executed as

$$\begin{aligned} \chi^k_{(i+1)}(p) &= \chi^k_{(i)}(p) + \Delta\chi^k_{(i+1)} \\ \Delta\chi^k_{(i+1)} &= \chi^k_{(i)}(p) + d_i \Delta\chi^k_{(i)} \end{aligned} \tag{8}$$

§ This research was sponsored by the European Union and the Dutch government as part of the ENIAC/MODERN project

## Experimental Results

The proposed method and all sparse techniques have been implemented in Matlab. All the experimental results are carried out on a PC with an Intel Core 2 Duo CPUs running at 2.66 GHz and with 3 GB of memory. To characterize the timing behavior, a lookup table-based library is employed which represents the gate delay and output transition time as a function of input arrival time, output capacitive load, and several independent random source of variation for each electrical parameter (i.e., $R$ and $C$). In each case, both driver and interconnect are included for the stage delay characterizations. The analytical delay distribution obtained using the quadratic interconnect model in a 45 nm CMOS technology is illustrated in Figure 1. The nominal value of the total resistance of the load and the total capacitance is chosen from the set 0.15kΩ-1kΩ and 0.4pF-1.4pF, respectively. The sensitivity of each given data point to the sources of variation is chosen randomly, while the total $\sigma$ variation for each data point is chosen in the range of 10% to 30% of their nominal value. The scaled distribution of the sources of variation is considered to have a skewness of 0.5, 0.75, and 1. When very accurate Gramians (e.g. low rank approximations to the solutions) are selected, the approximation error of the reduced system as illustrated in Figure 2 is very small compared to the Bode magnitude function of the original system. The lower two curves correspond to the highly accurate reduced system; the proposed model order reduction technique delivers a system of lower order, and the upper two denote fixed reduced orders. The transfer function of the system is denoted as $G$. The reduced order is chosen in dependence of the descending ordered singular values $\sigma_1,\sigma_2,\ldots \sigma_l$, where $l$ is the rank of factors, which approximate the system Gramians. For $n$ variation sources and $r$ reduced parameter sets, the full parameter model requires $O(n^2)$ simulation samples and thus has a $O(n^6)$ fitting cost. On the other hand, the proposed accuracy validation algorithm has a main computational cost attributable to the $O(n+r^2)$ simulations for sample data collection and $O(r^6)$ fitting cost for the quadratic model significantly reducing the required sample size and the fitting cost. Using 7000 Monte Carlo iterations (to guarantee a 99% confidence level with 0.5% accuracy) as a reference, the proposed algorithm evaluates the parameterized quadratic wire delay model with an accuracy within 1%, at the cost of at most 50 iterations (Figure 3), while achieving at least 16-fold *cpu*-time reduction. For each Monte Carlo sample over $\varepsilon$'s, the relative error is calculated as the difference between the delay result of the proposed approach and that of straightforward Monte Carlo simulation.

## Conclusion

This paper presents a highly efficient methodology for quadratic timing model timing accuracy verification flow of nonlinear logic gates and logic stage delays. By adopting parameter dimension reduction techniques and an accurate model verification flow, timing model extraction can be performed in a reduced parameter space, thus providing a significant reduction on the required number of simulation samples to construct accurate quadratic timing models. Extensive experiments are conducted on a large set of random test cases, showing very accurate results.

## References


[1] C. Forzan, D. Pandini, "Statistical static timing analysis: A survey," *Integration*, vol. 42, no. 3, pp. 409-435, 2009

[2] S.R. Nassif, "Modeling and analysis of manufacturing variations," *CICC*, pp. 223-228, 2001

[3] L. Zhang, W. Chen, Y. Hu, A. Gubner, C. Chen, "Correlation-preserved non-gaussian statistical timing analysis with quadratic timing model," *DAC*, pp. 83-88, 2005

[4] V. Veetil, D. Sylvester, D. Blaauw, "Efficient Monte Carlo based incremental statistical timing analysis," *DAC*, pp. 676-681, 2008

[5] Q. Tang, A. Zjajo, M. Berkelaar, N. van der Meijs, "RDE-based transistor-level gate simulation for statistical static timing analysis," *DAC*, pp. 787-792, 2010

[6] Y. Bi, K.-J. van der Kolk, J.F. Villena, L.M. Silveira, N. van der Meijs, "Fast statistical Analysis of RC nets subject to manufacturing variabilities," *DATE*, 2011, in press.

[7] J. Li, J. White, "Efficient model reduction of interconnect via approximate system Grammians," *ICCAD*, pp. 380-384, 1999

[8] B. C. Moore, "Principal component analysis in linear systems: controllability, observability, and model reduction," *Trans. Automat. Control*, vol. 26, pp. 17-31, 1981


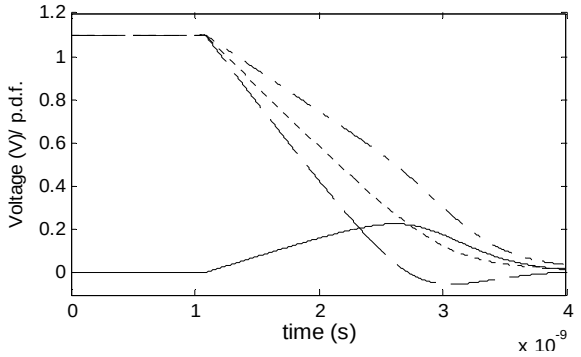


Figure 1: Analytical delay distribution in 45 nm CMOS technology. Solid line illustrates non-Gaussian distribution of the delay.

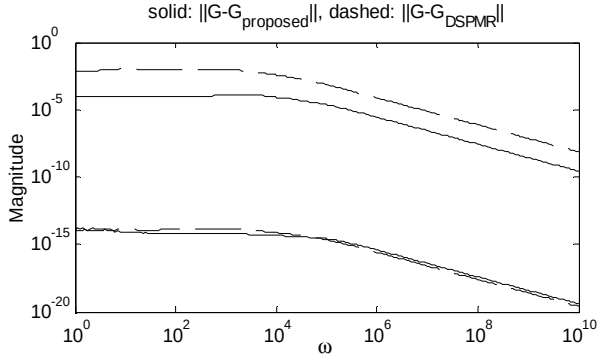


Figure 2: The Bode magnitude plot of the approximation errors.

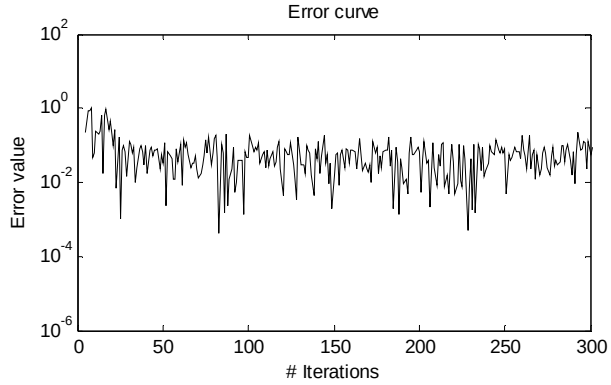


Figure 3: The least squares error for a 300 iterations.