Computationally Efficient Characterization of Standard Cells for Statistical Static Timing Analysis

by

Sharon H. Chou

Bachelor of Science in Electrical Engineering and Computer Science,

Massachusetts Institute of Technology

June, 2008

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

May, 2009

Author_____
Department of Electrical Engineering and Computer Science
May 22, 2009

Certified by_____
Alice Wang
Senior Member of Technical Staff, Texas Instruments, Inc.
VI-A Company Thesis Supervisor

Certified by_____
Dennis Buss
Chief Scientist, Texas Instruments, Inc.
VI-A Company Thesis Co-Supervisor

Certified by_____
Anantha Chandrakasan
Joseph F. and Nancy P. Keithley Professor of Electrical Engineering
M.I.T. Thesis Supervisor

Accepted by_____
Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses

Computationally Efficient Characterization of Standard Cells for Statistical Static Timing Analysis

by

Sharon H. Chou

Submitted to the
Department of Electrical Engineering and Computer Science

May 22, 2009

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

We propose a computationally efficient statistical static timing analysis (SSTA) technique that addresses intra-die variations at near-threshold to sub-threshold supply voltage, simulated on a scaled 32nm CMOS standard cell library. This technique would characterize the propagation delay and output slew of an individual cell for subsequent timing path analyses. Its efficiency stems from the fact that it only needs to find the delay or output slew in the vicinity of the ξ-sigma operating point (where ξ = 0 to 3) rather than the entire probability density function of the delay or output slew, as in conventional Monte-Carlo simulations. The algorithm is simulated on combinational logic gates that include inverters, NANDs, and NORs of different sizes. The delay and output slew estimates in most cases differ from the Monte-Carlo results by less than 5%. Higher supply voltage, larger transistor widths, and slower input slews tend to improve delay and output slew estimates. Transistor stacking is found to be the only major source of under-prediction by the SSTA technique. Overall, the cell characterization approach has a substantial computational advantage compared to SPICE-based Monte-Carlo analysis.

VI-A Company Thesis Supervisor: Alice Wang
Title: Senior Member of Technical Staff, Texas Instruments, Inc.

VI-A Company Thesis Co-Supervisor: Dennis Buss
Title: Chief Scientist, Texas Instruments, Inc.

MIT Thesis Supervisor: Anantha Chandrakasan
Title: Joseph F. and Nancy P. Keithley Professor of Electrical Engineering
Director, MIT Microsystems Technology Laboratory

Acknowledgment

My sincere gratitude goes to my thesis supervisors – Dr. Dennis Buss, Dr. Alice Wang, and Prof. Anantha Chandrakasan. I have learned a great deal from Dennis, who has been a constant source of inspiration for me, professionally and personally. Alice's timely reminders and insightful questions have been immensely helpful during the hardest phase of my thesis writing. I am especially grateful for the continual support from Anantha throughout my thesis research.

I would like to thank my colleagues at Texas Instruments – Dr. Jie Gu and Satyendra Datla for collaborating with my work and answering my numerous technical questions. They have helped make my relocation back to MIT glitch-free as well. I also thank Dr. Uming Ko and Dr. Gordon Gammie who helped make my Co-op term at TI's Wireless Terminal Business Unit possible.

My summer Co-op term would not be the same without current and former members of the Anantha-group who were/are working at TI – Mahmut Ersin Sinangil, Dr. Nathan Ickes, Joyce Kwong, Vivienne Sze, Dr. Brian Ginsburg, Dr. Raúl Blasquez-Fernandez, and of course Alice. Our group lunches, dinners and movies were lots of fun!

This thesis would not have been possible without funding from the VI-A program sponsored by the EECS Dept. at MIT and Texas Instruments, Inc.

Finally, I would like to thank my family for everything they have done to get me where I am now.

Contents

List of Figures

List of Tables

1. Introduction

Statistical process variations have long been an important design issue. But until recently, process variations have been assumed to be "global" in traditional static timing analysis (STA), i.e., transistor parameters may vary from die to die but are assumed to be constant within a die [1, 2]. With transistor geometries shrinking below 65nm, however, a new kind of statistical variation has become important for logic [3]. It is no longer valid to assume that transistor parameters are constant across a die because there are "local" or intra-die variations [1, 2]. Local variations have long been an important issue in SRAM design [4], but it is only recently that they have become significant enough to affect logic [3]. This has necessitated the development of new SSTA design techniques to address local variations.

Several physical mechanisms for local variations have been proposed [5], but for most process technologies, the predominant mechanism is random dopant fluctuations (RDF), which is a random variation in the number of dopant atoms in the channel of each transistor. On this basis, the transistor random variables are assumed to be independent, Gaussian random variables [6]. Furthermore, at nominal voltage, it is accurate to assume that circuit performance (propagation delay) is linear in transistor variation [7]. In this case, the circuit delay is Gaussian, and the standard deviation can be readily calculated from the standard deviations of the transistor parameters.

However, at low V_{DD} (near-threshold and sub-threshold), circuit delay is a nonlinear function of transistor random variables. This leads to a probability density function (PDF) for delay that is non-Gaussian. At sub-threshold V_{DD}, the delay exhibits an exponential dependence on variation of the threshold voltage (V_T), derived from the lognormal distribution that occurs frequently in analysis of sub-threshold circuit variability (Eq. 1). The current I_D is equal to I_0 when the gate-to-source voltage (V_{GS}) is equal to V_T without body effect (V_{To}), when the source-to-body voltage (V_{SB}) is zero. The body effect is captured by V_{SB} as well as the body-effect coefficient γ. There is an additional dependency on drain-induced barrier-lowering (DIBL) as

reflected in the drain-to-source voltage (V_{DS}) and the DIBL coefficient η. The factor *n* is the sub-threshold slope ideality parameter as defined by Eq. 2, and V_{th} is the thermal voltage [8].

$$I_D = I_o e^{\frac{V_{GS} - V_{To} - (\gamma V_{SB}) + \eta V_{DS}}{n V_{th}}} \left(1 - e^{\frac{-V_{DS}}{V_{th}}}\right) \qquad (1)$$

$$n = \frac{\Delta V_{GS}}{V_{th} \Delta \log(I_D) \ln(10)} \qquad (2)$$

The PDF of the delay at sub-threshold V_{DD} is a result of the exponential dependence of current on V_T and the assumption that V_T is normally distributed from local process variation. The lognormal distribution is asymmetric with a long tail on the right, which implies that below average circuit delays deviate only slightly from the mean, while above average delays can be several times the nominal value. This nonlinearity greatly complicates the statistical analysis because the circuit delay is no longer Gaussian.

At near-threshold V_{DD} but greater than threshold, the delay behavior transitions between an exponential dependence and a linear dependence. This region is highly nonlinear as well, and it is an important region of operation because low-power CMOS processes frequently operate in this region. To address the nonlinearity, we introduce an innovative approach to SSTA which can be used to characterize logic gates (cells) and perform timing path analysis in circuits. This thesis focuses the cell characterization step.

2. Previous and current SSTA procedures

Several approaches for SSTA have been proposed, ranging from numerical integration techniques [9] to Monte-Carlo based techniques [10, 11] to those based on probabilistic analyses. Though the methods based on numerical integration and Monte-Carlo techniques can provide a high level of accuracy, practical use of these methods becomes prohibitive because of the very high computational costs involved in these methods. This has led to the majority of research on SSTA being focused on probabilistic analysis-based procedures [12-20].

However, many of the approaches consider the delay PDF to be Gaussian and the delay to be a linear function of the variation sources [13, 17, 19]. With continuous scaling of technology and desired ultra-low voltage operation of the circuits for low-power applications, this assumption can no longer be justified. More recently, attempts have been made to address this issue of nonlinear delay variations resulting in non-Gaussian PDFs. Most of these methods rely on Taylor series expansion based polynomial representations to model the cell and timing path delays. The major drawback in such representations is that of high computational complexity in performing the MAX operation. An algorithm capable of handing non-Gaussian parameter distributions using independent component analysis and principal component analysis is proposed in [18], though it considers the delay to be linear with respect to the random variables. Quadratic delay model-based algorithms are proposed in [16, 20]. MAX is computed using a conditional linear approximation in [16] while considering the inputs to be Gaussian. The MAX operation in [20] is computed by a moment-matching technique where the moments are computed by computationally expensive numerical integration. A Taylor-series expansion-based polynomial delay model is proposed in [14]. It uses regression-based polynomial modeling of the MAX operation. Parameterized block-based SSTA approach is extended in [12] to account for nonlinear delays, where numerical integration is used to compute the MAX. A modified quadratic delay model is proposed in [15]. A moment-matching technique is applied for the computation of MAX while using a Fourier series-based approach to calculate the moments.

Two different basic approaches are to compute the delay distribution in path-based or block-based manner [3]. In path-based methods, it has been proposed to run a traditional STA first, and then analyze only the n most critical paths accurately using SSTA due to the high computational effort. The risk is that the statistically most critical path could be missed. Block-based approaches suffer from a lack of accuracy especially for the MAX/MIN operation.

As an example of an implementation for the path-based approach, the PrimeTime VX software provides a timing scheme that accounts for the probabilistic variations of the parameters. It takes as inputs the probability density functions of process parameters (e.g. oxide thickness, dopant level, flatband voltage) along with operating voltage and temperature (Figure 1). In this manner, PrimeTime VX considers the process variations in modeling the delay. However, it does not account for local variations and thus does not accurately model the delay behavior at low V_{DD}, where the delay probability density functions (PDF) are more lognormal than Gaussian. Additionally, since the worst-case delays tend to occur within a limited region of parameter variations, including the entire PDF would unnecessarily increase computation time.

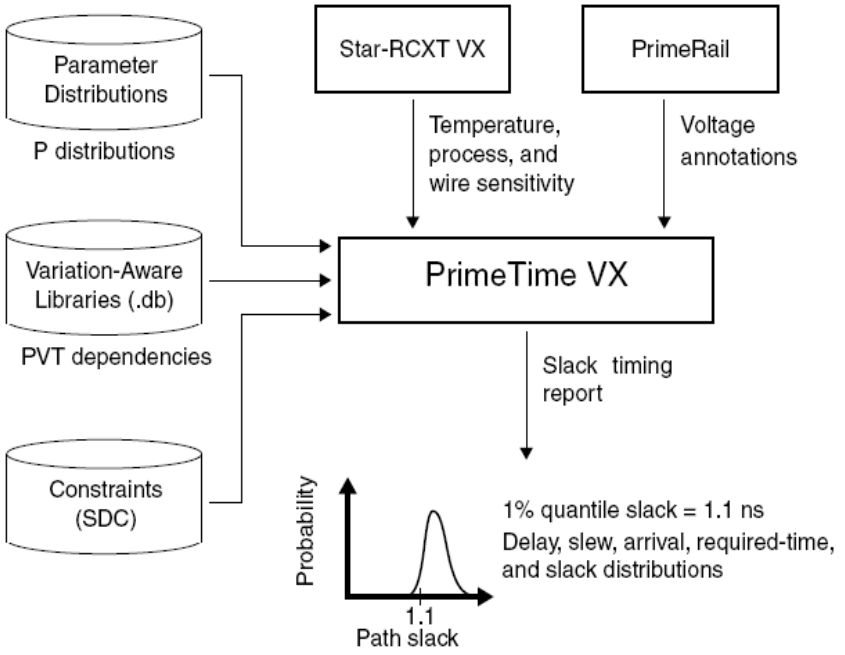


Figure 1. PrimeTime VX simulation flow

3. Theory of proposed SSTA algorithm

We now present a computationally efficient technique for performing SSTA in the regime where circuit performance is highly nonlinear in the space of transistor random variables. This technique is performed on a standard cell library implemented in a scaled 32nm CMOS technology. The algorithm is divided into two different SSTA tasks:

1. Statistical cell characterization
2. Statistical timing path analysis

This thesis focuses on the cell characterization step, but it is instructive to examine the theory for both steps simultaneously. To understand the principle of the SSTA technique, consider a timing path (TP) consisting of individual cells. The delay of the TP is the sum of the delays of the individual cells. The computational efficiency of the approach arises from the fact that, in most cases, we do not need to know the entire probability density function (PDF) of the TP delay. For example, we are usually only interested in the "3-sigma" (or "ξ-sigma") TP delay. As a result, we only need to know the PDF of the TP delay in the vicinity of 3-sigma. Since the TP delay is the sum of individual cell delays, the PDF of the TP delay is the convolution of the PDFs of cell delays. Furthermore, for PDFs of interest, this convolution integrand is significant only in a small region of the cell delay space. The integrand is a maximum at a point in space where the joint probability density of cell delays is a maximum; this point is defined as the "operating point". The computational efficiency results from restricting computations to calculating only the 3-sigma (or ξ-sigma) operating point for a TP.

Cell characterization is performed in an analogous manner as TP analysis. The goal of cell characterization is to determine the PDF of the stochastic cell delay D for each arc of each cell (each cell/arc) in the library, as well as the most probable output slew associated with every value of cell delay. An arc is defined by the input trigger edge (either rising or falling), input slew rate, and output capacitive load. Each cell/arc has an associated stochastic delay and

stochastic output slew, which both result from random variations in transistor parameters relative to their nominal values at the global (weak) corner. Let the PDF of the delay be $P_D(D)$. For the $P_D(D)$ of each cell/arc, we can define a cell/arc delay function (CADF) $D(\xi)$ that uniquely defines $P_D(D)$ and maps it onto a zero-mean, normalized ($\sigma = 1$) Gaussian parameter ξ. Figures 2 and 3 illustrate two cases – a linear delay function at nominal V_{DD} (Figure 2) and a nonlinear delay function at low V_{DD} (Figure 3). The CADF $D(\xi)$ uniquely defines $P_D(D)$, and the corresponding cell/arc slew function (CASF) $S(\xi)$ uniquely defines the output slew at each value of the cell delay. The CASFs have the same general morphology as their CADF counterparts – either linear at nominal V_{DD} or nonlinear at low V_{DD}. The functions $D(\xi)$ and $S(\xi)$ are the outputs of cell characterization.

In cell characterization, we typically characterize $D(\xi)$ and $S(\xi)$ over the range $0 < \xi < 3$ for subsequent TP analysis. For computational efficiency, it is necessary to characterize $D(\xi)$ and $S(\xi)$ as piece-wise linear curves with a limited number of linear segments. In this work, we do not explore the trade-off between accuracy and number of segments. We have thus characterized $D(\xi)$ and $S(\xi)$ with a resolution of 0.25σ in order to determine the inherent accuracy of this approach.

Variations in transistor parameters are specified in a SPICE model as independent zero-mean Gaussians, where there are n_v variables per transistor. In this study, $n_v = 2$, though theoretically n_v can be any arbitrary number depending on the transistor model. For a cell consisting of n_t transistors, there are $N = n_v n_t$ transistor random variables which we designate as x_i for $i = 1, 2 \ldots N$. Let us now consider the case where $D(\xi)$ is a linear function of transistor random variables.

3.1. Cell characterization in the linear case

At V_{DD} near nominal and for sufficiently small σ_i, the CADF is essentially linear (Figure 2). Let D be the stochastic delay and x_i be a transistor random variable represented by a zero-mean Gaussian. Because D is linearly related to x_i, it can be expressed as:

$$D = \sum_{i=1}^{N} \frac{dD}{dx_i} x_i = \sum_{i=1}^{N} \frac{dD}{d\zeta_i} \zeta_i \quad \text{, where} \quad \zeta_i = \frac{x_i}{\sigma_i}.$$

(3)

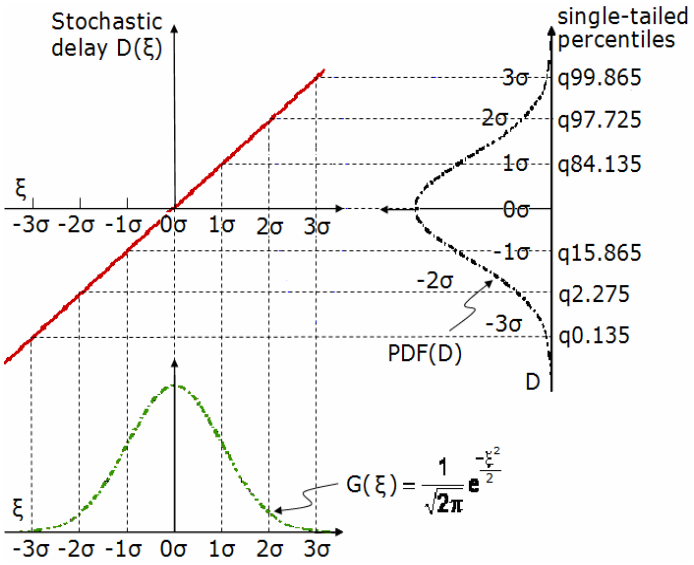


Figure 2. Gaussian $P_D(D)$ when $D(\xi)$ is linear at nominal V_{DD} (e.g. 0.9V)

We can define a weighting variable α_i as:

$$\alpha_i = \frac{dD}{d\zeta_i}$$

(4)

Hence the delay can also be expressed as:

$$D = \sum_{i=1}^{N} \alpha_i \zeta_i$$

(5)

N values of α_i are computed from 2N SPICE simulations as:

$$\alpha_i = \frac{D(\zeta_i + \Delta\zeta_i) - D(\zeta_i - \Delta\zeta_i)}{2\Delta\zeta_i}$$

(6)

Because x_i and hence ζ_i are statistically independent, the variance of delay D can be directly written as:

$$Var(D) = \sigma_D^2 = \sum_{i=1}^{N} \alpha_i^2$$

(7)

The resulting stochastic delay is thus a Gaussian random variable with zero mean and standard deviation σ_D. In the linear case, the CADF is $D(\xi) = \sigma_D \xi$ as shown in Figure 2, and the CASF is $S(\xi) = \sigma_S \xi$. Each cell/arc is completely characterized by the values σ_D and σ_S.

3.2. Cell characterization in the nonlinear case

This section describes the theory for the novel SSTA method that applies to the region of low-V_{DD} operation. For V_{DD} in the near-threshold or sub-threshold range, the stochastic delay is no longer a linear function of the transistor random variables. It is instead nonlinear in the transistor random variables x_i. As a result, the PDF for cell delay is non-Gaussian (Figure 3).

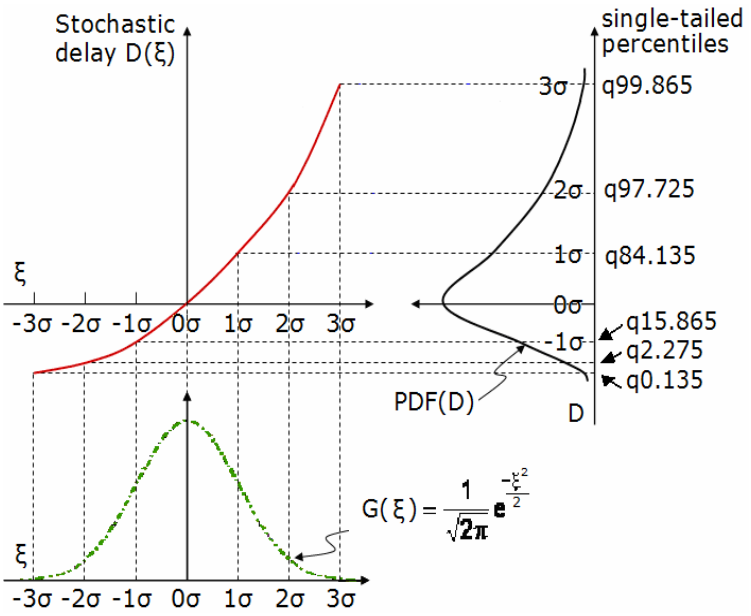


Figure 3. Non-Gaussian $P_D(D)$ when $D(\xi)$ is nonlinear at near/sub-threshold V_{DD} (e.g. 0.5V)

To obtain the nonlinear CADF and CASF for each cell/arc, the first step is to compute the sensitivity curves for each of the N random variables in each of the cell/arcs. Figure 4 shows typical sensitivity curves with respect to four different transistor random variables in an inverter. X1 and X2 correspond to PMOS variables while X3 and X4 correspond to NMOS variables.

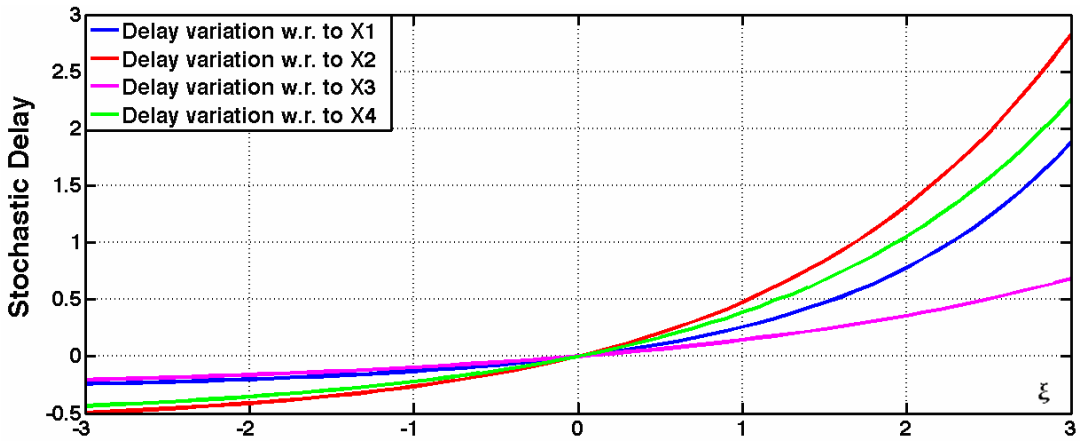


Figure 4. Typical sensitivity curves of an inverter at V_{DD} = 0.5V

We now introduce the concept of an *operating point* in ζ_i space. There is a different operating point for each value of ξ, and we refer to it as the ξ-sigma operating point. For each arc/cell, we would evaluate $D(\xi)$ at a select number of ξ values. For each value of ξ, there is an operating point in ζ_i space, which is the maximum of the joint PDF for the ζ_i that satisfy the condition $D = D_{\xi\sigma}$. In other words, the operating point is the point in ζ_i space where the function $D(\zeta_1, \zeta_2, \ldots \zeta_N) = D_{\xi\sigma}$ is tangent to the hyper-sphere of radius ξ and the value of $D_{\xi\sigma}$ is determined from the curve that is tangent at the operating point (Figure 5).

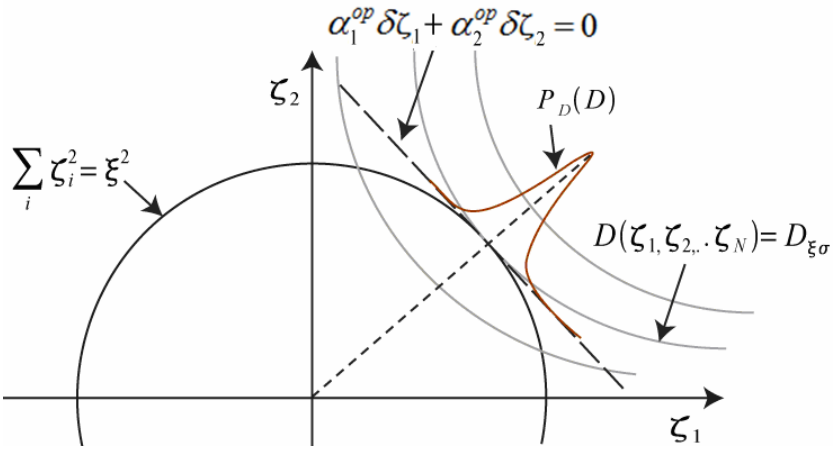


Figure 5. Cell operating point

The fundamental principle of this analysis is that although $D(\zeta_1, \zeta_2, \ldots \zeta_N)$ is a nonlinear function, it can be approximately linearized about any point in ζ_i space. In particular, it can be linearized about the operating point. We define ζ_i^{op} as the ξ-sigma operating point and define $\delta\zeta_i = \zeta_i - \zeta_i^{op}$ as the incremental variations in ζ_i about this operating point. Then we can express the total cell delay as:

$$D(\zeta_1, \zeta_2 \ldots \zeta_N) \approx D_{3\sigma} + \sum_{i=1}^{N} \left(\frac{dD}{d\zeta_i} \right)_{op} \delta\zeta_i = D_{3\sigma} + \sum_{i=1}^{N} \alpha_i^{op} \delta\zeta_i \tag{8}$$

Moreover, since the delay can be approximated as a linear function of the transistor random variables in the vicinity of the operating point, the delay PDF is approximately the convolution of the PDFs of the individual ζ_i. The integrand of the convolution integral peaks at the operating point and falls off sharply in all directions (Figure 5). As a result, in the region of ζ_i

space that makes the largest contribution to the integral, the linear approximation is valid. The ξ-sigma operating point can thus be determined as the point of tangency of the hyper-plane denoted as

$$\sum_{i=1}^{N} \alpha_i^{op} \delta\zeta_i = 0$$

(9)

with the hyper-sphere

$$\sum_{i=1}^{N} \zeta_i^2 = \xi^2$$

(10)

The operating point is determined by (11) which can be solved in an iterative manner.

$$\zeta_i^{op} = \frac{\xi \alpha_i^{op}}{\sqrt{\sum_{j=1}^{N} (\alpha_j^{op})^2}}$$

(11)

For each value of ξ, once the operating point is determined, the delay $D(\xi)$ and output slew $S(\xi)$ are simulated each by a single SPICE run at the operating point.

4. Cell characterization experimental setup

The cell-level SSTA algorithm described in Section 3 is implemented to simulate the ξ-sigma delays and output slews in the range $0.25 \le \xi \le 3$. The simulations are run at V_{DD} = 0.9V (nominal) and V_{DD} = 0.5V (low) for each cell/arc of variously-sized inverters, NAND gates, and NOR gates. The sensitivity functions (delay as a function of ζ_i) are linearized at $\zeta_i = 0$ by taking the slope of the line from the delay at $-\Delta\zeta_i = -0.1$ and $+\Delta\zeta_i = 0.1$ (Figure 6), with one SPICE simulation run at each point.

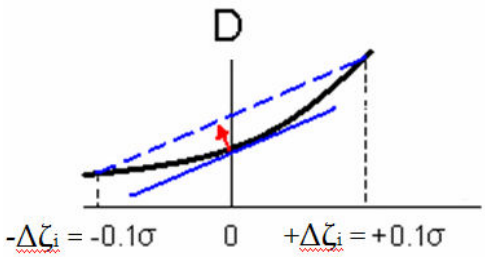


Figure 6. Linearization of the sensitivity function

The slope α_i is equal to the weighting factor as defined in Section 3.1, denoted as below, where ζ_i is a transistor variable and $\Delta\zeta_i = 0.2$.

$$\alpha_i = \frac{D(\zeta_i + \Delta\zeta_i) - D(\zeta_i - \Delta\zeta_i)}{2\Delta\zeta_i}$$, repeated from (6)

Each transistor variable has a ξ-sigma operating point ζ_i^{op} that is calculated as below:

$$\zeta_i^{op} = \frac{\xi\alpha_i^{op}}{\sqrt{\sum_{j=1}^{N}(\alpha_j^{op})^2}}$$, repeated from (11)

It has been experimentally determined that calculating Eq. 11 iteratively does not significantly improve the error estimation. We therefore compute Eq. 11 in a single pass with α_i evaluated at $\zeta_i = 0$.

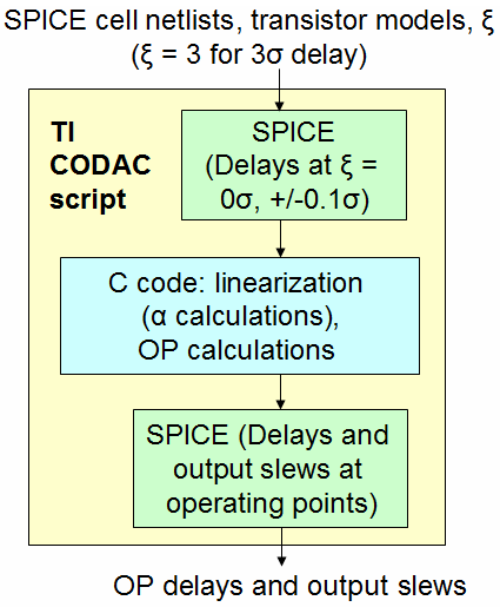


Figure 7. Cell characterization simulation flow

A single SPICE run is required to obtain the total delay $D_i(\xi)$ and output slew $S_i(\xi)$ at the operating point designated by the value of ξ. Different combinational CMOS cells are simulated under an array of input slews and output capacitance loads (Table 1). The propagation delays are measured from the time when the input waveform rises or falls to 50% of V_{DD} to the time when the output waveform falls or rises to 50% of V_{DD}, respectively. The input/output slew duration (ISD/OSD) is measured as the time interval between when the input/output waveform rises from 25% of V_{DD} to 75% V_{DD}, or falls from 75% of V_{DD} to 25% of V_{DD}. Table 1 lists the setup conditions for a cell with strength of 1x, which is determined by transistor size.

Table 1. Load/input slew conditions for all tested cells

V_{DD} = 0.9V	(Gate strength = 1)	ISD = 10ps	ISD = 100ps	ISD = 400ps
(nominal)	Load = 1 fF	0	1	2
	Load = 5 fF	3	4	5
	Load = 25 fF	6	7	8
V_{DD} = 0.5V	(Gate strength = 1)	ISD = 125ps	ISD = 1250ps	ISD = 5000ps
(near/sub-V_T)	Load = 1 fF	0	1	2
	Load = 5 fF	3	4	5
	Load = 25 fF	6	7	8

The output loads are scaled according to gate strength. The cells tested are listed in Table 2. The number before "x" measures gate strength – 1x indicates minimum-width transistors and *c*x indicates *c* times the minimum width, where *c* = number of fingers in each transistor.

Table 2. Listing of the tested cells by topology and size

	Inverters	2-input NAND	3-input NAND	2-input NOR	3-input NOR
Small	INVT_1x	NAND2_1x	NAND3_2x	NOR2_1x	NOR3_2x
Medium	INVT_5x	NAND2_2x	NAND3_4x	NOR2_2x	NOR3_4x
Large	INVT_10x	NAND2_4x	NAND3_8x	NOR2_4x	NOR3_8x

In any given run, the input waveform is connected to either Port A, B, or C, while the other input ports are either tied to V_{DD} in NANDs or tied to V_{SS} in NORs to enable output switching (Figure 8).

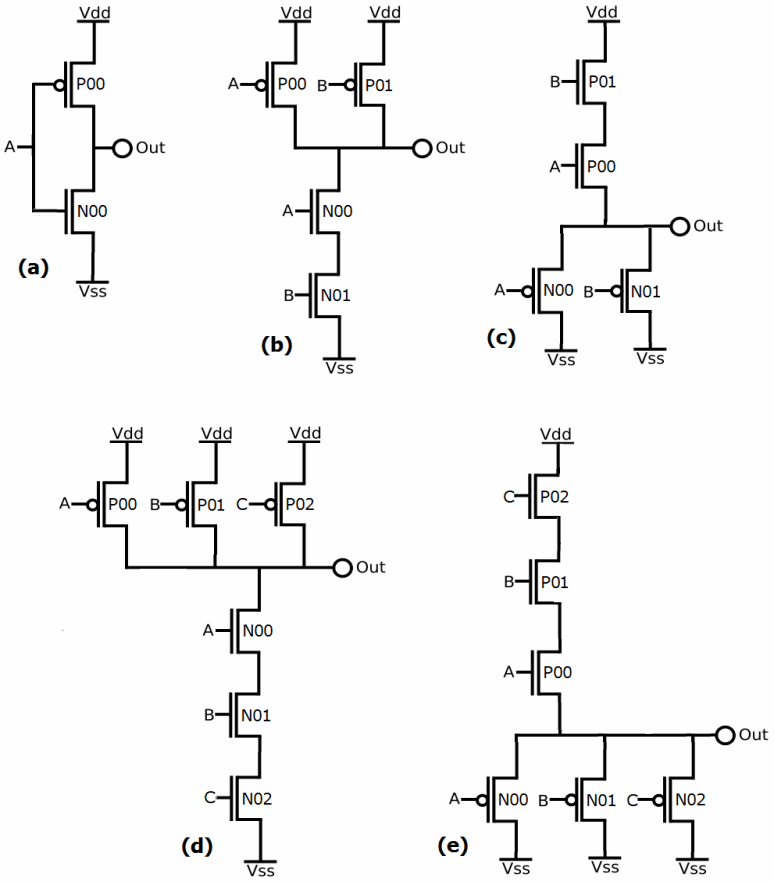


Figure 8. Schematics and port labels in inverters (a), NAND2 (b), NOR2 (c), NAND3 (d), and NOR3 (e)

5. Cell-level characterization results

Typical PDFs of the total cell delay and output slew for an inverter are shown for V_{DD} = 0.9V (Figure 9) and V_{DD} = 0.5V (Figure 10). The PDFs for NAND gates and NOR gates have very similar distributions. The corresponding stochastic CADF and CASF are shown in Figures 11 and 12. SPICE-based Monte-Carlo (MC) analysis with 10000 samples serves as a comparison.

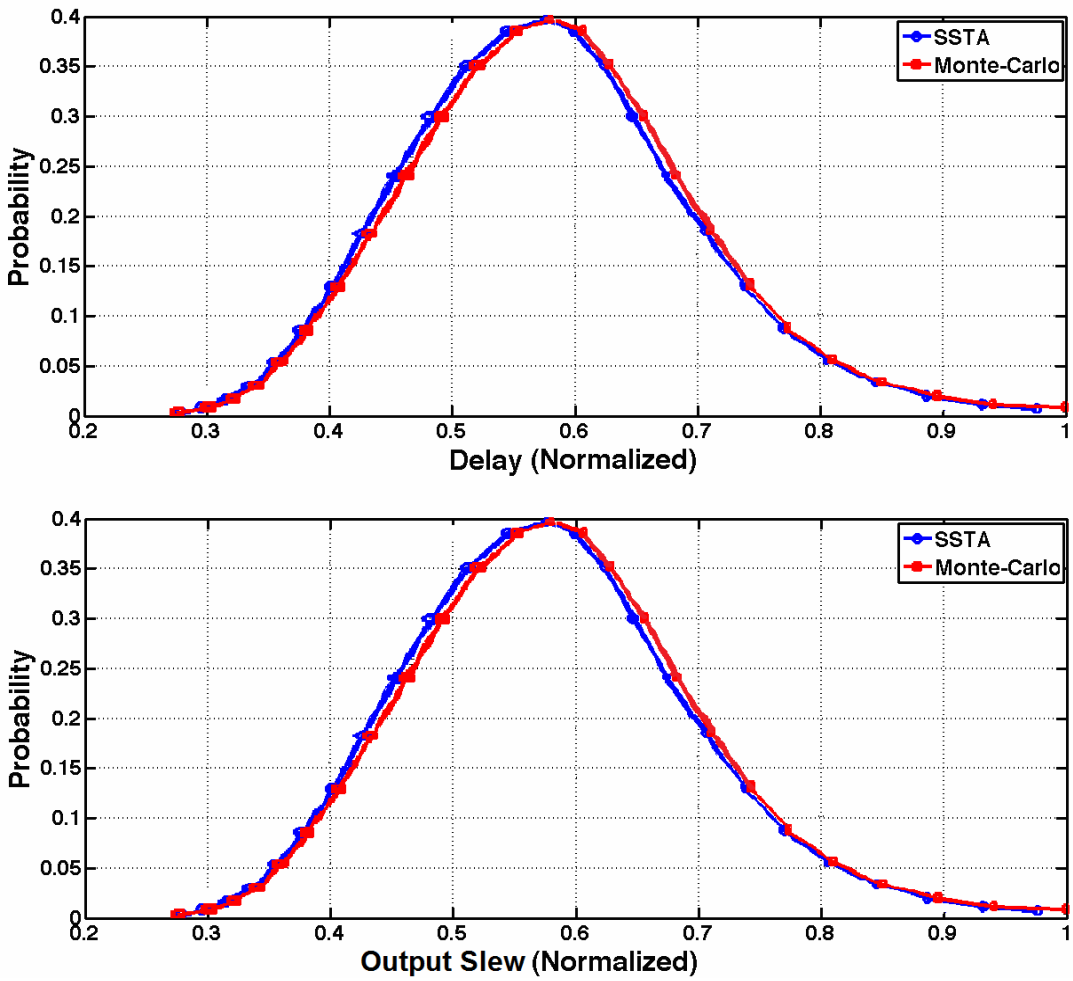


Figure 9. Typical delay and output slew PDF at V_{DD} = 0.9V

The PDFs for V_{DD} = 0.9V are approximately Gaussian as described in Section 3.1.

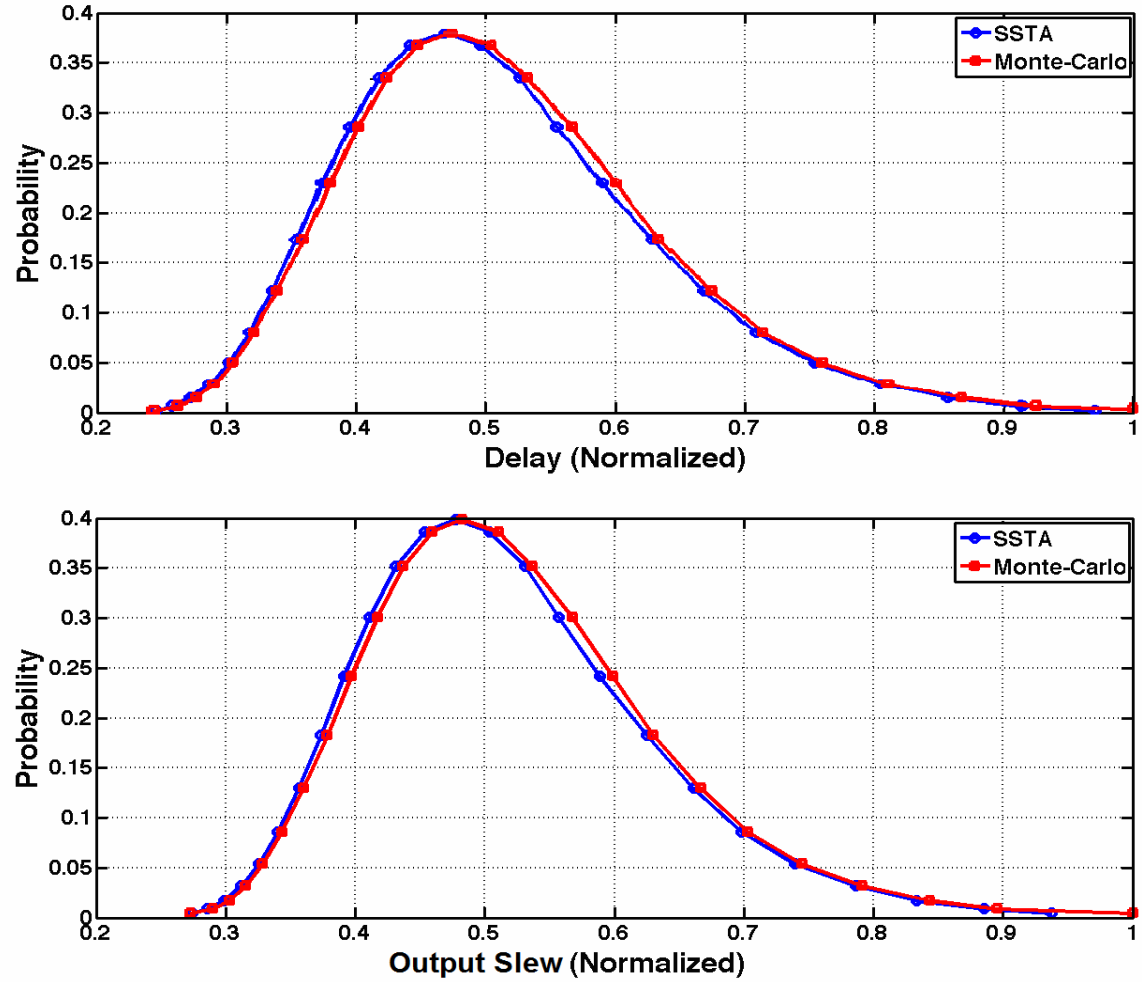


Figure 10. Typical delay and output slew PDF at V_{DD} = 0.5V

The PDFs in Figure 10 are skewed to the right as described in Section 3.2.

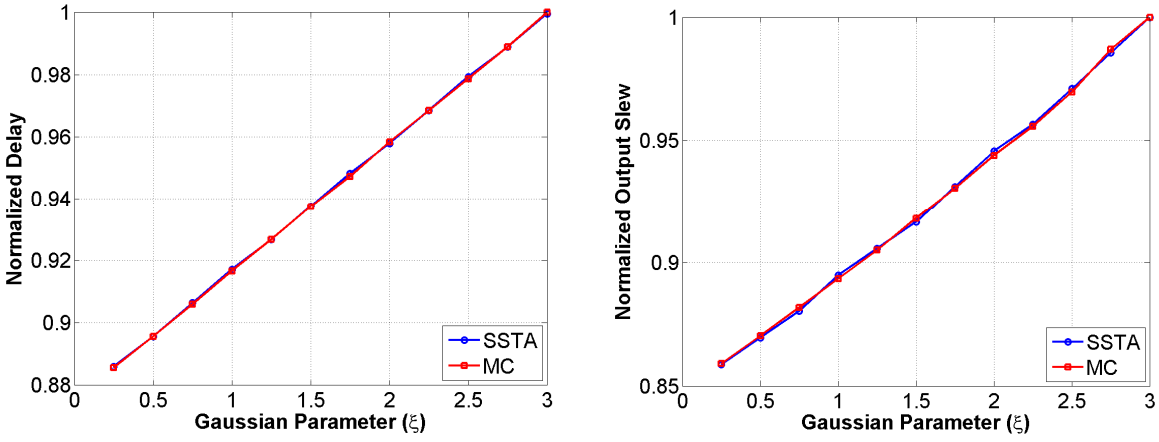


Figure 11. Typical CADF and CASF for a cell at V_{DD} = 0.9V

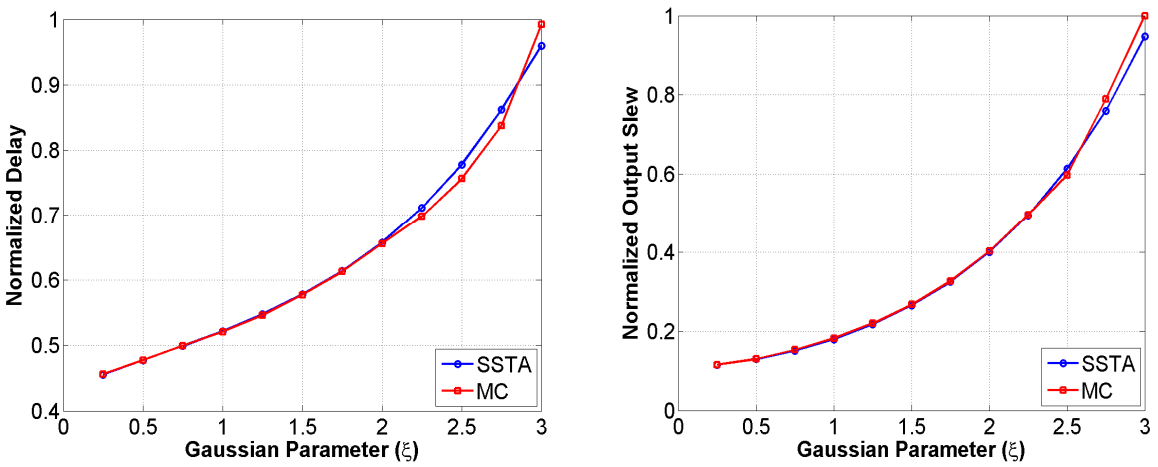


Figure 12. Typical CADF and CASF for a cell at V_{DD} = 0.5V

The CADFs and CASFs at V_{DD} = 0.9V are approximately linear (Figure 11) while those at V_{DD} = 0.5V are highly nonlinear (Figure 12). In the CADFs and CASFs of all cells, there is very close agreement between the SSTA and Monte-Carlo $\sigma_{D,i}(\xi)$ when $\xi < 2$. For some cells at V_{DD} = 0.5V, the SSTA delay and slew estimates deviate from those in MC analysis when $\xi > 2$. This could be attributed to the relatively sparse data samples at the tail end of the delay and slew PDFs, where $\xi = 2$ is located at the 97.725th percentile and $\xi = 3$ is located at the 99.865th percentile (227th and 13th largest data point from the MC distribution, respectively). The data shown in the subsequent sections is entirely from the results of sweeping the input waveform at Port A (Figure 8), which is representative of all data since the accuracy of the algorithm is found not significantly affected by the input ports.

5.1. Delay characterization results and discussion

The percentage errors in delay estimations are plotted for inverters (Figures 13a to 13d), NANDs (Figures 14a to 14d), and NORs (Figures 15a to 15d) across β ratio from 1.0 to 2.5 (ratio between PMOS width and NMOS width). Percentage errors are calculated at $\xi = 3$ as follows:

$$\text{Percent (\%) error} = (\text{SSTA delay} - \text{MC delay}) / \text{MC delay} \quad (12)$$

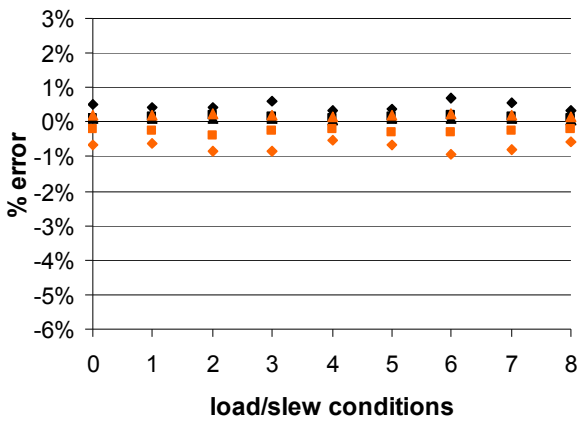


Figure 13a. Input-rise arc at V_{DD} = 0.9V

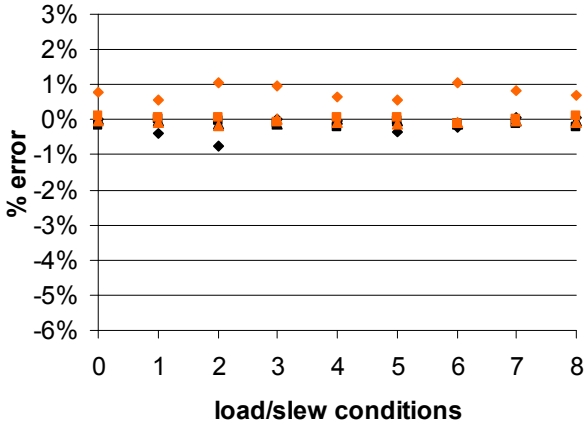
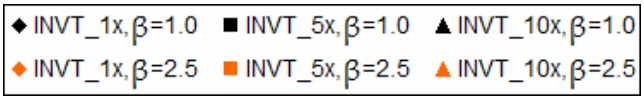


Figure 13b. Input-fall arc at V_{DD} = 0.9V



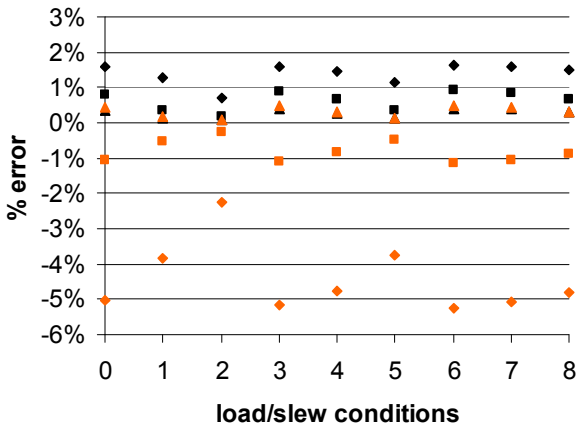


Figure 13c. Input-rise arc at V_{DD} = 0.5V

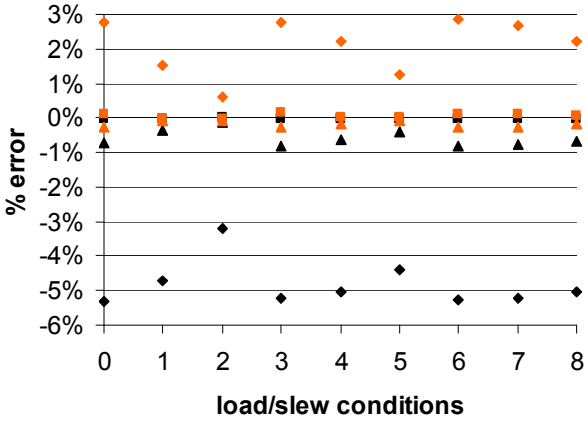


Figure 13d. Input-fall arc at V_{DD} = 0.5V

All inverters have errors within 2% for both the input-rise and input-fall arcs, except INVT_1x at V_{DD} = 0.5V at input-fall. A slower slew rate tends to improve accuracy, whereas output load does not significantly affect the percentage error estimates (Figures 13a-d).

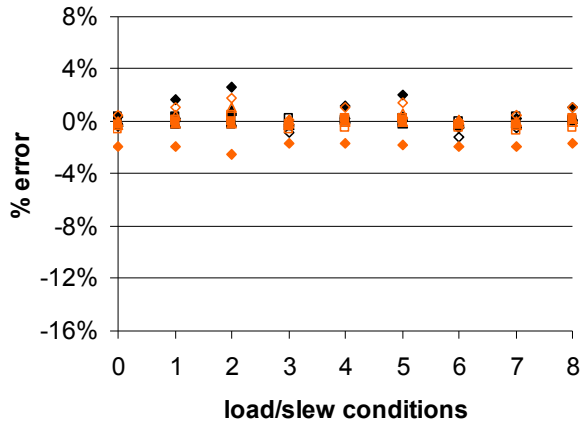


Figure 14a. Input-rise arc at V_{DD} = 0.9V

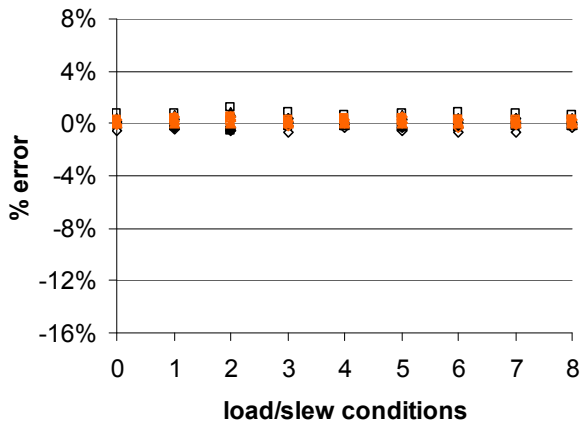
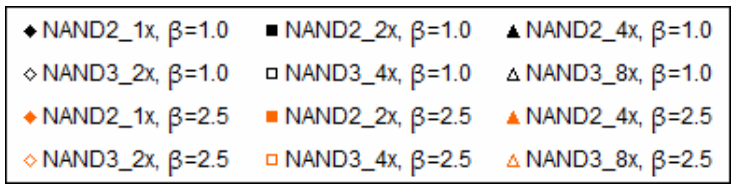


Figure 14b. Input-fall arc at V_{DD} = 0.9V



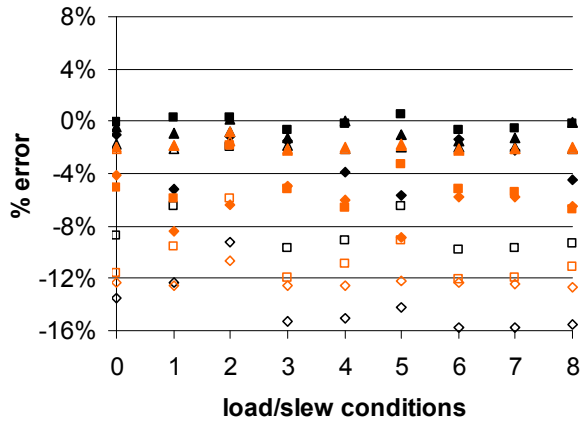


Figure 14c. Input-rise arc at V_{DD} = 0.5V

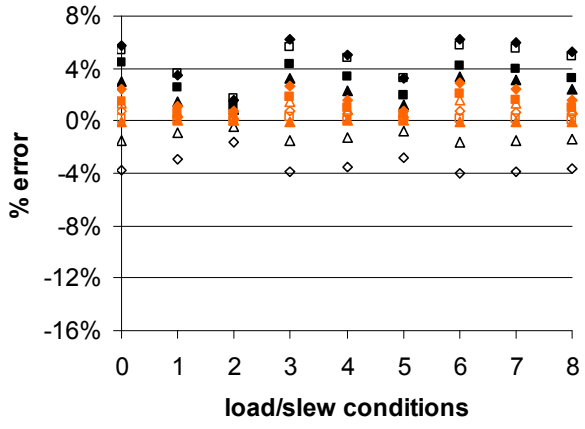


Figure 14d. Input-fall arc at V_{DD} = 0.5V

For the NAND gates, there is fairly good agreement at V_{DD} = 0.9V (Figures 14a-b). At V_{DD} = 0.5V, there is also good agreement (within 4~5%) except at the input-rise arc for the smaller two sizes of the NAND3 gates (Figures 14c-d). A slower slew rate tends to improve accuracy (Figures 14a-d). Output load does not significantly affect the percentage error estimates except

in NAND3 gates at the input-rise arc, where a greater output load increases percentage error (Figure 14c).

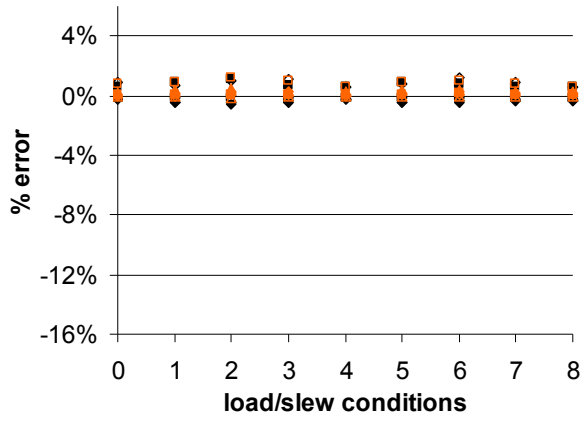


Figure 15a. Input-rise arc at V_{DD} = 0.9V

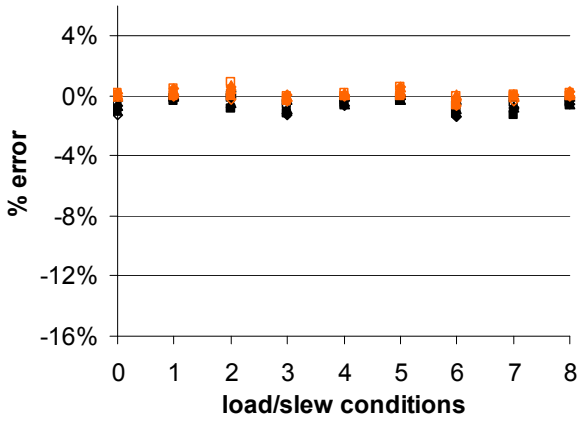
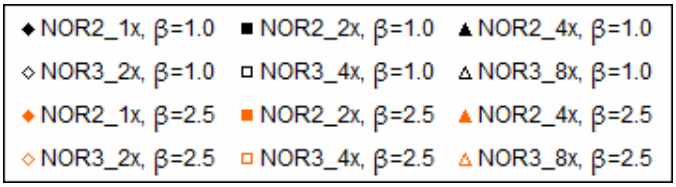


Figure 15b. Input-fall arc at V_{DD} = 0.9V



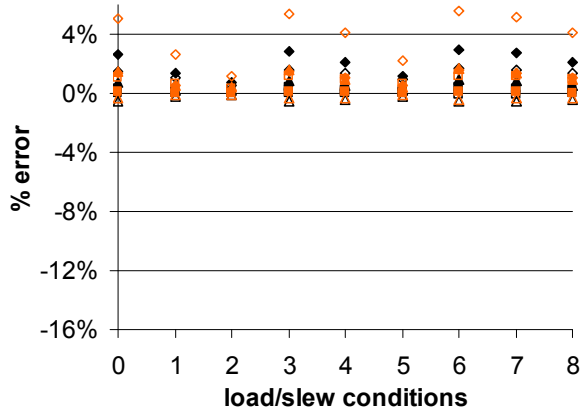


Figure 15c. Input-rise arc at V_{DD} = 0.5V

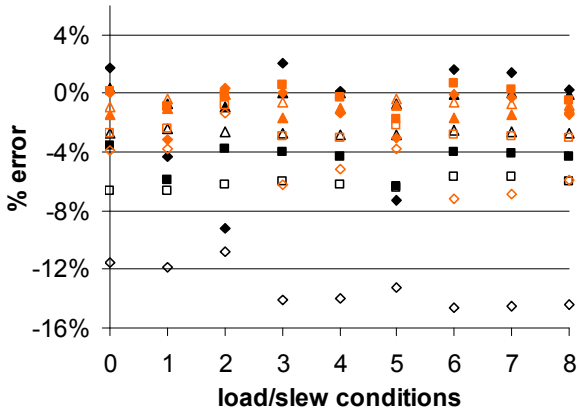


Figure 15d. Input-fall arc at V_{DD} = 0.5V

For the NOR gates, there is especially good agreement at V_{DD} = 0.9V (Figures 15a-b). At V_{DD} = 0.5V, there is also good agreement (within 4%) except at the input-fall arc for the smaller two sizes of the NOR3 gates (Figures 15c-d). A slower slew rate tends to improve accuracy (Figures

15c-d). Output load does not significantly affect the percentage error estimates except in NOR3 gates at the input-fall arc, where a greater output load increases percentage error (Figure 15d).

There are a number of trends observed: (1) higher V_{DD} results in less error, (2) larger transistor width (greater gate strength) leads to less error, (3) the effect of higher β ratio is more apparent at V_{DD} = 0.5V, where it generally results in less error, (4) for the input-rise arc vs. input-fall arc, larger stochastic delays and nonlinearity of the delay sensitivity curves lead to greater errors, (5) slower input slew rates generally result in less error, (6) output load does not significantly affect error except at V_{DD} = 0.5V for stacked transistors, and (7) the most significant source of error stems from transistor stacking in both NMOS (affecting NAND gates at input-rise arcs) and PMOS (affecting NOR gates at input-fall arcs). Note that the NMOS directly affects the delay at the input-rise arc and the PMOS directly affects the delay at the input-fall arc, since these gates are inverting and the output discharges through the NMOS and charges though the PMOS, respectively.

(1) Effect of V_{DD}

Since the SSTA characterization relies on linearizing operations, the estimation is more accurate when the delay is already approximately linear with respect to the transistor variables at V_{DD} = 0.9V. The nonlinearity of the delay sensitivity functions and hence CADFs at V_{DD} = 0.5V contributes to more error between the Monte-Carlo and SSTA results.

(2) Effect of Gate Strength

For the inverters, a larger transistor width results in less error. This is because local variations have less impact on larger transistors as the process fluctuations account for a smaller proportion of the total transistor parameters. This is generally true for the NAND and NOR gates as well, although in the cases with transistor stacking, the delay estimates are more strongly affected by the β ratio and properties of the sensitivity curves (described later in this section). It suggests that in general, larger gate size leads to better delay estimates.

(3) Effect of β Ratio

Increasing the β ratio makes the PMOS "stronger" since the wider PMOS draws more current and decreases propagation delay when the PMOS charges up to V_{DD} at the input-fall arc. For the inverters, higher β ratio leads to delay underestimates at the input-rise arc while lower β ratio leads to delay overestimates, which is especially apparent at V_{DD} = 0.5V (Figures 13a-d). This indicates that the accuracy of the SSTA algorithm is dependent on the relative strengths of PMOS and NMOS and input-rise/fall arc. For the NAND and NOR gates, higher β ratio slightly improves the SSTA estimates at the input-rise arc (Figures 14a, 14c, 15a, 15c) while it more significantly improves the SSTA estimates for the input-fall arc (Figures 14b, 14d, 15b, 15d), especially at V_{DD} = 0.5V. In these cases, the higher β ratio makes the PMOS stronger and compensates for the PMOS stacking. This results in better estimates for the input-fall arc. Similarly, stronger NMOS leads to better estimates for the input-rise arc.

(4) Input-Rise vs. Input-Fall Arcs

The error estimates are highly correlated to properties of the delay sensitivity curves, which affect the CADFs. Larger stochastic delays with respect to the NMOS/PMOS variables at the input-rise/fall arc and higher nonlinearity of the delay sensitivity curves lead to greater errors. An example can be seen for INVT_1x (Figure 16). In Figure 16, $P00_{p1}$ indicates the curve for the normalized total delay as a function of the p_1 parameter of PMOS00 (schematic in Figure 8), and similarly for the other curves. The stochastic delays can be measured by centering the intersection point of the sensitivity curves at zero (PMOS curves for the input-fall arc and NMOS curves for the input-rise arc). When β = 1.0, the sensitivity curves for the PMOS variables are steeper than those for the NMOS variables, and vice versa when β = 2.5. This corresponds to the results that the relatively stronger NMOS when β = 1.0 improves the delay estimates at the input-rise arc (Figure 13c), and the relatively stronger PMOS when β = 2.5 improves the delay estimates at the input-fall arc (Figure 13d).

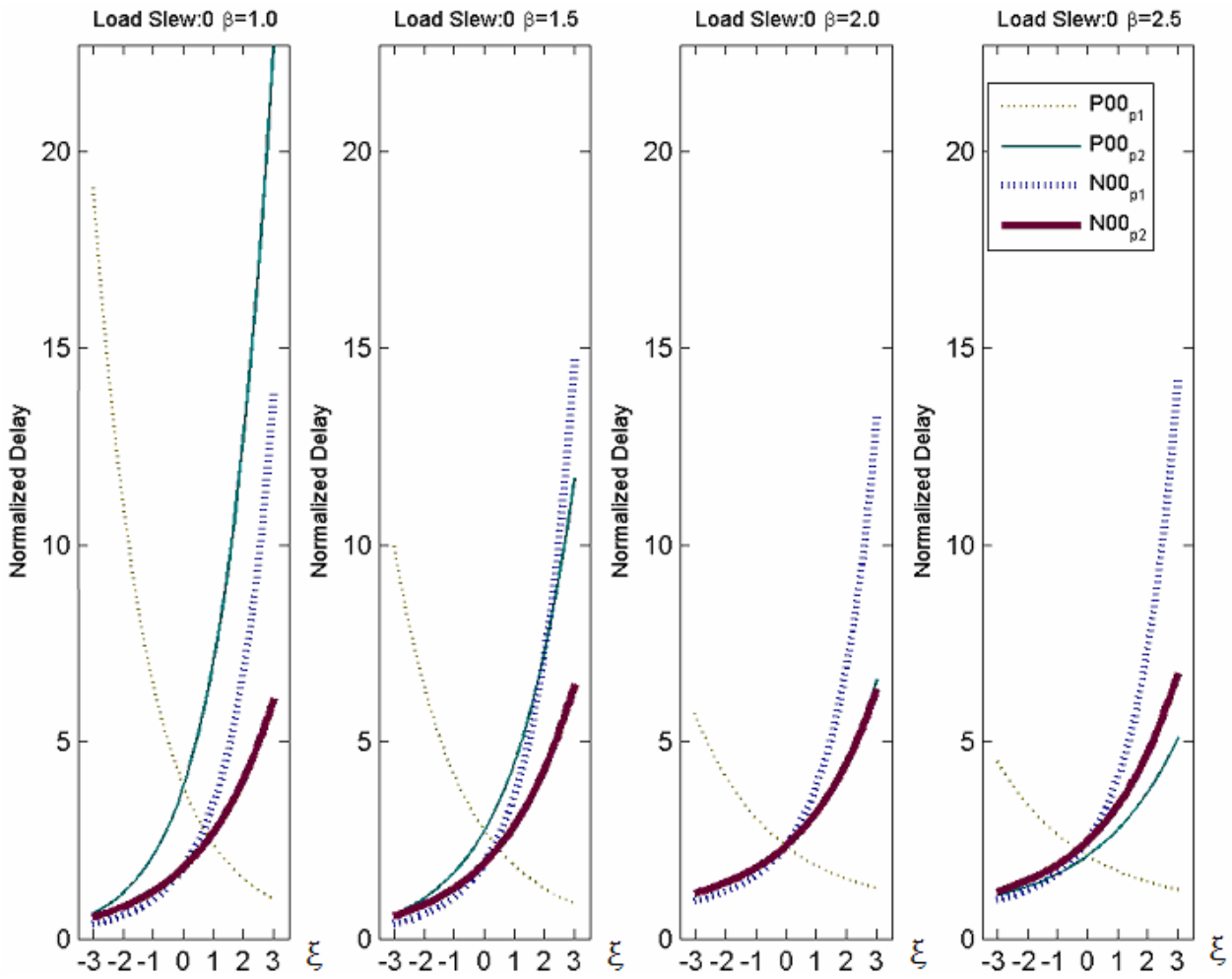


Figure 16. Sensitivity curves for INVT_1x at V_{DD} = 0.5V and load/slew condition 0, across β ratios

An example of how the linearity of the delay sensitivity curves affects delay estimates can be seen in Figure 17 for the input-fall arc of NOR2_1x at V_{DD} = 0.5V, where the PMOS curves are of interest ($P00_{p1}$, $P01_{p1}$, $P00_{p2}$, and $P01_{p2}$). The error is about 9% when β = 1.0, -2% when β = 1.5, and less than ±0.5% when β = 2.0 and 2.5. The most salient differences are the curvatures for $P01_{p1}$ and $P01_{p2}$, where they are high when β = 1.0 and low when β = 2.5.

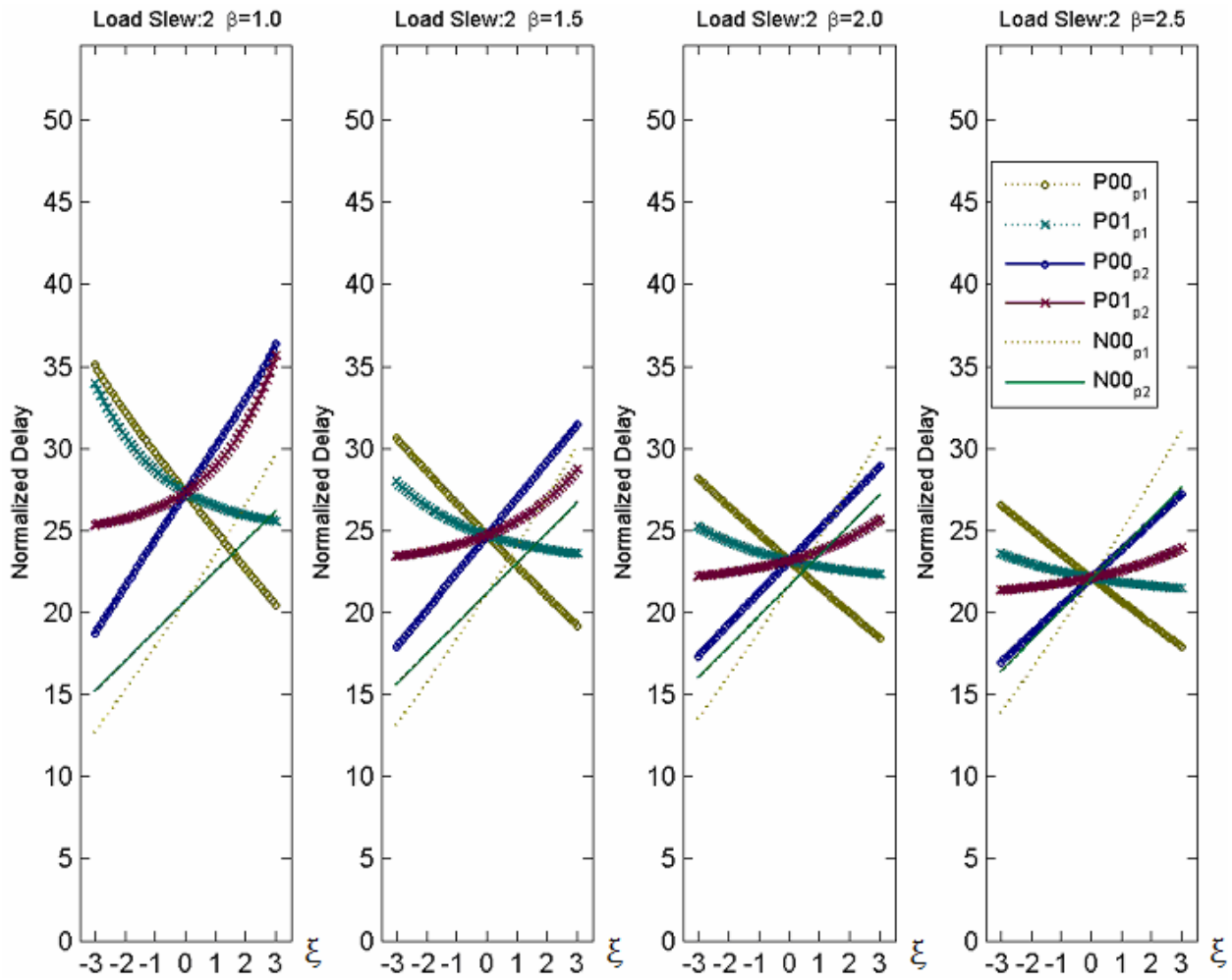


Figure 17. Sensitivity curves for NOR2_1x at V_{DD} = 0.5V and load/slew condition 2, across β ratios

This phenomenon mirrors the V_{DD} = 0.9V vs. V_{DD} = 0.5V case (Figures 2 and 3), where the CADFs with less curvature result in less error. Therefore, both the stochastic delay sensitivity curves of the PMOS variables *relative to* those of the NMOS variables as well as the linearity of the sensitivity curves strongly affect the accuracy of the delay estimates. The same interaction can also be observed in other gates, such as NOR3_4x (Figure 18) and NAND3_4x (Figure 19).

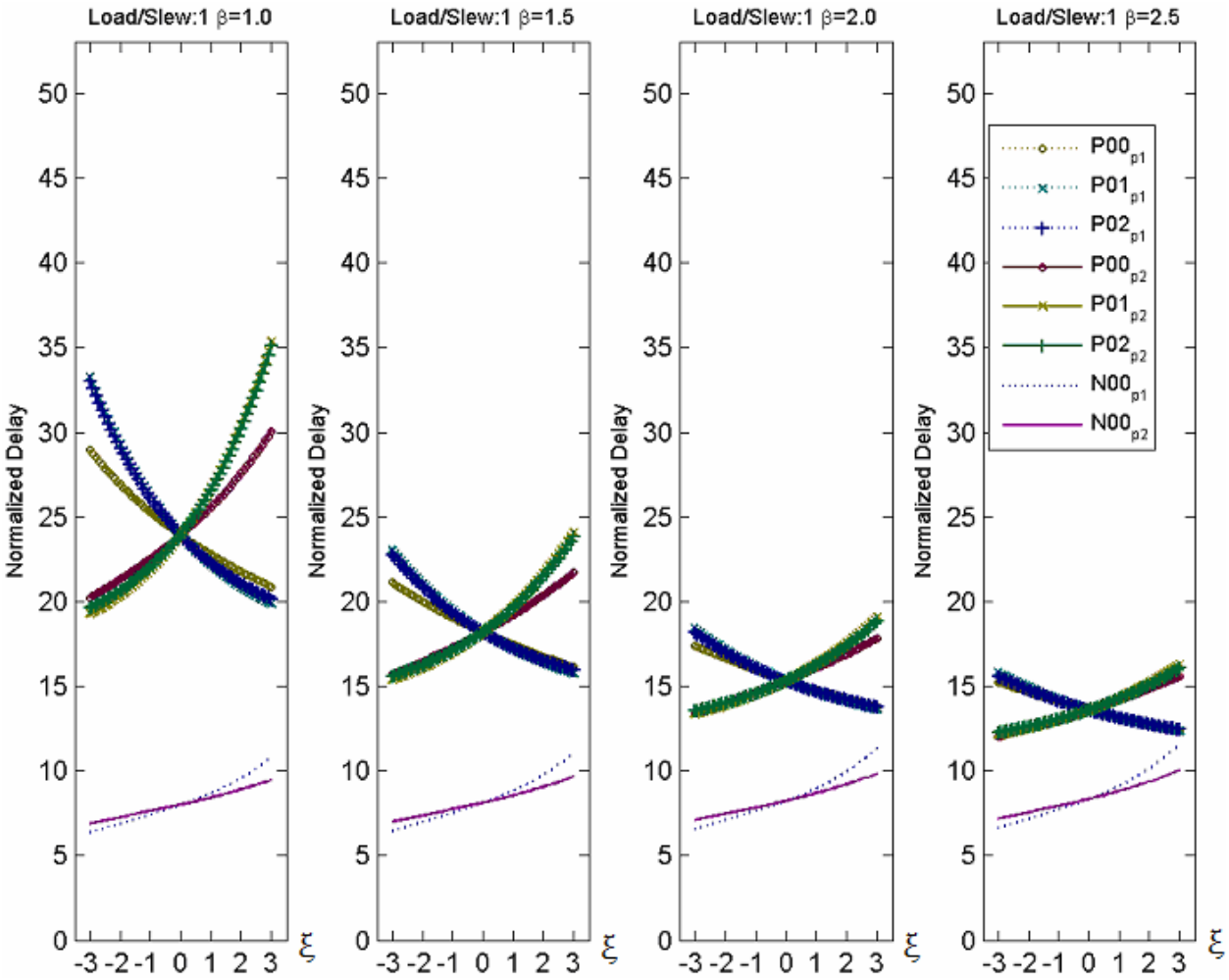


Figure 18. Sensitivity curves for NOR3_4x at V_{DD} = 0.5V and load/slew condition 1, across β ratios

The input-fall arc of NOR3_4x has an error of about -7% when β = 1.0 (where the PMOS is the weakest relative to NMOS), -5% when β = 1.5, -4% when β = 2.0, and -2% when β = 2.5 (where the PMOS is relatively the strongest and counters the stacking effect), as the PMOS sensitivity curves become more linear (Figure 18). A similar result can be seen at the input-fall arc of NAND3_4x, where the error was about 5% when β = 1.0 and less than 0.5% when β = 1.5, 2.0, and 2.5 (Figure 19).

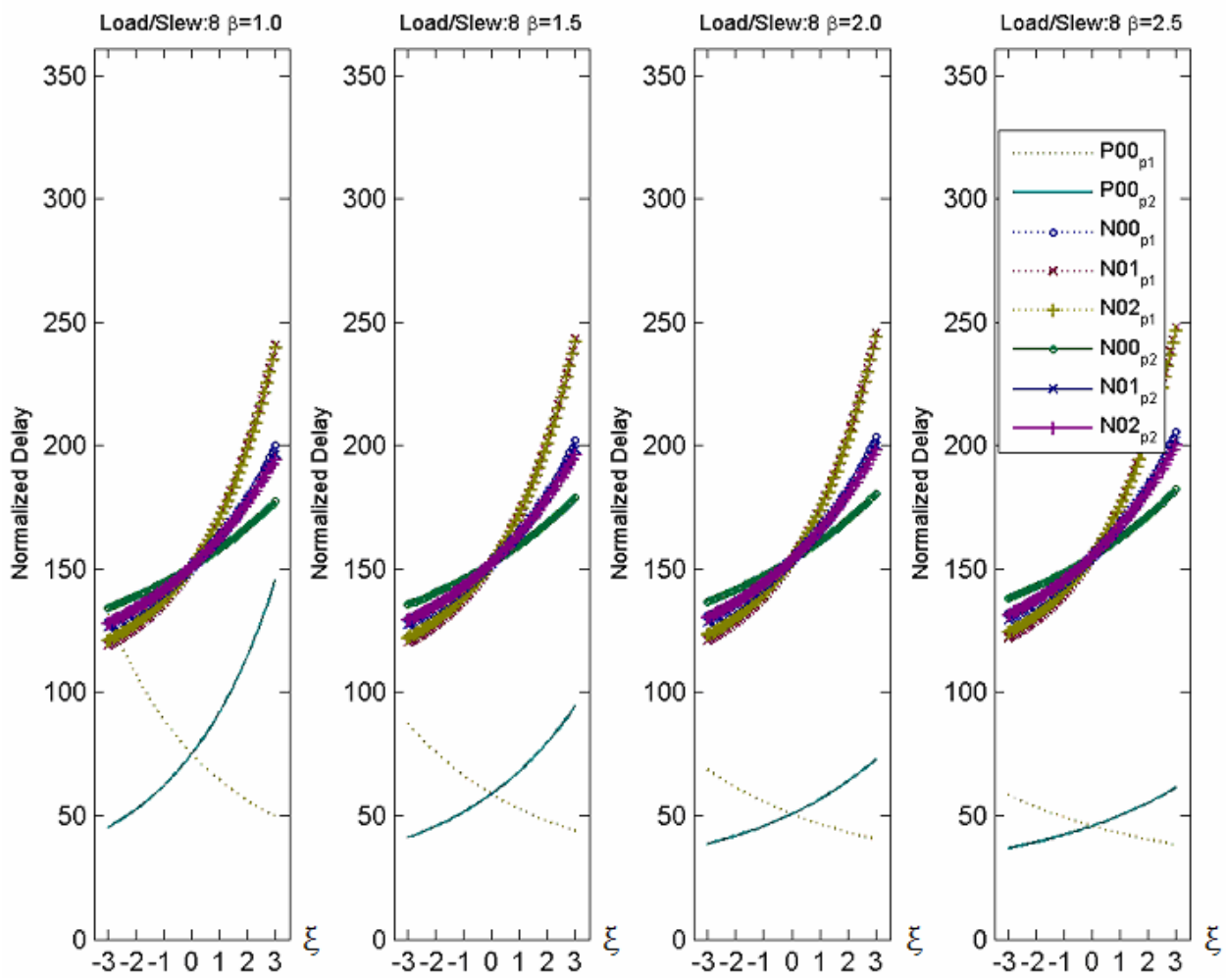


Figure 19. Sensitivity curves for NAND3_4x at V_{DD} = 0.5V and load/slew condition 8, across β ratios

(5) Effect of Input Slew

Slower slew rates in general yield less percentage error than faster slew rates. This is more pronounced in many cases of small capacitive loads at V_{DD} = 0.5V, e.g. load/slew condition 2 (Figures 13c-d, 14c-d, 15c-d). Sensitivity to input slew rates increases with smaller capacitive loads because propagation delays through internal transistor capacitances in such cases affect the overall delay more compared to the delay from output capacitance. In the case of load/slew condition 2 (i.e. the smallest output capacitance with the slowest input slew rate) for V_{DD} = 0.5V, the input and output voltage waveforms are "quasi-static" (Figure 20). This means that the input slew duration is much longer than the RC time constant at the output voltage node, and the propagation delay is thus almost independent of process variation and transistor behavior.

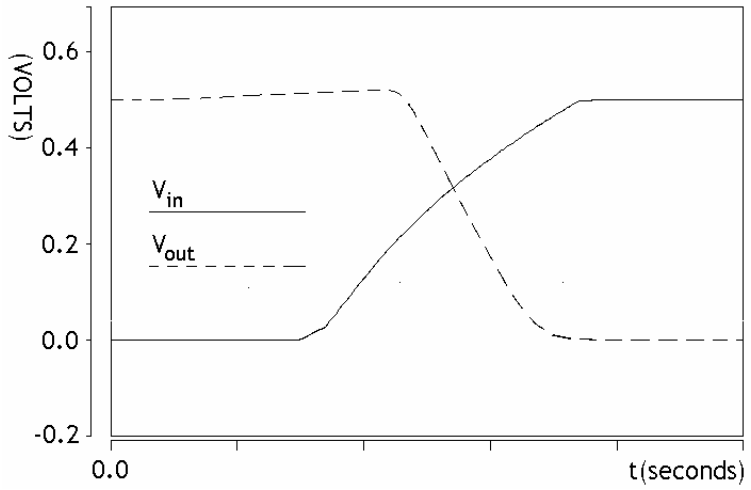


Figure 20. Input and output voltage waveforms for NAND3_2x at V_{DD} = 0.5V, load/slew condition 2

(6) Effect of Output Load

Capacitive load at the gate output does not have significant impact on the accuracy of the delay estimations, except at V_{DD} = 0.5V for small gates when charging or discharging through stacked transistors at a fast slew rate (Figures 14c-d, 15c-d, load/slew condition 6). In such cases, the delay behavior is almost exclusively dependent on the RC constant at the output voltage node (Figure 21), and is thus particularly sensitive to process variations.

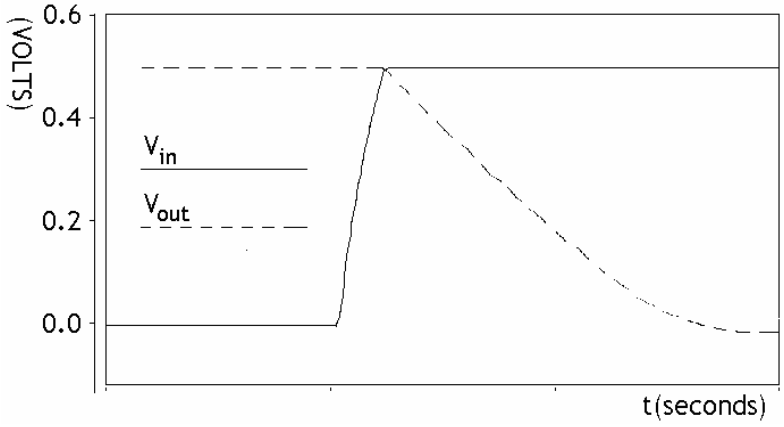


Figure 21. Input and output voltage waveforms for NAND3_2x at V_{DD} = 0.5V, load/slew condition 6

5.2. Output slew characterization results and discussion

The accuracy of the output slew estimation is correlated with the accuracy of the delay estimation, as both measurements are dependent on the output waveform. The percentage error estimates between SSTA and Monte-Carlo for the simulated gates are presented as follows: inverters (Figures 22a-d), NANDs (Figures 23a-d), and NORs (Figures 24a-d) across β ratio from 1.0 to 2.5. The percentage error is calculated at $\xi = 3$ as:

$$\text{Percent (\%) error} = (\text{SSTA output slew} - \text{MC output slew}) / \text{MC output slew} \quad (13)$$

There are a number of trends observed that are similar to those in cell delay characterization: (1) higher V_{DD} results in less error, (2) larger transistor width (greater gate strength) leads to less error, (3) higher β ratio generally yields slightly less error at $V_{DD} = 0.5V$ for NAND and NOR gates but has no significant effect at $V_{DD} = 0.9V$, (4) greater errors at the input-rise arc can be attributed to weaker NMOS relative to PMOS and vice versa, (5) slower input slew rate improves the percentage error in cases where the effect of transistor stacking is dominant, (6) output load does not significantly affect percentage error, and (7) transistor stacking contributes the most percentage error at $V_{DD} = 0.5V$, for the input-rise arcs of NAND gates in input-rise and input-fall arcs of NOR gates.

(1) Effect of V_{DD}

As with the delay characterization, the algorithm is more accurate at $V_{DD} = 0.9V$ for all gates (Figures 22 to 24), usually within 5%. At $V_{DD} = 0.5V$, the output slew duration for SSTA is more susceptible to other factors such as gate topology, transistor size, and input arc.

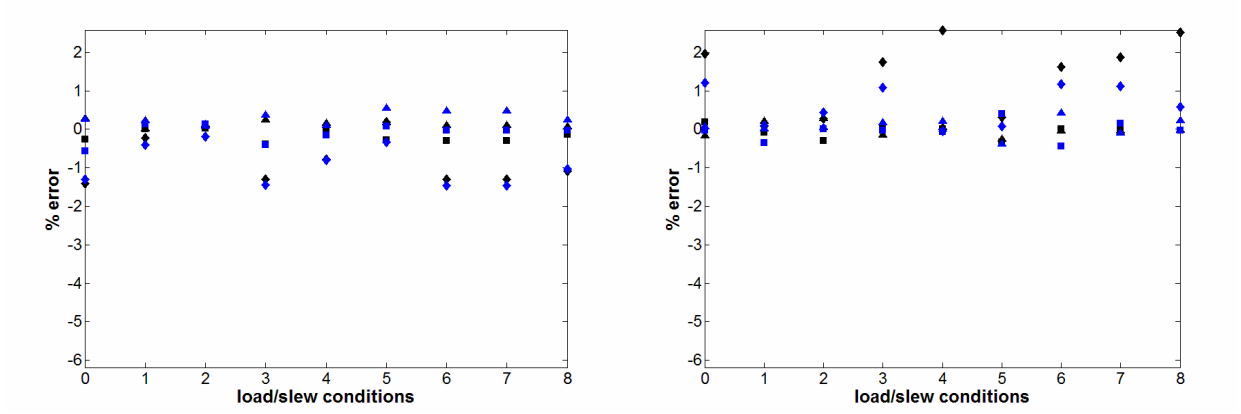


Figure 22a. Input-rise arc at V_{DD} = 0.9V **Figure 22b. Input-fall arc at V_{DD} = 0.9V**

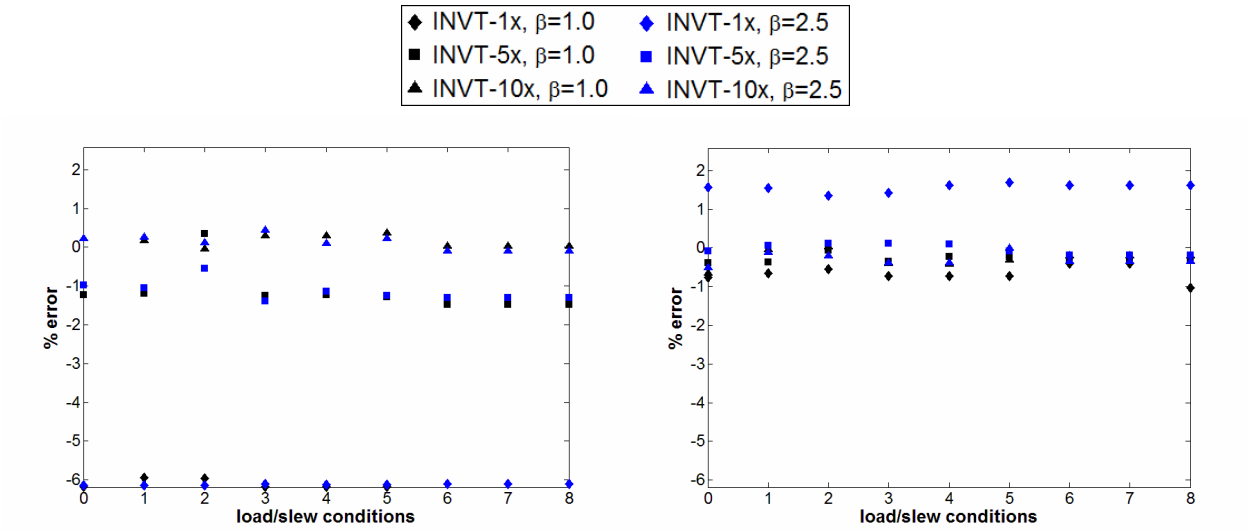


Figure 22c. Input-rise arc at V_{DD} = 0.5V **Figure 22d. Input-fall arc at V_{DD} = 0.5V**

(2) Effect of gate strength

For all gates, a larger transistor width within a single gate topology results in less error, similar to the case of delay. This effect is more pronounced at V_{DD} = 0.5V, where the minimum-sized gates are more susceptible to the effect of relative PMOS/NMOS strength. It can be observed in INVT_1x, where a greater β ratio improves the results at the input-fall arc and vice versa (Figures 22a-d). and the others in the form of "stack effect" (Section 5.3).

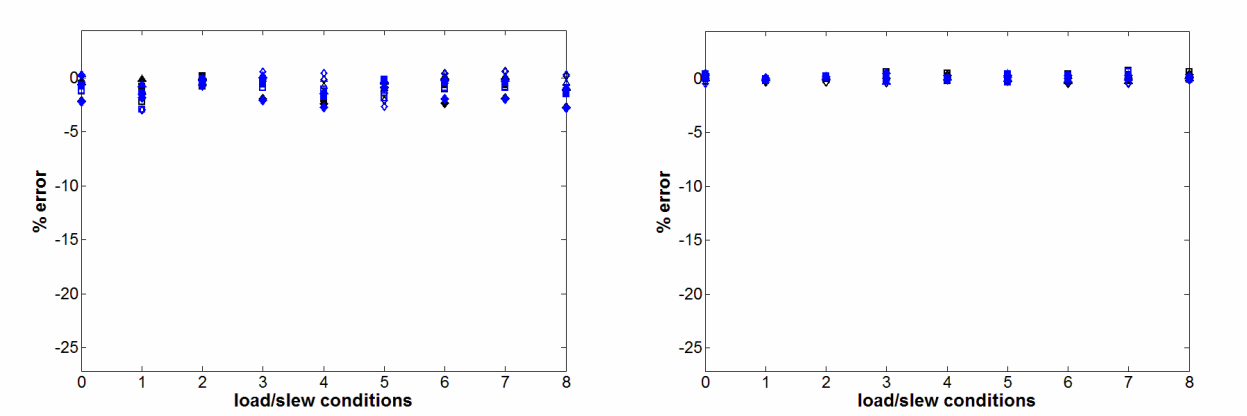


Figure 23a. Input-rise arc at V_{DD} = 0.9V

Figure 23b. Input-fall arc at V_{DD} = 0.9V

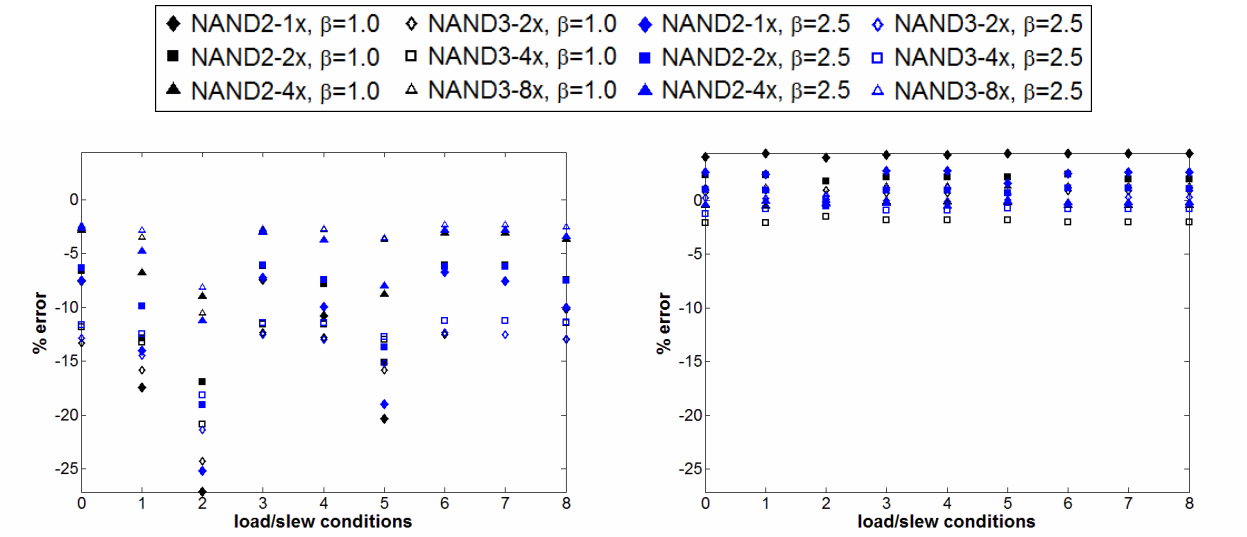


Figure 23c. Input-rise arc at V_{DD} = 0.5V

Figure 23d. Input-fall arc at V_{DD} = 0.5V

(3) Effect of β ratio and (4) input-rise vs. input-fall arc

β ratio does not significantly affect the accuracy at V_{DD} = 0.9V, although for INVT_1x at the input-fall arc and larger output loads, higher β ratio improves the accuracy (Figure 22b). At V_{DD} = 0.5V for INVT_1x, SSTA underestimates the output slew duration by about 6% at the input-rise arc regardless of β ratio (Figure 22c). However, at the input-fall arc, the algorithm underestimates by about 1% at β = 1.0, and overestimates by about 2% at β = 2.5 (Figure 22d). It indicates that since the β ratio only changes the PMOS width but not the NMOS width, the input-rise arc is unaffected while the input-fall arc percentage errors are less when the β ratio is higher (stronger PMOS). A similar tendency can be observed in NAND and NOR gates, where

the input-rise arc is unchanged by β ratio, while a larger β ratio improves the percentage error at the input-fall arc (Figures 23c-d, 24c-d).

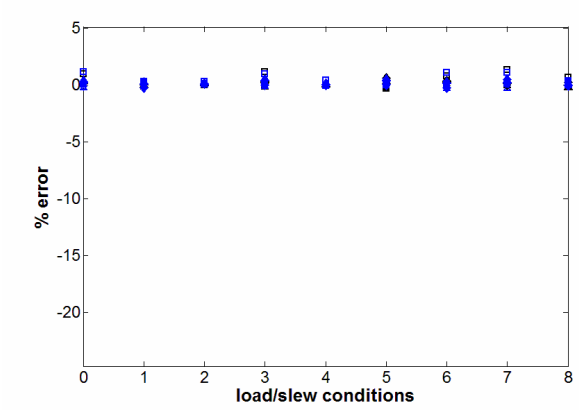


Figure 24a. Input-rise arc at V_{DD} = 0.9V

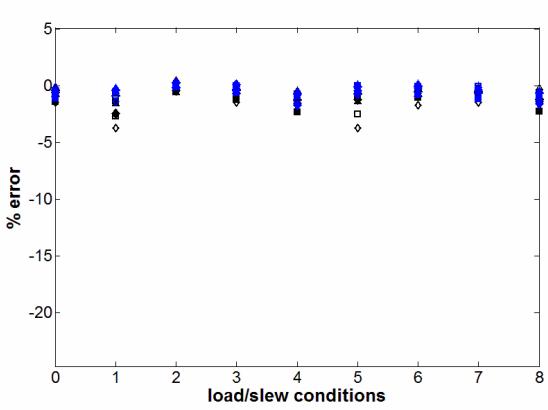
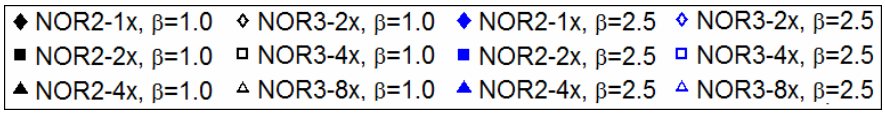


Figure 24b. Input-fall arc at V_{DD} = 0.9V



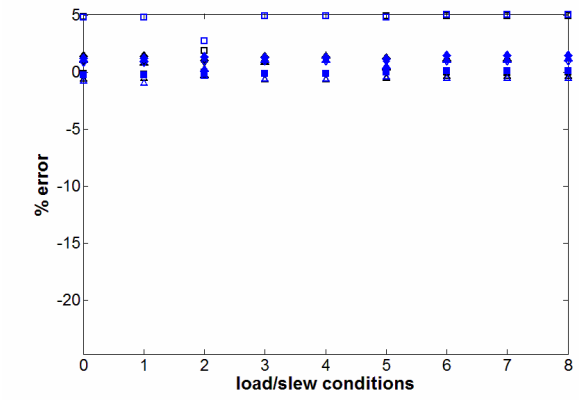


Figure 24c. Input-rise arc at V_{DD} = 0.5V

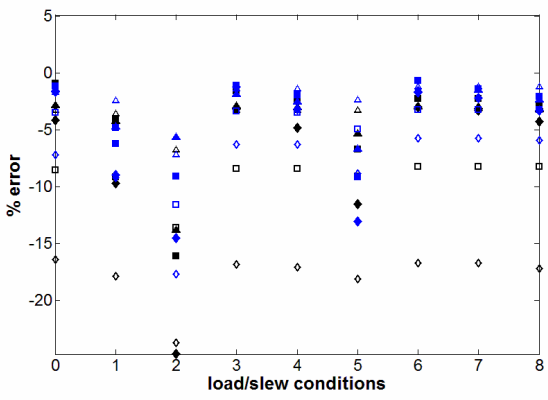


Figure 24d. Input-fall arc at V_{DD} = 0.5V

(5) Effect of input slew rate

For the inverters, faster input slews generally resulted in larger error, although at V_{DD} = 0.5V, the percentage error appears to be essentially unaffected by input slew rate (22c-d). For the NAND and NOR gates, the percentage error is visibly influenced by the input slew rate when there is charging or discharging through stacked transistors (Figures 23c, 24d) at small output loads, otherwise it is not significantly affected by the input slew rate.

(6) Effect of output load

The output load does not considerably influence percentage error in most cases, although for the stacked transistor arcs (Figures 23c, 24d), it mitigates the effect of slew rates. This is similar to the case for delay estimations.

5.3. Transistor stacking effect in cell characterization

We have observed in Sections 5.1 and 5.2 that the percentage error is greater when the output is being charged or discharged through a stack of transistors compared to the case when it is being charged/discharged through a single transistor. The analysis in this section shows that this increased error is the result of cross terms in the nonlinear expansion of $D(\zeta_1, \zeta_2, ..\zeta_N)$. The characterization technique presented in this thesis accounts for nonlinearities in each of the variables ζ_i but not for cross terms. The analysis in this section illustrates the origin of the error in the case of the stacked devices NMOS devices in a 3-imput NAND operating at $V_{DD} = 0.5V$.

Figure 25 shows the analysis with respect to two normalized transistor variables ζ_1 and ζ_2. Only the delay is analyzed here because the error in output slew is correlated with the error in delay, so the same analysis applies. The dashed black arc is the hyper-sphere of radius 3. The solid (red) curve represents the delay $D(\zeta_1, \zeta_2,.., \zeta_N) = D_{3\sigma}$, where only ζ_1 and ζ_2 vary and all the other ζ_i are kept at their operating points. The dashed (blue) line is the linear approximation of the delay at the 3-sigma operating point. The horizontal and vertical green Gaussian distributions are projections of the delay PDF when it is mapped onto the axes of ζ_1 and ζ_2, respectively.

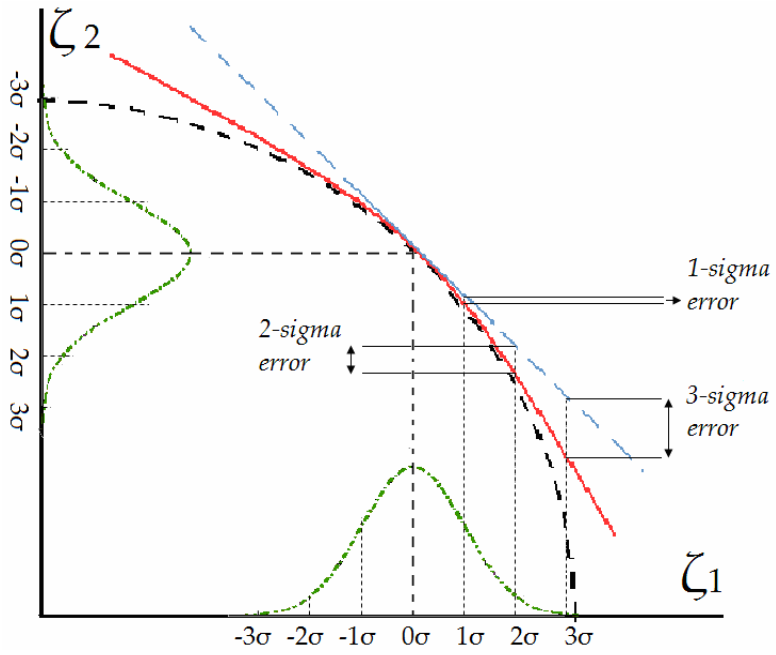
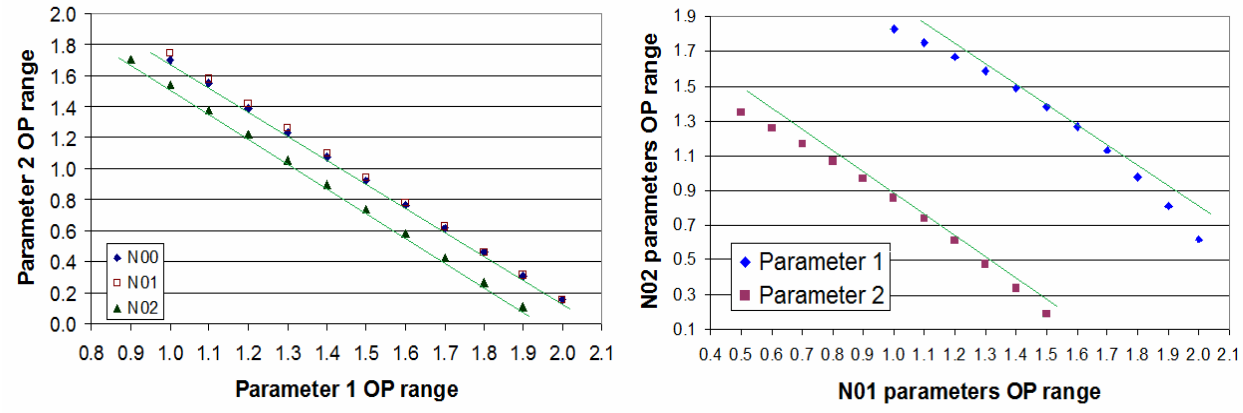


Figure 25. A 3-sigma iso-delay function with respect to normalized transistor variables ζ_1 and ζ_2

Experimentally, when we examine the space of transistor parameters p_1 and p_2 for a single transistor, the delay $D(\zeta_1, \zeta_2)$ is highly linear as shown in Figure 26a. N00, N01, and N02 refer to the three stacked NMOS devices in a 3-input NAND. However, in the space of random variables from different transistors in the stack, we find substantial nonlinearity in $D(\zeta_1, \zeta_2)$ in the region of the operating point as shown in Figure 26b.



(a) **(b)**

Figure 26. 3-sigma iso-delay functions with respect to two normalized transistor variables in NAND3_2x, (a) for p_1 and p_2 within a transistor, (b) $N01_{p1}$ vs. $N02_{p1}$ and $N01_{p2}$ vs. $N02_{p2}$

The solid curve in Figure 25 represents the p_1 and p_2 data in Figure 26b as both data sets are nonlinear. The corresponding delay percentage errors are as follows:

- Error at 1-sigma ≈ -0.5%
- Error at 2-sigma ≈ -6%
- Error at 3-sigma ≈ -16%

Within the range of values that contribute to the integrand of the convolution integral for the delay function, there is sufficient nonlinearity to account for the increased percentage error.

The analysis above shows that the discrepancy between SSTA and MC for cells whose delay is dominated by charge or discharge through stacked transistors is attributed to the increased nonlinearity. In the case of stacked transistors, the additional nonlinearity arises from the importance of cross terms in the delay function $D(\zeta_1, \zeta_2, \ldots \zeta_N)$. If we only consider the effect of two of the transistor variables, e.g. ζ_1 and ζ_2, the delay as a function of normalized transistor random variables can be approximated by:

$$D(\zeta_1, \zeta_2, \zeta_3^{op}, \zeta_4^{op}, \ldots, \zeta_N^{op}) = \alpha_1 \delta\zeta_1 + \alpha_2 \delta\zeta_2 + \beta_1 (\delta\zeta_1)^2 + \beta_2 (\delta\zeta_2)^2 + \gamma_{1,2} \delta\zeta_1 \delta\zeta_2$$

(14)

where $\delta\zeta_i = \zeta_i - \zeta_i^{op}$.

Our approach in this work is to account for the non-zero β_i coefficients by linearizing about the operating points. Nevertheless, the cross term (with γ coefficient) is not comprehended because doing so would increase computational complexity.

This thesis shows the accuracy that can be obtained under the assumption that cross terms are ignored. We see that the error due to this assumption is worst for cells with stacked devices, but even in this case, the delay percentage errors are shown to be less than about 16%. When error of this magnitude is not acceptable, cross terms can be included, and it is likely this can be done in an approximate way that is computationally efficient.

6. Conclusion

This work has described a computationally efficient statistical static timing analysis (SSTA) technique that addresses intra-die (local) variations when V_{DD} is at near-threshold or sub-threshold, simulated on a scaled 32nm CMOS standard cell library. The technique features an operating point approach that is computationally efficient compared to conventional Monte-Carlo analysis. Only 10~20 simulations are needed to completely characterize the propagation delay and output slew of a cell at a given input and output condition, ξ-sigma away from the global weak corner.

The percentage errors between the SSTA algorithm and the Monte-Carlo method mostly stay within 5% for all the cells tested – inverters, two-input and three-input NAND and NOR gates of different sizes. Cells simulated at V_{DD} = 0.9V have less percentage errors than the same cells at V_{DD} = 0.5V. Greater gate strength, i.e. larger transistor widths, results in less error. For inverters, the arc (input-rise or input-fall) that charges/discharges through the relatively stronger transistor has less error. For NANDs and NORs, stronger PMOS (higher β-ratio) improves accuracy in general. Transistor stacking results in the largest sources of error at V_{DD} = 0.5V. It is attributed to correlations among transistor parameters within the stack, which may be quantified and accounted for accordingly. The proposed algorithm thus demonstrates potential to be incorporated into CAD flow for IC verification for transistor nodes beyond 45nm at near-threshold to sub-threshold operation.

7. Future work

For cell characterization of stacked devices, the observed cross-terms (Section 5.3) that result from interactions among transistor variables in different transistors could be accounted for by systematically measuring the curvatures of the delay functions with respect to two of the transistor variables. This would call for more detailed work to investigate which cells have large cross terms in the delay or output slew functions, and to devise a computationally efficient way to approximately include the cross terms.

The proposed SSTA algorithm could be used to characterize all common cells, such as adders [21], latches, and registers [21]. In the case of sequential elements, it would be necessary to understand the potential issues that result form pass-gate logic. The algorithm has been extended to (1) multi-stage chains with cascaded cells, where each cell has its associated operating points analogous to those of transistor parameters within a cell, (2) hold-time analysis of cascaded sequential elements in the presence of local variations, and (3) timing of convergent paths, where the nominal delays might be similar but the operating point delays with local variations would differ greatly [21].

This work has characterized the cell/arc delay function (CADF) and cell/arc slew function (CASF) of cells with twelve points from $0.25 < \xi < 3$. In production characterization, this would require extensive computation. Future work would need to determine the trade off between the number of points in the piece-wise linear approximation to the CADF/CASF and resulting accuracy. Strategies for selecting the points on the CADF/CASF curve for the most accurate characterization would need to be determined. When these queries are addressed, it is possible to streamline the algorithm and apply it to the design flow of a deep submicron CMOS process operating at near-threshold to sub-threshold supply voltages.

8. References

[1] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations", *International Conference on Computer Aided Design*, 2007.

[2] B. Liu, "Gate level statistical simulation based on parameterized models for process and signal variations", *International Symposium on Quality Electronic Design*, 2007.

[3] M. Bühler, J. Koehl, J. Bickford, et al., "DFM/DFY design for manufacturability and yield – influence of process variations in digital, analog and mixed-signal circuit design", *DATE Special Session*, 2006.

[4] B. H. Calhoun and A.P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS", *IEEE Journal of Solid-State Circuits, vol. 41, no. 7*, 2006.

[5] P. Andrei and I. Mayergoyz, "Random doping-induced fluctuations of sub-threshold characteristics in MOSFET devices", *Solid-State Electronics, vol. 47,* (2003) 2055–2061.

[6] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture", *Design Automation Conference*, 2003, pp. 338–342.

[7] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: Variability characterization and modeling for 65- to 90-nm processes", *Custom Integrated Circuits Conference*, 2005, pp. 593–600.

[8] N. Drego, A. Chandrakasan, and D. Boning, "A test-structure to efficiently study threshold-voltage variation in large MOSFET arrays", *International Symposium on Quality Electronic Design*, 2007.

[9] J. A. G. Jess, K. Kalafala, S. R. Naidu, R. H. J. M. Otten, and C. Visweswariah, "Statistical timing for parametric yield prediction of digital integrated circuits", *Design Automation Conference*, 2003, pp. 932–937.

[10] A. Singhee, S. Singhal, and R. A. Rutenbar, "Practical, fast Monte Carlo statistical static timing analysis: why and how," *International Conference on Computer Aided Design*, 2008, pp. 190-195.

[11] V. Veetil, D. Sylvester, and D. Blaauw, "Efficient Monte Carlo based incremental statistical timing analysis", *International Conference on Computer Aided Design*, 2008, pp. 676-681.

[12] H. Chang, V. Zolotov, C. Visweswariah, and S. Narayan, "Parameterized block-based statistical timing analysis with non-Gaussian and nonlinear parameters", *Design Automation Conference*, 2005, pp. 71–76.

[13] H. Chang, and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal", *International Conference on Computer Aided Design*, 2003, pp. 621-625.

[14] V. Khandelwal and A. Srivastava, "A general framework for accurate statistical timing analysis considering correlations", *Design Automation Conference*, 2005, pp. 89-94.

[15] L. Cheng, J. Xiong, and L. He, "Nonlinear statistical static timing analysis for non-Gaussian variation sources", *Design Automation Conference*, 2007, pp. 250-255.

[16] L. Zhang, W. Chen, Y. Hu, J. A. Gubner, and C. C.-P. Chen, "Correlation-preserved non-Gaussian statistical timing analysis with quadratic timing model", *Design Automation Conference*, 2005, pp. 83-88.

[17] M. Orshansky and K. Keutzer, "A general probabilistic framework for worst case timing analysis", *Design Automation Conference*, 2002, pp. 556-561.

[18] J. Singh and S. Sapatnekar, "Statistical timing analysis with correlated non-Gaussian parameters using independent component analysis", *Design Automation Conference*, 2006, pp. 155-160.

[19] C. Visweswariah, C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, S. Narayan, D. K. Beece, J. Piaget, N. Venkateswaran and J. G. Hemmett, "First-order incremental block-based statistical timing analysis", *Design Automation Conference*, 2004, pp. 331-336.

[20] Y. Zhan, A. J. Strojwas, X. Li, and L. T. Pileggi, "Correlation-aware statistical timing analysis with non-Gaussian delay distribution", *Design Automation Conference*, 2005, pp. 77-82.

[21] R. Rithe, S. Chou, D. Buss, and A. Chandrakasan, "SSTA for ultra-low voltage operation", *International Conference on Computer Aided Design*, 2009 (in submission).