Transistor-Level Gate Model Based Statistical Timing Analysis Considering Correlations

Qin Tang, Amir Zjajo, Michel Berkelaar and Nick van der Meijs
Circuits and Systems, Delft University of Technology
e-mail: qintang@ieee.org

Abstract—To increase the accuracy of static timing analysis, the traditional nonlinear delay models (NLDMs) are increasingly replaced by the more physical current source models (CSMs). However, the extension of CSMs into statistical models for statistical timing analysis is not easy. In this paper, we propose a novel correlation-preserving statistical timing analysis method based on transistor-level gate models. The correlations among signals and between process variations are fully accounted for. The accuracy and efficiency are obtained from statistical transistor-level gate models, evaluated using a smart Random Differential Equation (RDE)-based solver. The variational waveforms are available, allowing signal integrity checks and circuit optimization. The proposed algorithm is verified with standard cells, simple digital circuits and ISCAS benchmark circuits in a 45nm technology. The results demonstrate the high accuracy and speed of our algorithm.

I. INTRODUCTION

In static timing analysis (STA), the need for accuracy has driven the development of delay models. A long time industry standard is the traditional nonlinear delay model (NLDM) [1], which models gate delay and output slew as a nonlinear function of input slew (S_in) and output effective capacitance (C_eff). This only represents the signal waveform very crudely, so more recently current source models (CSMs) [2]–[7] have gained attention. Instead of modeling gate delay directly, CSM models every gate with a current source and multiple capacitors, which depend on S_in and C_eff [2] or input and output voltages [3]–[7]. This CSM representation improves delay calculation accuracy thus there is a level of industry acceptance. However, most CSMs use the assumption that only one input is switching while others are static. As a consequence, some effects like multiple input simultaneous switching (MISS) are not modeled, leading to large errors [7]. Recently, even higher accuracy is achieved by transistor-level gate models [8]–[10] which can accurately model effects like MISS. Since most CSMs [3]–[7] and transistor-level gate models [8]–[10] have elements dependent on input and output voltages, they are called Voltage-in Voltage-out (ViVo) gate models in this paper.

The down-scaling of technology brings a significant increase in the device and interconnect manufacturing process variations, causing larger spreads in circuit timing uncertainty. To analyze the resulting variation in delay, STA can be performed at multiple corners. Although STA is accurate at every corner, the corner-based method is too pessimistic since it is close to impossible for all process parameters to have extreme values at the same time. Additionally, if the number of process variations is N_p, there are 2^{N_p} process corners—often too many to analyze. Consequently, statistical STA (SSTA) has been developed, which requires statistical gate modeling or statistical gate delay models.

Many published SSTA methods, denoted as function-based SSTA in this paper, model gate delay as a (non)linear function of process variations. The coefficients are stored in look-up tables with entries of S_in and C_eff. This modeling method is similar to the NLDM concept [1]. Most SSTA methods assume S_in and C_eff are fixed when calculating gate delay distributions. However, due to process variations in receiver and driver, both S_in and C_eff are variational. Not considering the statistical S_in and C_eff can result in 30% delay errors and even worse for bigger circuits [4]. Also, like NLDM, function-based SSTA models can not account for resistive interconnect loads and nonlinear input waveforms. Furthermore, the variational waveforms can not be obtained since only delay and slew variations are available. Additionally, function-based SSTA is entirely based on non-physical or empirical models, which is the major source of inaccuracy [3].

For these reasons, to increase accuracy, CSM models also have been extended for use in SSTA [3]–[6]. In [3], the variational voltages and all elements in CSM are modeled as a stochastic first-order expression in terms of process variations. Then the output voltage is treated as a Markovian process for delay distribution calculation. In [4], the current source value and capacitances in CSM are modeled as a quadratic Hermite function of process variations. Crossing time distributions are calculated by process variation sampling and linear interpolation. A CSM with parametric nonlinear voltage-dependent current source and parametric capacitance is used in [5] and [6]. The voltage in [5] is represented as a time-domain statistical variable and time-domain integration is performed. The gate output voltage distribution in [6] is obtained by Monte Carlo (MC) sampling. However, these methods are just verified on several simple single gates, and the correlations between input and output signals and among process variations are not considered.

To gain even higher accuracy than the above CSM methods,
and to be able to see the important effects such as MISS, in this paper we propose a statistical timing analysis solution based on statistical transistor level gate models to provide statistical information of any required crossing time. The proposed solution has the following features: 1) The variational waveform, which models several varying crossing times, is calculated and propagated through circuits; 2) In the proposed Random Differential Equation (RDE-) based statistical simulation, all input signals are considered together and calculated directly, thus fundamentally addressing MISS in statistical timing analysis; 3) As we use a common format for waveforms and elements in gate models, the correlations among input signals and between input and output signals are preserved during probability density function (pdf) computation; 4) Arbitrary distributions of process variations can be handled in pdf calculation. The proposed algorithm is verified on some simple circuits and ISCAS85 benchmark circuits considering correlations.

Compared to our previous publications on this topic [10]–[12], in this paper we contribute: 1) optimization on our transistor model for gate modeling; 2) improved algorithms to solve the RDE system in the simulation more efficiently; 3) consideration of correlations between different signals and different process parameters; 4) experiments including circuits, not just single gates.

II. TRANSISTOR-LEVEL GATE MODELING

Transistor-level gate models have been introduced for higher accuracy and faster characterization of STA [9], [10]. Since the gate models are constructed at the transistor level, the transistor model is a key issue which needs to have sufficient accuracy, account for the impact of process variations, while still being simple enough to be evaluated efficiently. In this paper, we use the table-based Statistical Simplified Transistor Model (SSTM) [10] for gate modeling. Every transistor in the circuit is modeled by a current source $i_{ds}$ and five capacitors as shown in Fig. 1. In SSTM, gate channel capacitances, $c_{gs}$, $c_{gb}$ and $c_{gd}$, are modeled as a function of $V_{gs}$ and $V_{ds}$ while junction depletion capacitances, $c_{sb}$ and $c_{db}$, are represented, for simplicity, by constant values. The current and capacitances in the SSTM are modeled as a linear function of the process variations of interest $\xi$.

For stage-by-stage timing analysis of large circuits, the input capacitance of every gate is required. In the SSTM-based gate models, the input capacitance of a gate at any input is the sum of the gate capacitances $C_g$ of all the transistors connected to that input, where $C_g$ is the sum of all gate channel capacitances in SSTM. We improved on [10] by characterizing the $C_g$ of every transistor in the library w.r.t. $V_g$ only, based on the following concerns: i) The $c_{gs}, c_{gb}$ and $c_{gd}$ in SSTM [10] depend on $V_{ds}$ which is unknown for the previous gate; ii) Evaluating a simple closed-form expression is much more efficient than interpolating from matrices three times.

The gate models are constructed by replacing every transistor in the gate by its corresponding SSTM. To reduce the complexity of the interconnect model after RC extraction, model order reduction techniques can be employed, such as [13]. Every resistance and capacitance in the interconnect model is also modeled as a linear function w.r.t. $\xi$. It should be noted that the statistical timing analysis method presented in Section III is independent of the type of ViVo-gate models. In other words, it can be used with other ViVo-gate models.

By using efficient threading algorithm and multiple processors, [9] shows that it is practical to use transistor level gate models for multi-million gate STA runs to reach the combination of accuracy and speed.

III. STATISTICAL TIMING ANALYSIS

For optimized ViVo-gate models, like CSMs and transistor-level gate models [3]–[10], nodal analysis (NA) or modified NA (MNA) is used for gate simulation. The NA/MNA equation solution is deterministic with all parameter values fixed. If process variations are included, typically a corner-based method is used as an outer loop. As mentioned in Section I, however, corner-based methods are pessimistic and time-consuming. In this section, we propose a RDE-based statistical simulation method which provides variational waveforms directly after simulating only once. The theoretical derivation in Section III-A is an extension based on our previous work [11].

A. RDE-based statistical simulation

The deterministic NA equation can be written in the compact format:

$$F(\dot{x}, x, t, p) = 0 \quad p = p_0, x(t_0) = x_0$$

where $x$ denotes node voltages, $\dot{x}$ is its time derivative and $p_0$ is the nominal value of the process parameter vector $p$. Denote $x_s(t)$ as the solution of (1) which satisfies:

$$F_s = F(\dot{x}_s, x_s, t, p_0) = 0 \quad x(t_0) = x_0$$

Since all process parameters have their nominal values $p_0$, $x_s(t)$ is deterministic, which means it can be solved similar to ViVo-gate model based STA methods, like [8]–[10]. However, if process variations are considered the solution becomes statistical.

If we take into account process variations, $p = p_0 + \xi$ where $\xi$ is the process variation vector which includes both global and local variations. Consequently, (1) becomes a random differential equation (RDE):

$$F_s = F(\dot{x}, x, t, \xi) = 0 \quad \xi = p - p_0, x(t_0) = x_0 + \delta_0$$

where $\delta_0$ denotes the initial condition variation caused by process variations. It is worth noticing that the main difficulty to solve (3) is the high nonlinearity w.r.t. the random variables.
ξ and the large number of process variations including local variations. If the number of local variations in the problem is very large, techniques exist to reduce it considerably [9]. In order to make (3) manageable, it is linearized\(^1\) by a truncated Taylor expansion around \(x_s\) and \(p_0\):

\[
F_x \approx F_x + \frac{\partial F_x}{\partial x_s}(t)(x(t) - x_s(t)) + \frac{\partial F_x}{\partial p_0}(t)\xi = 0 \quad (4)
\]

where \(F_x\) is defined in (2).

To simplify the notation, the variation of state variable \(x\) is denoted by \(y\), thus (4) can be rewritten as (5)

\[
y(y(t) = E(x_s) + F(x_s)\xi \quad y(t_0) = y_0 = \delta_0 \quad (5)
\]

\(C, E\) and \(F\) are \(N_y \times N_y, N_y \times N_x\) and \(N_x \times N_p\) matrices respectively, where \(N_y\) is the number of unknown nodes and \(N_p\) is the number of process variations. Consequently, the nonlinear equation (3) is converted to a linear RDE in \(y\) with \(x_s\)-dependent coefficient matrices. \(x_s(t)\) can be solved by well-known deterministic STA methods like in [8]–[10].

Unfortunately, the variation of state variable \(y\) can not be calculated directly from (5) since \(ξ\) is a random variable. According to the Random Differential Equation (RDE) theorem [14], (5) has a unique mean square solution which can be represented as (6)

\[
y(t) = Ψ(t,t_0)y_0 + Θ(t)ξ \quad (6)
\]

where \(Ψ(t,t_0)\) is the homogeneous solution of (5) satisfying

\[
C(x_s)Ψ(t,t_0) = E(x_s)Ψ(t,t_0) \quad (8)
\]

and \(Θ(t)\) is an integral in the range \(t_0, t\), which depends on \(Φ, C\) and each column of \(F\) [11]. If the initial condition \(x_0\) is deterministic, then \(y_0\) is zero. Since the voltage variation can be considered as zero when the signal is not switching for delay calculation, the initial condition for our problem is deterministic. Even if the initial condition \(y_0\) is statistical due to process variations, it can also be represented as a first-order function w.r.t. ξ. Therefore, \(y(t)\) is rewritten as \(Ψ(t)ξ\) in (7) where \(Ψ(t)\) is a \(N_y \times N_p\) matrix.

We obtain \(Ψ(t)\) by substituting (7) into (5):

\[
C(x_s)Ψ(t) = E(x_s)Ψ(t) + F(x_s) \quad (9)
\]

After solving \(x_s\) and \(Ψ(t)\), \(x(t)\) can be obtained based on \(x(t) = x_s(t) + y(t)\) and \(y = Ψ(t)ξ\) in (7):

\[
x(t) = x_s(t) + Ψ(t)ξ \quad (10)
\]

Equation (10) is used to calculate the time-varying moments of the output voltage. The first two central moments and covariance are expressed in (11)-(13), where the correlation coefficient \(ρ\) between every two process variations are included in the \(E\{ξξ^T\}\) calculation. Since both input and output voltages have the same model w.r.t. ξ, the correlations among input and output voltages and the correlations among process variations can be easily considered during moment calculation. For more efficient memory consumption, a smaller number of normal voltage \(x_s(t)\) points and their corresponding coefficients at the node of interest can be saved and propagated.

\[
E\{x(t)\} = x_s(t) \quad (11)
\]

\[
Var\{x(t)\} = Ψ(t)E\{ξξ^T\}Ψ^T(t) \quad (12)
\]

\[
Cov\{x(t_a), x(t_b)\} = Ψ(t_a)E\{ξξ^T\}Ψ^T(t_b) \quad (13)
\]

B. Analysis flow

The Delay distribution analysis procedure is shown in Algorithm 1. The implementation details of steps 1-6 are presented below.

**Step 1.** Initial condition \(x_0\) of every gate is obtained from the data characterized in library according to the switching of nominal input signals (rising, falling or static).

**Step 2.** The nominal waveform \(x_s(t)\) is computed by a method as commonly in CSM-based STA. In our simulation, instead of Newton-Raphson iterations, Broyden’s method is used at each integration step, as it is a better fit for our table-based representation of the SSTM. Additionally, for higher efficiency, we choose linear interpolation based on triangulation [15].

**Step 3.** At every time point, once \(x_s\) is known, \(C, E\) and \(F\) are updated and function (9) can be solved to obtain \(Ψ\). However, the high dimensionality of \(Ψ\) and \(F\) poses an additional difficulty, which is solved in Step 4.

---

\(^1\)Higher accuracy can be obtained by using higher-order models or piece-wise linear simulation method at the cost of complexity.
Step 4. Based on moment matching, (9) is split into \( N_p \) ordinary differential equations (ODEs):
\[
C(x_s)\dot{\Psi}_j(t) = E(x_s)\Psi_j(t) + F_j(x_s) \quad j = 1 : N_p
\]
where \( F_j \) and \( \Psi_j \) are the \( j \)th column of \( F \) and \( \Psi \), respectively. After using a numerical integration method, due to \( x_s \)-dependent coefficients \( C \), \( E \) and \( F_j \), (14) becomes a linear algebraic equation (LAE), which means that the LAE can be solved fast without the necessity of root-finding iterations. Only LU decomposition, and forward and backward substitution are needed to solve the LAE. Additionally, the same coefficients \( C \) and \( E \) of \( N_p \) ODEs in (14) requires LU decomposition only once to solve these \( N_p \) ODEs.

Step 5. The \( k \)th node voltage, which needs to be stored and propagated (denoted as \( v(t) \)), can be expressed as:
\[
v(t) = x_{sk}(t) + \Psi_{(k)}(t)\xi
\]
where \( x_{sk}(t) \) and \( \Psi_{(k)}(t) \) are the \( k \)th element of \( x_s(t) \) and the \( k \)th row of \( \Psi(t) \), respectively.

C. Computing the delay distribution

For timing analysis, the problem of interest is to compute the moments of arrival time, gate delay or in general crossing time. The crossing time \( t_\eta \) is defined as the first time for voltages to cross the threshold voltage \( V_\eta = \eta \% \cdot V_{dd} \). The \textit{cdf} of crossing time is calculated when the nominal voltage is in transition. For a rising transition this is expressed as:
\[
F_n = P(t_\eta \leq t_n) = 1 - P(t_\eta > t_n) = 1 - G_n
\]
\[
G_n = P(v_1 \leq V_n \cap v_2 \leq V_n \cap \cdots \cap v_n \leq V_n)
\]
\[
P(v_n \leq V_n | v_{n-1} \leq V_n, \ldots, v_1 \leq V_n) \cdot G_{n-1}
\]
\[
P(v_n \leq V_n | v_{n-1} \leq V_n) \cdot G_{n-1} (n = 2 : N)
\]
\[
P(v_n \leq V_n | v_{n-1} \leq V_n) = \frac{G_{n-1}}{P(v_{n-1} \leq V_n)}
\]
where \( v_i \) is the voltage of interest at time \( t_i \) and \( F_n \) denotes the \textit{cdf} of crossing time at time \( t_n \). Equation (18) is rewritten in (19) since the voltages are modeled as Markovian processes [3, 12]. Based on (16) to (20) an iteration method is used to calculate the \textit{cdf} of the corresponding crossing time with initial condition \( G_1 = 1 \). Given the moments and covariances calculated in the RDE-based statistical simulator in (11)-(13), the joint probability and single probability in (20) can be obtained easily.

The relationship between the \textit{cdf} \( F(t) \) and the discretized \textit{pdf} \( f(t) \) in our algorithm is illustrated in Fig. 2. To simplify the calculations, the \textit{cdfs} and \textit{pdfs} have these properties: i) \( F = 1 \) if \( F \geq F_{\text{max}} \) and \( F = 0 \) if \( F \leq F_{\text{min}} \). The time \( t_{\text{start}} \) and \( t_{\text{end}} \) correspond to \( F_{\text{min}} \) and \( F_{\text{max}} \) shown in Fig. 2, respectively; ii) \( f(t) \) is calculated during the period \( [t_{\text{start}}, t_{\text{end}}] \), hence the \textit{pdf} has values only on the definite interval \( [t_{\text{start}}, t_{\text{end}}] \). Let \( t_n' = (t_{n-1} + t_n)/2 \), then the discretized \textit{pdf} is approximated by the following method:
\[
f(t_n') = \int_{t_{n-1}}^{t_n} f(t)dt = F(t_n) - F(t_{n-1}) \text{ where } f(t_1) = 0.
\]
The effective \textit{cdf} is defined as the \textit{cdf} within \( [t_{\text{start}}, t_{\text{end}}] \). If the simulation uses a non-uniform time step algorithm, the effective \textit{cdf} needs to be uniformly sampled for \textit{pdf} computation. After uniformly sampling and interpolating from the effective \textit{cdf} with \( N_s \) samples, the \( N_s \times 1 \) time and \textit{cdf} vectors are obtained and denoted as \( T_1 \) and \textit{cdf}_a, respectively. These vectors are used to calculate the \textit{pdf} vector \( \Omega \) with element \( t_k = \text{cdf}_a u_k - \text{cdf}_a u_{k-1} \) \( (\Omega = 0, k = 2 : N_s) \).

The last step is to calculate the moments of crossing time (mean \( \mu \), standard deviation \( \sigma \) and skewness \( \gamma \)). Denoting \( T_1^T \) as the transposition of the column vector \( T_1 \), the calculation method can be formulated as following:
\[
\mu = T_1^T \Omega \quad \sigma = T_1^T \Omega - \mu^2
\]
\[
\gamma = (\Gamma - 3\mu^2 - \mu^3)/(\sigma^3) \quad (\Gamma = T_3 \Omega^T)
\]
The relationships between the elements of \( T_2 \) and \( T_3 \) with \( T_1 \) are \( T_2(k) = T_1^2(k) \) and \( T_3(k) = T_1^3(k) \) for \( k = 1 : N_s \).

The calculation method for a falling transition is similar to the above methods with the only difference in (17) where \( v_i \) is replaced by \( V_{dd} - v_i \). If the waveform is non-monotonic and crosses \( V_i \) multiple times, the method above is used to iteratively find all crossing times.

D. Complexity analysis

As shown in Algorithm 1, the majority of the runtime is consumed in step 2 to calculate the nominal value \( x_s \) and in step 4 to compute the sensitivities \( \Psi \). Therefore, \( T_{\text{STA}} \approx T_{\text{STA}} + T_{\Psi} \). where \( T_{\text{STA}} \) is the runtime of the whole statistical timing analysis algorithm, \( T_{\text{STA}} \) is the runtime of step 2 and \( T_{\Psi} \) is the time of step 3-4. Step 2 can be solved by ViVo-gate model based STA procedures [3]–[6], [8]–[10], and its complexity depends on the gate models used. For the proposed SSTM-based gate models, the method proposed in [8], [9] also can be used. Compared to traditional ViVo-gate model based STA, our statistical timing analysis method requires extra runtime \( T_{\Psi} \).

Step 4 has complexity \( O(N_p) \). There are 5-7 most important process parameters such as length and threshold voltage. Fortunately, the local variations can be collapsed into a much smaller number of variations (or even one variation [9]) after using Principle Component Analysis-like methods [16]. In our method, after using numerical integration the solving procedure of the \( N_p \) ODEs in (14) requires LU decomposition only once. Furthermore, to calculate \( \Psi \), no root-finding iterations are necessary. Therefore, compared to \( T_{\text{STA}}, T_{\Psi} \)
is approximately proportional to \( \frac{N}{N_{iter}} \cdot T_{STA} \) if the average number of iterations at each integration step in step 2 is \( N_{iter} \). This is more efficient than a corner-based method.

IV. CORRELATIONS OF VARIATIONAL WAVEFORMS

During statistical timing analysis, the correlation of signals caused by process variations and path re-convergence should be considered and efficiently simulated. Fig. 3 indicates the delay standard deviation (\( \sigma \)) and delay skewness (\( \gamma \)) of a NAND2 with respect to different correlation coefficient (\( \rho \)) of input arrival times, with different nominal arrival time difference (\( dt \)). \( \rho \) changes from 0 to 0.9. Input signals are simple ramps with the same arrival time variance (\( \sigma_t = 10\text{ps} \)), but different arrival time means (\( \mu_t \)). It should be noted that, when \( \mu_t \)s are far away from each other (\( dt = 6\sigma_t \)), the correlation has significantly less impact on delay distribution and therefore can be ignored.

![Fig. 3. The importance of correlations](image)

If more than one input switches in a multi-input gate, the 50% crossing time cdfs of the switching inputs can be calculated and the corresponding effective cdf range mentioned in section III-C is used. Take NAND2 as a case and denote the effective cdf range of two inputs as \([t_{start1}, t_{end1}]\) and \([t_{start2}, t_{end2}]\), separately. If \( t_{start2} - t_{end1} > \varepsilon \) or \( t_{start1} - t_{end2} > \varepsilon \), the correlation between two inputs will be ignored and the latest/earliest input or inputs will be propagated while the other is assumed static. However, if the effective cdf ranges are overlapping, all stochastic correlated inputs must be considered.

V. EXPERIMENTAL RESULTS

The effectiveness and accuracy of the proposed approach was evaluated on some most commonly used standard cells and ISCAS85 benchmark circuits using the GVT library in the latest Nangate 45nm package [17]. The SSTM is characterized based on the simulated data using a full BSIM4 model, and every gate model is constructed by replacing every transistor in the gate by its corresponding SSTM. The whole algorithm is implemented in Matlab in a computer with single processor.

**SSTM-based deterministic delay calculation for STA:** Since the statistical simulation depends on the nominal value computation (\( x_s \) in (2)), the accuracy of the proposed SSTM-based gate models for deterministic timing analysis (no process variations) is important. It was tested by the minimum-sized standard cells listed in Table I. Every switching input signal is a ramp with input slew varying from 7.5ps to 600ps and the load capacitance changes from 0.40fF to 25.6fF. Both rising and falling inputs are simulated. The \( \mu \) and \( \sigma \) of gate delay relative errors and output slew relative errors after thousands of simulations are listed in the Table I. The scenarios that all input signals switch at the same time are also included in the experiments. The results indicate a high accuracy of deterministic delay and slew calculation by using our SSTM-based gate modeling.

<table>
<thead>
<tr>
<th>Standard cells</th>
<th>delay error</th>
<th>slew error</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV</td>
<td>0.2135%</td>
<td>0.1882%</td>
</tr>
<tr>
<td>NAND2</td>
<td>0.4828%</td>
<td>0.2024%</td>
</tr>
<tr>
<td>NOR2</td>
<td>0.4521%</td>
<td>0.2659%</td>
</tr>
<tr>
<td>AND2</td>
<td>0.7842%</td>
<td>0.0304%</td>
</tr>
<tr>
<td>XOR2</td>
<td>0.0782%</td>
<td>0.7842%</td>
</tr>
<tr>
<td>BUF</td>
<td>0.7165%</td>
<td>0.8048%</td>
</tr>
<tr>
<td>MUX2</td>
<td>0.2346%</td>
<td>0.3633%</td>
</tr>
<tr>
<td>AOR21</td>
<td>0.1722%</td>
<td>0.5285%</td>
</tr>
<tr>
<td>AOI211</td>
<td>0.2109%</td>
<td>0.5285%</td>
</tr>
<tr>
<td>NAND4</td>
<td>0.9568%</td>
<td>1.6800%</td>
</tr>
</tbody>
</table>

**Statistical timing analysis considering MISS:** In order to evaluate the capability of our statistical simulation method for multiple variational inputs, we applied our approach in circuits with up to four inputs. All inputs of every gate are variational with signal correlations and have high possibilities to switch near-simultaneously (MISS). The multi-input cells are NAND2, NOR2, NOR3, NAND3, AOR21, AOI211, AOI22 and NAND4. Every variational input signal is modeled as a ramp signal of 40ps input transition time with voltage variations. The \( \sigma \) of voltages and the arrival time differences among input signals are varied to obtain results at diverse scenarios. The correlation among every two voltage variations varies from 0 to 0.8. All the statistical simulation results are compared to 10K Spectre Monte Carlo (MC) simulations. Fig. 4 illustrates the relative errors of all the experiments. Most of \( \mu \) relative errors are within 1% while AOI211 has over 1% relative errors when correlation coefficient is 0.8 and variance is large. All the \( \sigma \) relative errors are within 6% except two biggest \( \sigma \) cases (6.42% and 6.71%) coming from NAND4 and AOI21 respectively. All of the skewness errors are within 8%. The average \( \mu \), \( \sigma \) and \( \gamma \) relative errors are 0.38%, 2.30% and 2.87% respectively. Fig. 5 shows the discrete pdf with 50 samples and the histogram of MC simulation in Spectre of AOI21. All inputs have the exact same mean value of arrival times (MISS). The discrete pdf was scaled to provide a straightforward shape comparison.

![Fig. 4. All moment percentage relative errors comparison](image)

**Statistical timing analysis with \( L_{eff} \) and \( V_{th} \) variations:** Effective length \( L_{eff} \) and threshold voltage \( V_{th} \) are chosen as the representative process variables, which both have 3\( \sigma \) equal to 20% of the mean value with correlation coefficients of 0, 0.2, 0.5 and 0.8. We firstly applied the proposed method.
to nine common standard cells with different input transitions. Fig. 6 illustrates the average relative errors (absolute values) of $\mu$ and $\sigma$ for nine common standard cells. The worst $\sigma$ relative errors are $-4.03\%$ and $3.04\%$ from AOI211 and XOR2 with falling output respectively. Fig. 7 shows what the variational waveforms look like. The discrete pdf of the 50% crossing time is shown on the upper right corner of Fig. 7.

Secondly, we used the proposed transistor-level statistical timing analysis method for some circuits listed in Table II, where the absolute $\mu$ and $\sigma$ relative errors of delay distribution calculation are included. The gates with more than 3 inputs in C432 and C499 are replaced with several logic gates with no comparison. The correlation coefficient ($\rho$) among process variations are preserved during simulation since the voltages and all elements in gate models have the same model format. Furthermore, the multiple input switching problem is addressed by considering all input signals together for output information. The variational waveforms of the gate output are calculated by RDE-based statistical simulations, which is used for delay distribution calculation. The experiments demonstrate the high accuracy and efficiency of the proposed method for both deterministic delay calculation and statistical timing analysis.

VI. CONCLUSIONS

In this paper, we have presented a new transistor-level gate model based statistical timing analysis method. The gate models are constructed based on statistical simplified transistor models for higher accuracy. Correlations among input signals and among process variations are preserved during simulation since the voltages and all elements in gate models have the same model format. Furthermore, the multiple input switching problem is addressed by considering all input signals together for output information. The variational waveforms of the gate output are calculated by RDE-based statistical simulations, which is used for delay distribution calculation. The experiments demonstrate the high accuracy and efficiency of the proposed method for both deterministic delay calculation and statistical timing analysis.

| TABLE II |
| THE ABSOLUTE VALUES OF DELAY $\mu$ AND $\sigma$ RELATIVE ERRORS (UNIT: %) OF SOME CIRCUITS WITH DIFFERENT CORRELATION COEFFICIENTS $\rho$, COMPARED WITH 10K SPECTRE MC RESULTS |
| $\rho$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| name | 0 | 0.2 | 0.5 | 0.8 |
| Adder | 0.01 | 0.05 | 0.54 | 0.40 | 1.00 | 2.28 | 1.04 | 2.63 |
| C432 | 0.18 | 2.00 | 0.71 | 1.46 | 0.88 | 1.04 | 1.15 | 0.90 |
| C499 | 0.81 | 2.19 | 0.37 | 0.95 | 0.47 | 2.32 | 1.07 | 2.97 |

REFERENCES