Interpreting SSTA Results with Correlation

Zeqin Wu, Philippe Maurine, Nadine Azemard, Gilles R. Ducharme

Interpreting SSTA Results with Correlation

Zeqin Wu[1], Philippe Maurine[1], Nadine Azemard[1], and Gille Ducharme[2]

[1] LIRMM, UMR CNRS/University of Montpellier II, (C5506),
161 rue Ada, 34392 Montpellier, France
{azemard,pmaurine,wu}@lirmm.fr
[2] Dept. Math University of Montpellier II
Place Eugène Bataillon, 34095 Montpellier, France
ducharme@math.univ-montp2.fr

Abstract. ***Statistical Static Timing Analysis*** (SSTA) is becoming necessary; but has not been widely adopted. One of those arguments against the use is that results of SSTA are difficult to make use of for circuit design. In this paper, by introducing conditional moments, we propose a path-based statistical timing approach, which permits us to consider gate topology and switching process induced correlations. With the help of this gate-to-gate delay correlation, differences between results of SSTA and those of ***Worst-case Timing Analysis*** (WTA) are interpreted. Numerical results demonstrate that path delay means and standard deviations estimated by the proposed approach have absolute values of relative errors respectively less than 5% and 10%.

Keywords: Conditional Moment, Worst-case Timing Analysis (WTA), Statistical Static Timing Analysis (SSTA), Gate-to-gate Delay Correlation (GDC).

1 Introduction

Traditional ***Worst-case Timing Analysis*** (WTA) assumes that all physical and environmental parameters are at their worst or best conditions simultaneously. From the point of view of probability theory, this conservative case is next to impossible to appear in reality. Consequently, such an assumption induces pessimism in delay estimation, and thereby in circuit design. As the magnitude of process variations grows, this pessimism increases significantly, leading to the understanding that traditional corner-based design methodologies will not meet the needs of designers in the near future. Thus, ***Statistical Static Timing Analysis*** (SSTA), where process variations and timing characteristics are considered as ***Random Variables*** (RV), has gained favor in the past six years.

The authors of [1]-[2] propose non-linear parametric models handling Gaussian and non-Gaussian process variations, which is a significant progress relative to the linear dependency on Gaussian process parameters presented in [3]-[4]. These approaches are based on first or second order approximation of Taylor expansion, which describes gate-level timing behavior, and is capable of capturing process variations. Then, SUM and MAX operations corresponding to each proposed parametric model are performed on the random timing variables, such as arrival time and gate delay.

J. Monteiro and R. van Leuken (Eds.): PATMOS 2009, LNCS 5953, pp. 16–25, 2010.

It seems that SSTA is becoming a promising alternative by introducing more and more advanced elements from probability theory and statistics, such as Reduced Rank Regression (RRR) [1], Fourier series [2], etc. However, statistical timing techniques have not been widely adopted. First, as suggested in [5], the accuracy of published approaches is not clear due to the fact that ***Monte-Carlo*** (MC) simulations used for validations are based on the same assumptions used in SSTA. Next, the dependency of gate delay on input slope and output load has not received much attention [5]. In fact, most of the proposed approaches either make a worst-case estimate of slope, or propagate the latest arriving slope, each of which can lead to significant error of estimation [6]. At last, it takes time to understand and make use of statistical terms, like correlation, for circuit design.

In this paper, we propose a path-based statistical timing engine to propagate iteratively means and variances of gate delays with the help of conditional moments. These moments conditioning on input slope and output load are pre-characterized by MC simulation, and organized as a tree of lookup tables, called statistical timing library. This timing engine may: a) avoid gate delay modeling errors, b) take into account the effects on gate delay: input pin, output edge, input slope, and output load, and c) deal with a large number of process parameters. Moreover, in a first attempt, we propose an empirical technique to estimate ***Gate-to-gate Delay Correlation*** (GDC). With this statistical term, we explain the incoherence between the critical paths arrangement of SSTA and that of WTA.

Figure 1 gives us an overview on the flow of the approach. First, given a statistical process model, we characterized conditional moments of timing variables with HSPICE [7] under certain power supply voltage (1.1V, 1.2V, 1.3V) and temperature ($-45C^o$, $25C^o$, $125C^o$) conditions. Then statistical timing library is constructed as a tree of lookup tables with the statistical computing tool R [8]. Each of these lookup tables contains conditional moments of timing variables with input slope and output load as indices. In the second step, a certain number of critical paths were extracted from the considered circuits using WTA under the software RTL Compiler [9]. Once the two steps above are finished, we can perform an analysis and generate a statistical timing.

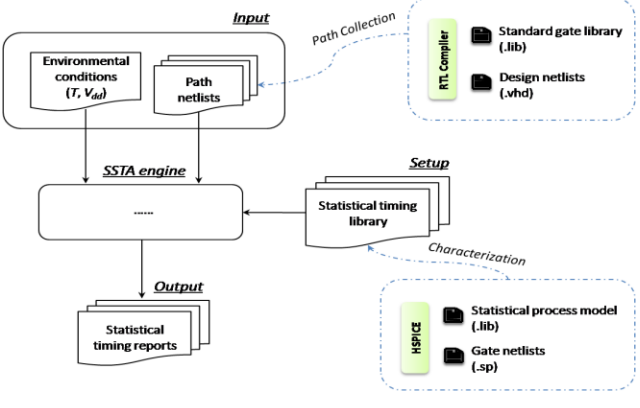

Fig. 1. Flow of the proposed approach

The rest of the paper is organized as follows. Section 2 presents the path-based timing approach, especially the skills to estimate GDC. Section 3 explains why results of SSTA and those of WTA are incoherent. The validation of the approach is given in Section 4. At last, the paper is concluded in Section 5.

2 Statistical Timing Engine

This section introduces the statistical timing engine basing on moment propagation. For gate-level delays, we make no assumption on their distributions, and just propagate means and variances; whereas path delays are assumed to be Gaussian distributed. Such an assumption is somewhat reasonable according to the Central Limit Theorem.

2.1 Conditional Moments

The mean and variance of a random variable, if they exist, are respectively the first and second central moment. A ***conditional moment*** is the moment of one random variable conditioning on another random variable. If X and Y are two random variables, then the ***conditional mean*** $E(X|Y=y)$ is the mean of X given the value $Y=y$. Unlike the conventional mean, which is a constant for a specific probability distribution, $E(X|Y=y)$ is a function of y, that is to say, the conditional mean varies along with the condition Y. Similarly, the ***conditional variance*** $Var(X|Y=y)$ is the variance of X given the value $Y=y$. If Y follows a continuous distribution with ***Probability Density Function*** (PDF) $f(y)$, then we have:

$$\mu_X = E(X) = E[E(X|Y)] = \int E(X|Y=y) \cdot f(y)dy \tag{1}$$

$$\begin{aligned}\sigma_X^2 &= Var(X) = E[Var(X|Y)] + Var[E(X|Y)] \\ &= \int \{Var(X|Y=y) + [E(X|Y=y) - \mu]^2\} \cdot f(y)dy\end{aligned} \tag{2}$$

where μ_X and σ_X^2 are respectively the mean and variance of X.

Equations (1) and (2) give us an alternative way to calculate μ_X and σ_X^2 if these two moments cannot be obtained directly (e.g. the probability distribution of X is unknown). These two equations imply the dependency between X and Y, which permits us to implement the idea of moment propagation.

In (1) and (2), if X, Y represent respectively the output slope τ_{out} and the input slope τ_{in} of a considered gate, given (3) – (4):

$$E(\tau_{out}|\tau_{in}) = a_1 + a_2 \cdot \tau_{in} \tag{3}$$

$$Var(\tau_{out}|\tau_{in}) = b_1 + b_2 \cdot \tau_{in} \tag{4}$$

where a_1, a_2, b_1, b_2 are values to identify. Then, we can compute $\mu_{\tau_{out}}$, $\sigma_{\tau_{out}}^2$ with:

$$\mu_{\tau_{out}} = a_1 + a_2 \cdot \int \tau_{in} \cdot f(\tau_{in})d\tau_{in} = a_1 + a_2 \cdot \mu_{\tau_{in}} \tag{5}$$

$$\begin{aligned}\sigma_{\tau_{out}}^2 &= \int \left[b_3 + b_4 \cdot \tau_{in} + \left(a_2 \cdot \tau_{in} - a_2 \cdot \mu_{\tau_{in}}\right)^2\right] \cdot f(\tau_{in})d\tau_{in} \\ &= \left(b_1 + b_2 \cdot \mu_{\tau_{in}}\right) + \left(a_2 \cdot \sigma_{\tau_{in}}\right)^2\end{aligned} \tag{6}$$

where $f(\tau_{in})$ is the PDF of τ_{in}. Note that in Equations (5) – (6), $f(\tau_{in})$ is not explicitly known, while $\mu_{\tau_{in}}$ and $\sigma^2_{\tau_{in}}$ are required.

2.2 Moment Propagation

This subsection presents the technique to propagate moments of timing variables iteratively along a timing path. First of all, we assume that all timing variables follow continuous distributions.

Let us define the problem of moment propagation: for the considered gate, given mean $\mu_{\tau_{in}}$ and variance $\sigma^2_{\tau_{in}}$ of input slope, and the output load $C_{out} = K$, we expect to get the output slope moments $\mu_{\tau_{out}}$, $\sigma^2_{\tau_{out}}$ and the gate delay moments μ_{gd}, σ^2_{gd}. Note that K represents the nominal value of output load. Its variations have been captured during timing characterization [10]. Besides, only the moments of timing variables instead of distributions are known, i.e. slope and gate delay may follow any distribution.

After the timing characterization presented in [10], we construct the lookup tables in the statistical timing library with the structure as follow: (a) input slope index T_i $(i = 1, \dots, I)$, (b) output load index C_j $(j = 1, \dots, J)$, and (c) lookup values including conditional moments: $E(\tau_{out}|\tau_{in} = T_i, C_{out} = C_j)$, $Var(\tau_{out}|\tau_{in} = T_i, C_{out} = C_j)$, $E(d|\tau_{in} = T_i, C_{out} = C_j)$, and $Var(d|\tau_{in} = T_i, C_{out} = C_j)$.

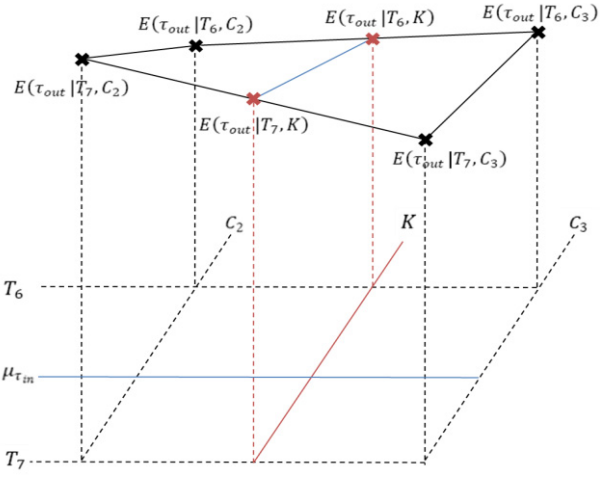

Fig. 2. Estimating $\mu_{\tau_{out}}$ with bilinear interpolation

For simplicity, $E(\tau_{out}|\tau_{in} = T_i, C_{out} = C_j)$ will be denoted as $E(\tau_{out}|T_i, C_j)$ in the rest of this paper, and the other three conditional moments will use the similar notation. Typically, suppose $\mu_{\tau_{in}} \in (T_6, T_7)$ and $K \in (C_2, C_3)$, then as shown in Figure 2, the estimation is done using the bilinear interpolation technique, which is an

extension of linear interpolation for interpolating functions of two variables on a regular grid. First, in terms of the output load $C_{out} = K$, the interpolation gives:

$$E(\tau_{out}|T_i, K) = \frac{K - C_2}{C_3 - C_2} \cdot [E(\tau_{out}|T_i, C_3) - E(\tau_{out}|T_i, C_2)] + E(\tau_{out}|T_i, C_2) \quad (7)$$

$$Var(\tau_{out}|T_i, K) = \frac{K - C_2}{C_3 - C_2} \cdot [Var(\tau_{out}|T_i, C_3) - Var(\tau_{out}|T_i, C_2)] + Var(\tau_{out}|T_i, C_2) \quad (i = 6, 7) \quad (8)$$

With equations (7) – (8), we can further interpolate in the direction of τ_{in} as:

$$E(\tau_{out}|\tau_{in}, K) = \frac{\tau_{in} - T_6}{T_7 - T_6} \cdot [E(\tau_{out}|T_7, K) - E(\tau_{out}|T_6, K)] + E(\tau_{out}|T_6, K) \quad (9)$$

$$Var(\tau_{out}|\tau_{in}, K) = \frac{\tau_{in} - T_6}{T_7 - T_6} \cdot [Var(\tau_{out}|T_7, K) - Var(\tau_{out}|T_6, K)] + Var(\tau_{out}|T_6, K) \quad (10)$$

Combining equations (3) – (4) with (9) – (10), we have:

$$\begin{cases} a_1 = \dfrac{T_7 \cdot E(\tau_{out}|T_6, K) - T_6 \cdot E(\tau_{out}|T_7, K)}{T_7 - T_6} \\ a_2 = \dfrac{E(\tau_{out}|T_7, K) - E(\tau_{out}|T_6, K)}{T_7 - T_6} \\ b_1 = \dfrac{T_7 \cdot Var(\tau_{out}|T_6, K) - T_6 \cdot Var(\tau_{out}|T_7, K)}{T_7 - T_6} \\ b_2 = \dfrac{Var(\tau_{out}|T_7, K) - Var(\tau_{out}|T_6, K)}{T_7 - T_6} \end{cases} \quad (11)$$

Similarly, μ_d and σ_d^2 are estimated by replacing the conditional moments $E(\tau_{out}|T_i, K)$, $Var(\tau_{out}|T_i, K)$ respectively with $E(d|T_i, K)$ and $Var(d|T_i, K)$ in equation (11).

2.3 Gate-to-Gate Delay Correlation

For a timing path of L gates, if the moment propagation technique allows iteratively computing gate delay moments $\mu_{d_l}, \sigma_{d_l}^2, (l = 1, 2, \ldots, L)$, then the path delay d_{path}, which is the sum of all gate delays, has the mean and variance given by:

$$\begin{cases} \mu_{d_{path}} = \sum_{l=1}^{L} \mu_{d_l} \\ \sigma_{d_{path}}^2 = \sum_{l=1}^{L} \sum_{k=1}^{L} \rho_{kl} \cdot \sigma_{d_k} \sigma_{d_l} \end{cases} \quad (12)$$

where ρ_{kl} is the correlation $cor(d_k, d_l)$. Assuming that path delay is a Gaussian RV, then to get the distribution $N\left(\mu_{d_{path}}, \sigma_{d_{path}}^2\right)$, all that remains is to estimate ρ_{kl}.

A common way to estimate ***Gate-to-gate Delay Correlation*** (GDC) is to approximate the dependency of gate delay on process parameters with Taylor expansion, and then to translate the parameter-space correlation to the performance-space correlation. Theoretically, apart from process parameters, all factors that affect gate delay, like gate type, output load, etc., should be considered. Table 1 demonstrates that GDC varies with gate type, output load ($1fF$, $10fF$, $100fF$) and input/output edge ($R \to F$, $F \to R$, $R \to R$, $F \to F$). The CDC coefficients are estimated with data from MC simulations. As shown in Table 1, the effects of gate type and input/output edge on CDC are obvious. In addition, it seems that coefficients are brought down by increasing output load. Thus, GDC is impacted by several parameters that can be classified in four categories: (a) switching induced parameters such as input slope, output load and edge applied on gate inputs; (b) gate topology parameters such as width of transistors, gate polarity (inverting or non-inverting gate) etc; (c) environmental parameters power supply voltage and temperature; and (d) distance separating gates. In the rest of the paper, spatial correlation is neglected because of the lack of data from industry to construct statistical process model.

Table 1. GDC varying with gate type, output load and input/output edge

			F_IVLL		*CTBUFLLP*	
			$10fF$	$10fF$	$10fF$	$10fF$
			$R \to F$	$F \to R$	$R \to R$	$F \to F$
$AN2LLX05$	$1fF$	$R \to R$	0.88	0.93	0.99	0.99
		$F \to F$	0.81	0.92	0.99	0.98
	$10fF$	$R \to R$	0.91	0.82	0.98	0.99
		$F \to F$	0.90	0.86	0.97	0.97
	$100fF$	$R \to R$	0.86	0.58	0.85	0.84
		$F \to F$	0.96	0.59	0.78	0.83
$NR2LLX05$	$1fF$	$R \to F$	0.97	0.76	0.90	0.94
		$F \to R$	0.61	0.97	0.94	0.90
	$10fF$	$R \to F$	0.99	0.75	0.91	0.95
		$F \to R$	0.66	0.99	0.95	0.92
	$100fF$	$R \to F$	0.99	0.62	0.89	0.88
		$F \to R$	0.64	0.99	0.89	0.87

As gate delay depends on a number of factors, which affects GDC as well, we propose a technique to estimate directly GDC instead of translating parameter-space correlation. Suppose that process parameters $X_1, \dots, X_n$ are classified into three groups: $\{X_1^N, \dots, X_{n_1}^N\}$ are parameters describing only N-transistors; $\{X_1^P, \dots, X_{n_2}^P\}$ are those only for P-transistors; and $\{X_1^S, \dots, X_{n_3}^S\}$ characterizing behaviors of both N- and P-transistors. Note that $n = n_1 + n_2 + n_3$. Adopting this classification, each process parameter X_i is further divided into a global component $X_{g,i}$ and a local component $X_{l,i}$, which are independent to each other.

Once this dichotomy achieved, variability of timing metrics is computed for each gate considering separately the three groups of process parameters, and at the same time global and local variations, as illustrated by:

$$\begin{cases} \left(\sigma_d^N\right)^2 = \left(\sigma_{g,d}^N\right)^2 + \left(\sigma_{l,d}^N\right)^2 \\ \left(\sigma_d^P\right)^2 = \left(\sigma_{g,d}^P\right)^2 + \left(\sigma_{l,d}^P\right)^2 \\ \left(\sigma_d^S\right)^2 = \left(\sigma_{g,d}^S\right)^2 + \left(\sigma_{l,d}^S\right)^2 \end{cases} \tag{13}$$

With such definitions, the total variance σ_d^2 of gate delay can be decomposed as:

$$\sigma_d^2 = (\sigma_d^N)^2 + (\sigma_d^P)^2 + (\sigma_d^S)^2 \tag{14}$$

Owing to this decomposition, correlation can then be estimated according to:

$$\rho_{kl} = \frac{cov(d_k, d_l)}{\sigma_{d_k} \cdot \sigma_{d_l}} \tag{15}$$

where

$$cov(d_k, d_l) \approx \sigma_{g,d_k}^N \cdot \sigma_{g,d_l}^N + \sigma_{g,d_k}^P \cdot \sigma_{g,d_l}^P + \sigma_{g,d_k}^S \cdot \sigma_{g,d_l}^S \tag{16}$$

From the above formulas, an immediate drawback appears: $\sigma_{g,d}^N$, $\sigma_{g,d}^P$, $\sigma_{g,d}^S$ must be characterized. However, the characterization step is only a one-off job, i.e. the high time-cost simulation is only needed to build the statistical timing library.

3 Application and Interpretation

In this section, we address the problem on arrangement of critical paths. Given a circuit block and the desired cycled time, we collect the top 100 paths in decreasing order of worst path delays under the 1.1V (supply voltage) and 125C° (temperature) operating conditions. As shown in Figure 3, the dashed line is the result ordered by worst path delays. Next, under the same environmental conditions, for each critical path, we calculate the corresponding statistical 3σ corners with the proposed timing engine, which is plotted with the continuous line in Figure 3.

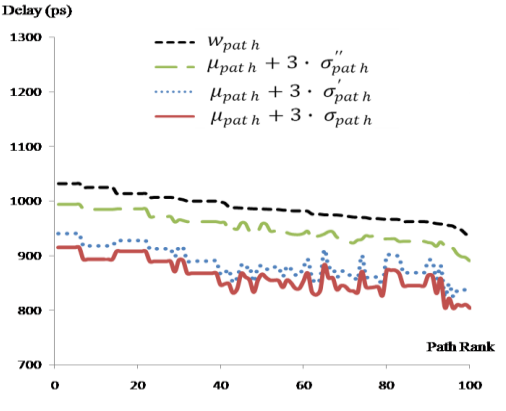

Fig. 3. Arrangement of critical paths

Obviously, the two arrangements obtained respectively by SSTA and WTA are not coherent. To interpret the difference, we suppose a timing path of N gates, each of which has his corresponding gate delay mean μ_i and variance σ_i^2, and define the worst gate delay w_i by:

$$w_i = \mu_i + \theta_i \cdot \sigma_i \qquad (i = 1, \ldots, N) \tag{17}$$

where θ_i are parameters to identify. Then, according to (12) and (17), path delay 3σ corner from SSTA and worst path delay w_{path} using WTA can be decomposed as:

$$\mu_{path} + 3 \cdot \sigma_{path} = \sum_{i=1}^{N} \mu_i + 3 \cdot \sqrt{\sum_{l=1}^{N}\sum_{j=1}^{N} \rho_{ij} \cdot \sigma_i \sigma_j} \tag{18}$$

$$\begin{aligned} w_{path} &= \sum_{l=1}^{N} w_l = \sum_{i=1}^{N} \mu_i + 3 \cdot \left(\sum_{l=1}^{N} \frac{\theta_i \cdot \sigma_i}{3} \right) \\ &= \sum_{i=1}^{N} \mu_i + 3 \cdot \sqrt{\sum_{l=1}^{N}\sum_{j=1}^{N} 1 \cdot \left(\frac{\theta_i \cdot \sigma_i}{3}\right) \cdot \left(\frac{\theta_j \cdot \sigma_j}{3}\right)} \end{aligned} \tag{19}$$

Comparing (18) with (19), we can find that the incoherence between the arrangement of SSTA and that of WTA comes from two factors: (a) GDC coefficients, in other words, $\rho_{ij} \neq 1$ if $i \neq j$ in (18) while the corresponding value in (19) is set to a constant "1"; and (b) standard deviation of gate delay, to be more precise, there exists at least one indicator i so that $\theta_i \cdot \sigma_i / 3 \neq \sigma_i$.

In order to eliminate respectively one of the two factors for more detailed comparison, we compute σ'_{path} and σ''_{path} with:

$$\sigma'_{path} = \sqrt{\sum_{l=1}^{N}\sum_{j=1}^{N} 1 \cdot \sigma_i \sigma_j} \tag{20}$$

$$\sigma''_{path} = \sqrt{\sum_{l=1}^{N}\sum_{j=1}^{N} \rho_{ij} \cdot \left(\frac{\theta_i \cdot \sigma_i}{3}\right) \cdot \left(\frac{\theta_j \cdot \sigma_j}{3}\right)} \tag{21}$$

According to Figure 3, we can conclude that: (a) the violation of path ranks is mainly from the way with which we estimate standard deviation of gate delay, in other words, from the gaps $\sigma_i \cdot (\theta_i/3 - 1)$; (b) it is feasible to attack problems like yield analysis and statistical optimization in terms of gate-level delay correlation. For example, the continuous line will move closer to the dotted line if we can propose techniques to increase ρ_{ij}, such as the use of low process sensitivity gates.

4 Validation

For the validation, we apply the path-based SSTA flow to the ITC99 benchmark circuits implemented respectively in 130nm and 65nm process. Results from the

statistical timing engine were compared to those delivered by MC simulations which are performed using the same statistical process model under the typical operating conditions (1.2V and 25C°).

In Figure 4, points above the 45° straight line indicate that values are overestimated; and those below the line are underestimated. To sum up, for mean of path delay, relative errors $|(\hat{\mu}-\mu)/\mu| \times 100\%$ are less than 5%; and as regards standard deviation, less than 10%. These errors are acceptable in the context of timing analysis. Moreover, most of the standard deviations are a little overestimated, which reduces the probability of the violations of the setup and hold time constraints.

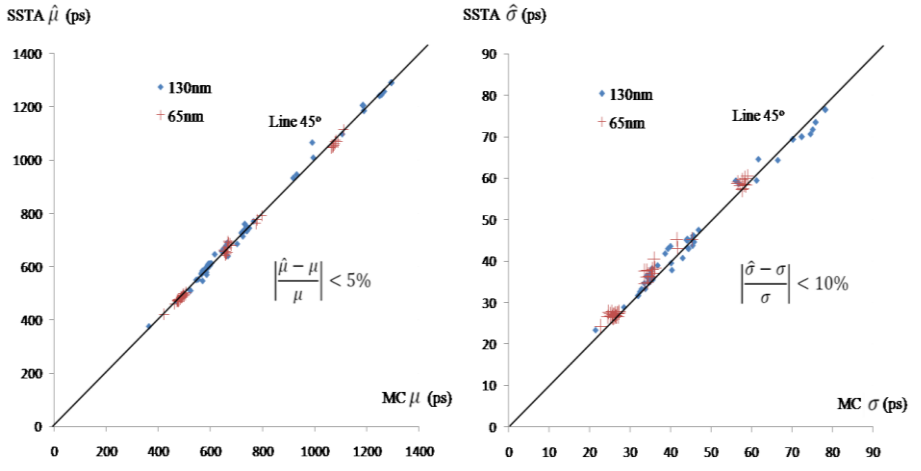

Fig. 4. Validation on estimating path delay probability distribution

In addition to accuracy, Table 2 gives some examples to demonstrate the significant CPU time gain of the SSTA engine compared to MC simulation. The ratio st/et means that the time needed to simulate one path is enough for us to perform SSTA on over 10^5 paths of the same length with the engine.

Table 2. Computational cost of MC simulation and the SSTA engine

path	logical depth	CPU time (s)		st/et (simulation time : st SSTA time : et)
		simulation (1500 runs)	SSTA	
1	5	2794.02	0.02	1.40×10^5
2	10	5245.12	0.03	1.75×10^5
3	15	6914.28	0.06	1.15×10^5
4	20	9881.50	0.08	1.24×10^5
5	25	12020.70	0.11	1.09×10^5

5 Conclusions

In this paper, we present a statistical timing engine considering effects of gate topology and switching induced correlation. A procedure to estimate gate-to-gate

delay correlation along path has been introduced for this purpose. What is more, with the help of this statistical dependency "gate-level correlation", differences between results of SSTA and those of WTA can be reasonably interpreted. The proposed SSTA flow gives us acceptable estimates of path delay distributions with absolute values of relative errors 5% and 10% respectively on mean and on standard deviation.

References

[1] Feng, Z., Li, P., Zhan, Y.: Fast second-order statistical static timing analysis using parameter dimension reduction. In: DAC 2007 (2007)
[2] Cheng, L., Xiong, J., He, L.: Non-linear statistical static timing analysis for non-Gaussian variation sources. In: DAC 2007 (2007)
[3] Chang, H., Sapatnekar, S.S.: Statistical timing analysis considering spatial correlations using a single PERT-like traversal. In: ICCAD 2003 (2003)
[4] Visweswariah, C., Ravindran, K., Kalafala, K.: First-order parameterized block-based statistical timing analysis. In: DAC 2004 (2004)
[5] Ripp, A., et al.: Design for Manufacturability and Yield – influence of process variations in digital, analog and mixed-signal circuit design. In: DATE 2006 (2006)
[6] Ramalingam, A., Singh, A.K., et al.: An Accurate Sparse Matrix Based Framework for Statistical Static Timing Analysis. In: ICCAD 2006 (2006)
[7] http://www.synopsys.com/home.aspx
[8] http://www.r-project.org002F
[9] http://www.cadence.com/
[10] Wu, Z., Maurine, P., Azémard, N., Ducharme, G.: SSTA with Correlations Considering Input Slope and Output Load Variations. In: VLSI-SOC 2008 (2008)