Advances in Computation of the Maximum of a Set of Random Variables

Debjit Sinha, Hai Zhou, and Narendra V. Shenoy†
EECS, Northwestern University, Evanston, IL 60208, USA
{debjit, haizhou}@ece.northwestern.edu
†ATG, Synopsys Inc., Mountain View, CA 94043, USA
narendra.shenoy@synopsys.com

Abstract

This paper quantifies the approximation error in Clark's approach [1] to computing the maximum (max) *of Gaussian random variables; a fundamental operation in statistical timing. We show that a finite Look Up Table can be used to store these errors. Based on the error computations, approaches to different orderings for pair-wise* max *operations on a set of Gaussians are proposed. Experiments show accuracy improvements in the computation of the* max *of multiple Gaussians by up to 50% in comparison to the traditional approach. To the best of our knowledge, this is the first work addressing the mentioned issues.*

1 Introduction

Analytical approaches to statistical static timing analysis have emerged as an active research topic [2–5]. Recent literature consider gate delays as Gaussian random variables since it facilitates fast analytical evaluation. Chang *et al.* propose a statistical timing analysis approach under this assumption which considers spatial correlations [6]. A timing analysis algorithm that accounts for correlations and accommodates dominant interconnect coupling is proposed by Le *et al.* in [7]. A first order incremental block based statistical timing analyzer is presented by Visweswariah *et al.* in [8].

Propagation of Gaussian distributions in block-based statistical timing analysis involves operations like *add* and *max*. It is required that the output of these operations be a Gaussian for further propagation. An *add* operation on Gaussians yields another Gaussian and is accurate. However, the *max* of multiple Gaussians is not a Gaussian, and approximating its distribution with a Gaussian introduces inaccuracies. Clark's approach [1] is used to approximate the *max* of two Gaussians with another Gaussian by matching the first two moments of their distributions. For multiple Gaussians, the *max* operation is performed a pair at a time. Each of these pair-wise operations introduce errors by approximating the resulting distribution with a Gaussian. These approximation errors can propagate and affect accuracy. We observe that the final loss in accuracy in the *max* of multiple Gaussians is dependent on the order in which pair-wise *max* operations are performed. Prior work does not describe the impact of ordering on the inaccuracy of the process.

Our contributions in this paper are summarized as follows.

- We quantify the error in the approximation of the *max* of two arbitrary Gaussians with a Gaussian. The closed form expression for the PDF of the true *max* is derived, and used to develop an analytical expression which quantifies the approximation error.

- We present a transformation to obtain the *max* of two random variables from the *max* of a new pair of derived random variables, parameters of which can be bounded. In addition, we show that approximation error of the *max* operation is an invariant of our transformation.

- We introduce the idea of using a finite Look Up Table (LUT) to store quantified approximation errors in the *max* operation on any Gaussian pair.

- We study the approximation errors as functions of the given Gaussians and propose good orderings for pairwise *max* operations on a given set of Gaussians. The orderings attain to reduce the loss in accuracy, without significant increase in run times.

Due to the symmetry of *max* and *min* operations, the *max* operation is considered in the paper. The approaches presented are extensible for the case of *min* operations. Experiments results show timing estimation accuracy improvements of up to 50% in comparison to the traditional approach. To the best of our knowledge, this is the first work addressing the mentioned issues.

2 Background

Statistical timing analysis involves propagation of delay distributions through the circuit. Statistical modeling is performed to model gate delays as a function of the sources of variation. Based on the work in [6–8], we consider circuit delays as Gaussian random variables. A Gaussian random variable X is formally expressed as $G(\mu_X, \sigma_X)$, with mean μ_X and variance σ_X^2.

The *add* operation on Gaussian variables is performed easily and yields another Gaussian. The *max* operation, on the other hand, is an intricate operation, and for a given set of Gaussians, is performed a pair at a time. We next show Clark's moment matching approach [1] to computing the *max* of two Gaussians X and Y. ρ is used to represent the correlation coefficient between X and Y. We define the following.

$$\phi(x) \triangleq \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2}) \tag{1}$$

$$\Phi(y) \triangleq \int_{-\infty}^{y} \phi(x) dx \tag{2}$$

$$a \triangleq (\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)^{1/2} \tag{3}$$

$$\alpha \triangleq \frac{\mu_X - \mu_Y}{a} \tag{4}$$

The mean μ_Z and variance σ_Z^2 of $Z \triangleq max(X, Y)$ is expressed analytically as follows.

$$\mu_Z = \mu_X \Phi(\alpha) + \mu_Y \Phi(-\alpha) + a\phi(\alpha) \tag{5}$$

$$\sigma_Z^2 = (\sigma_X^2 + \mu_X^2)\Phi(\alpha) + (\sigma_Y^2 + \mu_Y^2)\Phi(-\alpha) + (\mu_X + \mu_Y)a\phi(\alpha) - \mu_Z^2 \tag{6}$$

Z is approximated to a Gaussian variable $Z_G \triangleq G(\mu_Z, \sigma_Z)$ for delay propagation. The first and second order moments of Z are matched to obtain Z_G, while the higher order moments of Z are ignored. This is the first and foremost source of inaccuracy in the approach. The non-linearity of the *max* operation causes Z to have an asymmetric density function. However Z_G has a symmetric density function. We next quantify the error introduced in the above approximation in the following section.
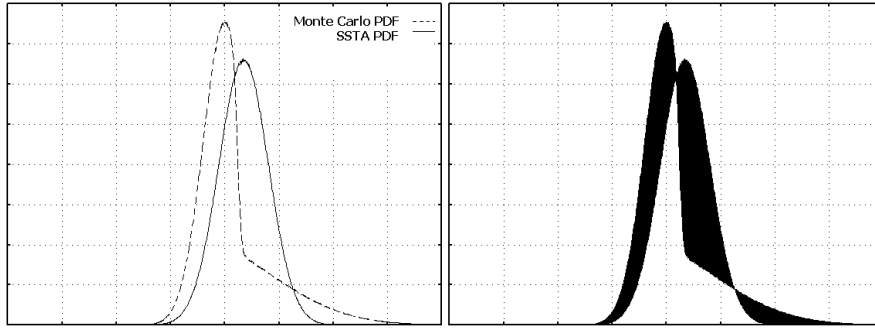
Figure 1: Error Ξ between two random variables with given PDFs is represented by the area of the shaded region

3 Approximation errors in the max operation

3.1 Error definition

A formal comparison of two given distributions requires a metric that quantifies the dis-similarity (or similarity) between them. Given two random variables X, Y, and their probability density functions (PDFs) φ_X, φ_Y respectively, we quantify the dis-similarity or the error Ξ_{XY} between the variables as the total area under the non-overlapped region of their PDFs. This is formally expressed as follows.

$$\Xi_{XY} \triangleq \int_{-\infty}^{\infty} |\varphi_X(t) - \varphi_Y(t)| dt \qquad (7)$$

Figure 1 shows PDFs of two random variables, the dis-similarity between which we attain to quantify. The area of the shaded region represents the error between them. Since the area under each PDF is 1,

$$0 \leq \Xi_{XY} \leq 2.$$

3.2 Error in approximating the max of two Gaussians

We consider the max of two Gaussians $Z \triangleq max(X, Y)$, where $X \triangleq G(\mu_X, \sigma_X)$, $Y \triangleq G(\mu_Y, \sigma_Y)$, and ρ denotes their correlation coefficient. Z is approximated to a Gaussian $Z_G \triangleq G(\mu_Z, \sigma_Z)$, after computing μ_Z and σ_Z from (5)−(6). Based on (7), we formally quantify the error introduced in this approximation as the following.

$$\Xi_{(Z)(Z_G)} \triangleq \int_{-\infty}^{\infty} |\varphi_Z(t) - \varphi_{Z_G}(t)| dt \qquad (8)$$

φ_{Z_G} denotes the PDF of the Gaussian Z_G. Mathematically,

$$\varphi_{Z_G}(t) \triangleq \frac{1}{\sqrt{2\pi}\sigma_Z} e^{-\frac{(t-\mu_Z)^2}{2\sigma_Z^2}} = \frac{1}{\sigma_Z}\phi\left(\frac{t-\mu_Z}{\sigma_Z}\right) \qquad (9)$$

We derive a closed form for $\varphi_Z(t)$ which denotes the true PDF of $max(X, Y)$ as the following. The proof is presented in Appendix I.

$$\varphi_Z(t)$$
$$= \frac{1}{\sigma_Y}\phi\left(\frac{t-\mu_Y}{\sigma_Y}\right)\Phi\left[\frac{\left(\frac{t-\mu_X}{\sigma_X}\right) - \rho\left(\frac{t-\mu_Y}{\sigma_Y}\right)}{(1-\rho^2)^{1/2}}\right] +$$
$$\frac{1}{\sigma_X}\phi\left(\frac{t-\mu_X}{\sigma_X}\right)\Phi\left[\frac{\left(\frac{t-\mu_Y}{\sigma_Y}\right) - \rho\left(\frac{t-\mu_X}{\sigma_X}\right)}{(1-\rho^2)^{1/2}}\right] \qquad (10)$$

$\Xi_{(Z)(Z_G)}$ can now be evaluated from (9) and (10) using numerical integration.

3.3 Errors in canonical form

We consider two properties of a generic max operation.

1. Scaling property

$$max(\lambda X, \lambda Y) = \lambda \cdot max(X, Y) \quad \forall \lambda \geq 0$$

2. Shift-invariance property

$$max(X + \theta, Y + \theta) = max(X, Y) + \theta$$

Consider Gaussians $X \triangleq G(\mu_X, \sigma_X)$ and $Y \triangleq G(\mu_Y, \sigma_Y)$, with correlation coefficient ρ. Without any loss of generality, we assume $\sigma_X \geq \sigma_Y$. Application of the above properties on $max(X, Y)$ (shifting by μ_X and subsequent scaling by σ_X) results in the following.

$$max(X, Y) = \mu_X + \sigma_X \cdot max(X', Y') \qquad (11)$$

where,

$$X' \triangleq G(\mu_{X'}, \sigma_{X'}) = \frac{X - \mu_X}{\sigma_X} = G(0, 1)$$
$$Y' \triangleq G(\mu_{Y'}, \sigma_{Y'}) = \frac{Y - \mu_X}{\sigma_X} = G\left(\frac{\mu_Y - \mu_X}{\sigma_X}, \frac{\sigma_Y}{\sigma_X}\right)$$

ρ' denotes the correlation coefficient between X' and Y'. Since $\sigma_{Y'} = \sigma_Y/\sigma_X$ and $\sigma_X \geq \sigma_Y$, we have the following.

$$0 \leq \sigma_{Y'} \leq 1 \qquad (12)$$

Lemma 1 $\rho' = \rho$

Proof: The covariance (cov) of two Gaussians is independent of their means μ and is directly proportional to their standard deviations σ. Since $\sigma_{X'} = \frac{\sigma_X}{\sigma_X}$ and $\sigma_{Y'} = \frac{\sigma_Y}{\sigma_X}$, we have $cov(X', Y') = \frac{cov(X,Y)}{\sigma_X^2}$. From definition, $\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$.

$$\rho' = \frac{cov(X', Y')}{\sigma_{X'}\sigma_{Y'}} = \frac{\frac{cov(X,Y)}{\sigma_X^2}}{1 \cdot \frac{\sigma_Y}{\sigma_X}} = \rho$$

□

Based on the definitions in (1)-(4), we define the following.

$$a' \triangleq \left(\sigma_{X'}^2 + \sigma_{Y'}^2 - 2\rho'\sigma_{X'}\sigma_{Y'}\right)^{1/2} \qquad (13)$$
$$\alpha' \triangleq \frac{\mu_{X'} - \mu_{Y'}}{a'} \qquad (14)$$

Lemma 2 $\alpha' = \alpha$

Proof: From (3) and (13), we have $a' = \frac{a}{\sigma_X}$. Thus,

$$\alpha' = \frac{\mu_{X'} - \mu_{Y'}}{a'} = \frac{-\mu_{Y'}}{a'} = \frac{-\frac{\mu_Y - \mu_X}{\sigma_X}}{\frac{a}{\sigma_X}} = \alpha$$
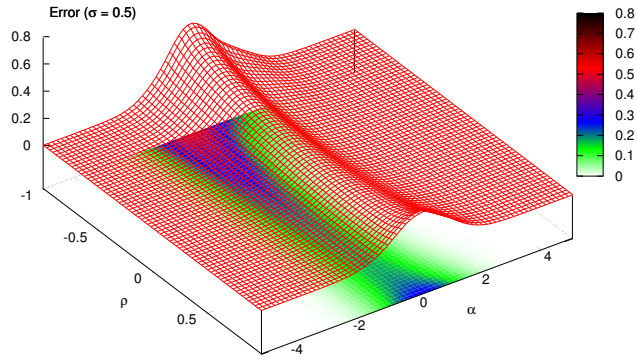
□

Figure 2: $\Xi_{(Z')(Z'_G)}$ as a function of ρ and α ($\sigma_{Y'} = 0.5$)

Figure 3: $\Xi_{(Z')(Z'_G)}$ as a function of $\sigma_{Y'}$ and ρ ($\alpha = 1.0$)

We denote the error in approximating $Z' \triangleq max(X', Y')$ with a Gaussian $Z'_G = G(\mu_{Z'}, \sigma_{Z'})$ to be $\Xi_{(Z')(Z'_G)}$ and prove that it is exactly equal to the error $\Xi_{(Z)(Z_G)}$ in approximating $Z = max(X, Y)$ with Gaussian $Z_G = G(\mu_Z, \sigma_Z)$. The proof is presented in Appendix II. Mathematically, we prove the following.

$$\Xi_{(Z')(Z'_G)} = \Xi_{(Z)(Z_G)} \qquad (15)$$

The approximation error in the *max* of any two Gaussians can thus be estimated from the approximation error in the *max* of the derived Gaussians, one of which is the unit normal Gaussian. The error is therefore a function of $\mu_{Y'}$, $\sigma_{Y'}$ and $\rho'(=\rho)$. Since $\alpha'(=\alpha)$ is a function of $\mu_{Y'}$, the error can be expressed as a function of α, σ'_Y and ρ. It is known that $\phi(\alpha) \approx 0$ for $|\alpha| \geq 4$, $\Phi(\alpha) \approx 0$ for $\alpha \leq 4$, and $\Phi(\alpha) \approx 1$ for $\alpha \geq 4$. Consequently, for $|\alpha| \geq 4$, $max(X, Y)$ almost identically resembles the dominating Gaussian [8]. The approximation error in this case is negligible. Thus, the region of interest for the parameters that affect the approximation error is bounded by the following.

$$
\begin{array}{lll}
-4 \leq & \alpha & \leq 4 \\
0 \leq & \sigma_{Y'} & \leq 1 \quad \text{from (12)} \\
-1 \leq & \rho & \leq 1 \quad \text{from definition}
\end{array}
$$

Experiments are performed to study the dependence of the approximation error $\Xi_{(Z')(Z'_G)}$ on the above parameters. It is observed that $\Xi_{(Z')(Z'_G)}$ decreases when either of the Gaussians dominate the other, *i.e.*, $|\alpha| \geq 3$ and increases for Gaussians that contribute almost equally to the *max i.e.*, α in the neighborhood of 0. $\Xi_{(Z')(Z'_G)}$ is found to increase with decreasing $\sigma_{Y'}$ and is convex with respect to the correlation coefficient. Figures 2−4 show the surface plots of $\Xi_{(Z')(Z'_G)}$ as functions of α, $\sigma_{Y'}$ and ρ. The presented plots reveal that while the *max* of two Gaussians can be very well approximated with a Gaussian in some cases, the approximation in other cases yields large errors.

3.4 Look Up Table for error storage

The computation of the mean and variance of the *max* of two Gaussians involve the evaluation of a definite integral $\Phi(\alpha)$ and an exponential $\phi(\alpha)$. Numerical computations for their accurate estimation is CPU expensive. We consider the
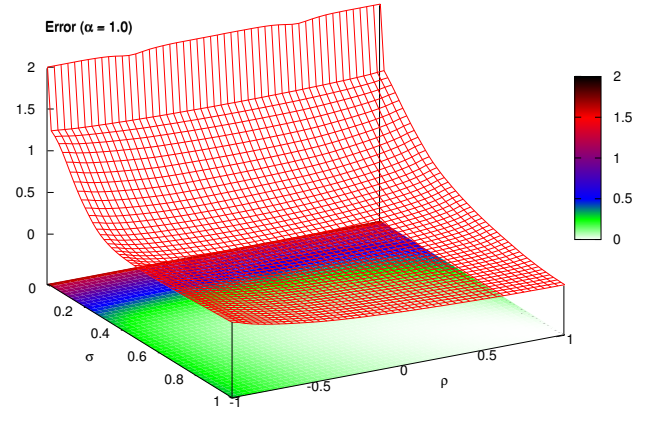

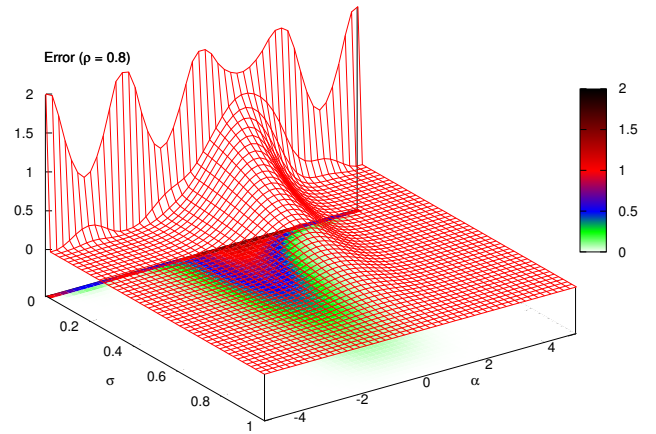Figure 4: $\Xi_{(Z')(Z'_G)}$ as a function of $\sigma_{Y'}$ and α ($\rho = 0.8$)

infinite Taylor series expansion [9] about a point k of $\Phi(\alpha)$.

$$\Phi(\alpha) = \Phi(k) + \ldots \frac{\Phi^{(n)}(k)}{n!}(\alpha - k)^n \ldots \qquad (16)$$

$\Phi^{(n)}(k)$ and $\phi^{(n)}(k)$ represent the nth derivatives of $\Phi(k)$ and $\phi(k)$ respectively. It is observed that the nth derivative of $\Phi(\alpha)$ (and $\phi(\alpha)$), for any $n > 0$ is a product of $\phi(\alpha)$ and a polynomial in α. Mathematically, we have the following[1].

$$\Phi^{(n+1)}(\alpha) = \phi^{(n)}(\alpha) = \phi(\alpha)(-1)^n \, n! \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^i \, \alpha^{n-2i}}{2^i \, i! \, (n-2i)!} \qquad (17)$$

We use the above Taylor series expansion to compute $\Phi(\alpha)$ and $\phi(\alpha)$ in the region $|\alpha| < 4$. We propose to pre-compute $\Phi(k)$ and $\phi(k)$ for multiple values of $k \in [0, 4)^2$, and store the values in two Look Up Tables (LUTs). Thereafter, $\Phi(\alpha)$ ($\phi(\alpha)$) is computed by a table lookup on the closest k in the vicinity of α to obtain $\Phi(k)$ ($\phi(k)$), followed by a finite Taylor

---

[1] Please contact the authors for a formal proof.
[2] It is known that $\Phi(-k) = 1 - \Phi(k)$ and $\phi(-k) = \phi(k)$

series expansion about k. For a uniformly sampled LUT with step-size p, it is observed that a very high degree of accuracy can be obtained by expanding few (typically 3 to 4) terms of the Taylor series expansion. We can obtain any desired accuracy in the computation of $\Phi(\alpha)$ and $\phi(\alpha)$ by either decreasing p while keeping n constant or by increasing n for a given p.

We extend this idea to using LUTs for storing quantified approximation errors next. Given that the approximation error for any two Gaussians is a function of three bounded parameters for all practical purposes, we accurately evaluate approximation errors at discrete points in the bounded space of these parameters. These evaluated errors are stored in a three dimensional finite LUT. Since time is not a constraint in construction of the table, the LUT can be created with as much accuracy as desired. The error estimation for a given point is evaluated from $(8)-(10)$. Approximation error estimation for a max operation is now performed very efficiently by a simple transformation to evaluate $\sigma'_Y = \frac{\sigma_Y}{\sigma_X}$ and a subsequent table lookup.

4 Error minimization problem

We consider the max operation on n given Gaussian random variables $X_0, X_1, \ldots, X_{n-1}$, such that

$$X_M \triangleq max(X_0, X_1, \ldots, X_{n-1}).$$

Pair-wise max operations are performed on the given Gaussians to yield a Gaussian $X_G = G(\mu_{X_G}, \sigma_{X_G})$, which is used to approximate X_M. The loss in accuracy of the final result is dependent on the ordering of the pair-wise max operations. This is because the introduced inaccuracy for each pair-wise max operation is a function of the Gaussian parameters of the pair themselves and can accumulate or get reduced. The max operation on n Gaussians is analogous to the construction of a binary tree with n leaves such that each internal node computes the max of its two children. We refer to this tree as a *Max Binary Tree* (MBT) in the rest of this paper.

Given n Gaussians, the max (X_M) of which we want to estimate and approximate with X_G, the Error minimization problem is to create an MBT that yields some X_G at its root such that $\Xi_{X_M X_G}$ is minimized.

According to Knuth [10], the total number of different *labeled oriented* binary trees with n leaves is $\binom{2n-1}{n-1}(2n-2)!/2^{n-1}$. In an MBT, only the leaves are labeled. Therefore, the total number of different MBTs is

$$\frac{\binom{2n-1}{n-1}(2n-2)!}{2^{n-1}(n-1)!} > (2n-1)^{n-1}.$$

Thus an exhaustive enumeration is prohibitive in solving this problem. Consequently, we consider good MBT construction approaches for error reduction.

5 Intelligent Max Binary Tree construction approaches

In this section, we present novel approaches for constructing good MBTs based on the study in the previous sections. We assume that a max operation takes $\Theta(1)$ time in complexity analysis.

5.1 Simple Max Binary Tree

This approach constructs the MBT as a skewed binary tree. A max is performed on two of the n given Gaussians to yield a new approximated Gaussian. Another max operation follows, which evaluates the max of the max obtained in the previous stage and one of the remaining $(n-1)$ given Gaussians. This process is repeated $(n-1)$ times to obtain a Gaussian, which is used to approximate the max of the n given Gaussians. The complexity of this approach is $\Theta(n)$.

5.2 Partition Max Binary Tree

The Partition MBT approach attains to reduce the depth of approximation errors accumulated in the $(n-1)$ stages of the previous approach by constructing a balanced binary tree. The given set of Gaussians is randomly partitioned into two subsets. The subsets are then further bi-partitioned, and the process is done recursively, until the subset contains no more than two Gaussians. A max operation is then performed on the Gaussians in each subset. The results are now propagated backward to evaluate the max of these values bottom up. The MBT formed using this approach is balanced, and reduces the depth of accumulation of approximation errors from $(n-1)$ to $(log\ n)$. The complexity of this approach is maintained at $\Theta(n)$.

5.3 Greedy Max Binary Tree

This MBT construction method involves a greedy approach to reduce approximation errors. Based on the study in the previous sections, the Greedy MBT approach iteratively computes the max of two Gaussians from the given set, such that the approximation error for that pair is the least in comparison to all other pairs. The computed max is then returned to the original set and the process continues for $(n-1)$ similar iterations. The method is analogous to a graph reduction problem. Consider a fully connected graph with n nodes, each representing a given Gaussian. Edges of the graph contain weights that denote the approximation error in the max operation of the pair of nodes it joins. Adjacent nodes of the edge with the least weight are combined using a max operation. The corresponding nodes are combined into one and edge weights are incrementally recomputed. This process is repeated $(n-1)$ times, after which the graph is left with a single node, which contains the approximated max of the given Gaussians. A LUT is used to evaluate the edge weights in the graph. An alternate metric for any edge weight could be the skewness of the max of its adjacent nodes. The Greedy MBT approach reduces given identical Gaussians (those having same means, variances and $\rho = 1$) to a single one. The complexity of this approach is $\Theta(n^3)$.

5.4 Cluster Max Binary Tree

The Cluster MBT approach is constructed as a combination of the Partition MBT and the Greedy MBT approaches. The max operation is performed in a greedy way on a Gaussian pair that yields the minimal error in approximation among all other pairs. However, the computed max is not sent back to the set of the given Gaussians as in the previous approach. A new pair is selected from the set of given Gaussians for the max operation iteratively till at most one Gaussian is left. The process restarts with the computed max distributions as the given set of Gaussians henceforth. The approach ensures that the constructed MBT is balanced and tries to reduce accumulation of approximation errors by constraining the maximum depth of the tree to $(log\ n)$. This approach reduces identical Gaussians to fewer ones, but is guaranteed to reducing them to one only when the number of identical Gaussians is 2^k for some positive integral value of k. The complexity of this approach is $\Theta(n^2)$.

6 Experimental results

We present experimental results of the proposed MBT construction approaches in this section. We construct a LUT to store approximation errors as presented in Section 3.4. Experimental results presented are for a LUT having 2×10^6 entries ($100 \times 100 \times 200$ for $\sigma_{Y'}$, ρ and α respectively). The simulations to generate the LUT take 4 hours on a Pentium 2.4GHz machine, with 1GB RAM. The Greedy and the Cluster MBT approaches use the LUT to pick Gaussian pairs. The standard deviation of each Gaussian has been constrained to at most 20% of its mean value.

We consider the *max* operation on randomly generated sets of 3 to 100 Gaussians and compare the error of the *max* obtained in different approaches with the distribution obtained from Monte Carlo simulations, which we assume golden. 1000 runs are performed for each set to obtain an average. The obtained experimental results reveal that the proposed Partition MBT, Greedy MBT, and the Cluster MBT approaches reduce the loss in accuracy of Ξ by up to 24% with respect to the results obtained from the Simple MBT approach.

In statistical timing, we are concerned about the estimation accuracy of specific probability points in the CDF. We define a probability point $V_{Pr=p}$ for a random variable X as

$$V_{Pr=p} \triangleq x : (Pr(X \leq x) = p), \quad p \in [0,1].$$

We compare accuracy gains obtained from the proposed approaches with respect to the Simple MBT approach. The absolute value of the difference in $V_{Pr=p}$ points obtained from the CDF of the constructed MBT and the Monte Carlo (MC) CDF is normalized by $V_{Pr=p}^{MC}$ and multiplied by 100 to denote the error percent. The error percent of the Simple MBT approach is used as a reference for comparisons. For each of our proposed approaches, we denote the value by which the error percent of the Simple MBT method exceeds the error percent of the proposed approach as the *%Gain*. Given a probability point p, the *%Gain* obtained in approach A (could be Partition, Cluster or Greedy) is formally defined as the following.

$$\%Gain \triangleq \tag{18}$$
$$(\frac{|V_{Pr=p}^{Simple} - V_{Pr=p}^{MC}|}{V_{Pr=p}^{MC}} \times 100) - (\frac{|V_{Pr=p}^{A} - V_{Pr=p}^{MC}|}{V_{Pr=p}^{MC}} \times 100)$$

The *%Gain* reflects the absolute value of the gain in accuracy and does not reflect the ratio of the gain in accuracy of an approach over the other. Tables 1−3 present *%Gains* of the proposed approaches for the probability points 0.5 (mean), 0.95 and 0.998 respectively. We present the number of Gaussians (N) in the given set, the average *%Gain* obtained (*Gain*) and the maximum *%Gain* (*MaxG*) obtained in the simulations. Table 4 presents the %Gain obtained in the accuracy of estimating the variance of the *max*.

The proposed approaches are found to improve *%Gain*. Since the average error percent in estimation of the $V_{Pr=p}$ points in the Simple MBT approach is experimentally found to be less than 2%, the *%Gain* values obtained are relatively significant. For example a *%Gain* of 0.5% implies a $\frac{0.5}{2} \times 100 = 25\%$ improvement in relative accuracy. The proposed approaches are heuristics and do not guarantee optimality. The Partition MBT approach performs better than the Simple MBT approach on the average and indicates that the depth of cumulative error accumulation causes a difference. The proposed approaches that use the LUT show absolute maximal improvements by up to 5.2% in improving the accuracy of estimation of the critical probability points in the CDF and by up to 50% in estimation of the variance of the *max*. On the average, we find the Greedy approach performs the best. Run times are comparable for sets having up to 30 Gaussians. For the *max* of 50 Gaussians, run times for the Simple, Partition, Cluster and Greedy MBT approaches are found to be 0.00007 seconds, 0.0002 seconds, 0.003 seconds and 0.29 seconds respectively. Though the Cluster and Greedy MBT approaches are computationally more expensive than the Simple MBT approach, given that we intend to use this method for computing the *max* of a finite number of delay distributions in statistical timing, the runtime overheads are acceptable. All experiments are performed on a Pentium 2.4GHz, 1GB RAM machine, running Red Hat Linux 9.0.

Table 1: % Accuracy gain results in $V_{Pr=0.5}(Mean)$

N	Partition		Cluster		Greedy	
	Gain	MaxG	Gain	MaxG	Gain	MaxG
3	0.00	0.5	0.01	0.5	0.01	0.5
5	0.00	0.5	0.02	1.0	0.02	1.0
7	0.01	0.8	0.03	1.0	0.03	1.0
9	0.03	1.1	0.06	1.1	0.06	1.1
12	0.05	0.8	0.07	1.1	0.07	1.4
15	0.08	0.9	0.11	1.2	0.09	1.4
20	0.14	0.9	0.18	1.3	0.16	1.5
30	0.28	1.1	0.35	1.7	0.31	1.6
50	0.55	1.3	0.70	2.9	0.58	2.9
100	1.04	1.9	1.29	2.7	1.12	2.5

Table 2: % Accuracy gain results in $V_{Pr=0.95}$

N	Partition		Cluster		Greedy	
	Gain	MaxG	Gain	MaxG	Gain	MaxG
3	0.02	2.0	0.05	2.0	0.05	2.0
5	0.01	3.1	0.08	3.2	0.09	3.2
7	0.04	2.2	0.07	2.3	0.15	2.7
9	0.03	2.8	0.09	2.3	0.20	2.3
12	0.02	2.7	0.08	2.5	0.23	3.0
15	0.04	2.3	0.14	2.3	0.24	3.1
20	0.05	2.1	0.15	2.8	0.26	2.8
30	0.09	2.8	0.35	3.1	0.31	3.1
50	0.28	3.2	0.33	2.9	0.40	3.2
100	0.57	3.1	0.62	3.4	0.68	3.2

Table 3: % Accuracy gain results in $V_{Pr=0.998}$

N	Partition		Cluster		Greedy	
	Gain	MaxG	Gain	MaxG	Gain	MaxG
3	0.01	2.7	0.07	2.7	0.07	2.7
5	0.01	3.8	0.17	3.8	0.18	3.8
7	0.00	3.4	0.10	3.9	0.23	3.8
9	0.06	4.5	0.28	3.6	0.31	4.0
12	0.06	3.7	0.16	4.6	0.31	3.6
15	0.08	3.6	0.14	3.6	0.38	4.3
20	0.10	3.0	0.25	3.6	0.41	4.7
30	0.12	3.3	0.27	4.1	0.50	5.2
50	0.04	3.4	0.23	4.0	0.37	4.3
100	0.08	3.5	0.17	4.1	0.34	3.8

Table 4: % Accuracy gain results in variance estimation

N	Partition		Cluster		Greedy	
	Gain	MaxG	Gain	MaxG	Gain	MaxG
3	0.09	42.4	0.54	41.4	0.54	41.4
7	0.08	28.8	0.86	31.6	1.87	31.6
9	0.43	35.7	2.32	27.7	2.54	30.6
12	0.53	36.0	1.35	50.0	2.42	33.1
15	0.68	33.5	1.22	27.6	3.05	34.5
20	0.78	22.6	1.88	33.4	3.06	32.5
30	0.67	24.9	2.04	36.6	3.63	39.8
50	0.24	29.4	2.31	33.9	2.92	32.3
100	0.00	23.3	0.92	28.5	1.50	26.6

7 Conclusions

In this paper, we quantify the approximation error in Clark's approach [1] to computing the *max* of Gaussians. We propose approaches to different orderings for pair-wise *max* operations on a set of Gaussians based on error computations. Prior research [11] has shown that the average errors in estimating the mean and standard deviation of a circuit's arrival time distribution in comparison to Monte Carlo simulations are $\approx 1.8\%$ and 13.7% respectively. Similar numbers are shown in [6]. The proposed approaches significantly improve the accuracy in variance estimation on the average, errors in which could be large otherwise. They also improve the estimation accuracy of specific probability points, which could have significant errors (about 4.56% on the average for $V_{Pr=0.99}$ [6]).

We believe that the proposed approaches would increase the accuracy in the estimation of node and edge criticalities [8], and would thereby guide statistical timing optimization better. In addition, expressions for the CDF and PDF of the true *max* of two Gaussians would help in accurate yield estimation, when considering both timing and power [11].

Appendix I

The joint probability density function $\varphi(x,y)$ for Gaussians $X \triangleq G(\mu_X, \sigma_X)$ and $Y \triangleq G(\mu_Y, \sigma_Y)$ is defined as follows.

$$\varphi(x,y) \triangleq \frac{e^{\left\{\frac{-1}{2(1-\rho^2)}\left[(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2\right]\right\}}}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}}$$

• Expression for the CDF of Z

The CDF $\Psi(t)$ of Z is defined as the probability that Z is at most t. Since $Z = max(X,Y)$, $\Psi(t)$ is given by the probability that both X and Y are not more than t. This is evaluated by integrating the joint probability density function $\varphi(x,y)$ in the region where both $X \leq t$ and $Y \leq t$. Thus,

$$\Psi_Z(t) \triangleq Pr(Z \leq t) = \int_{-\infty}^{t}\int_{-\infty}^{t} \varphi(x,y)\ dx\ dy$$

$$= \iint_{-\infty}^{t}\int^{t} \frac{e^{\frac{-1}{2(1-\rho^2)}\left[(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2\right]}}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} dx\ dy$$

$$= \iint_{-\infty}^{t}\int^{t} \frac{e^{\frac{-1}{2(1-\rho^2)}\left[(\frac{x-\mu_X}{\sigma_X}) - \rho(\frac{y-\mu_Y}{\sigma_Y})\right]^2} e^{\frac{-1}{2}\left[(\frac{y-\mu_Y}{\sigma_Y})^2\right]}}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}}\ dx\ dy$$

$$= \frac{\int_{-\infty}^{t}\phi(\frac{y-\mu_Y}{\sigma_Y})\int_{-\infty}^{t}\phi(\frac{[\frac{x-\mu_X}{\sigma_X}] - \rho[\frac{y-\mu_Y}{\sigma_Y}]}{[1-\rho^2]^{1/2}})\ dx\ dy}{\sigma_X\sigma_Y(1-\rho^2)^{1/2}}$$

$$= \frac{1}{\sigma_Y}\int_{-\infty}^{t}\phi(\frac{y-\mu_Y}{\sigma_Y})\Phi\left[\frac{(\frac{t-\mu_X}{\sigma_X}) - \rho(\frac{y-\mu_Y}{\sigma_Y})}{(1-\rho^2)^{1/2}}\right]\ dy.$$

• Closed form expression for the PDF of Z

We next derive a closed form expression for the PDF $\varphi_Z(t)$, which is defined as the derivative of $\Psi(t)$ with respect to t.

$$\varphi_Z(t) \triangleq \frac{d}{dt}[\Psi_Z(t)] = \frac{d}{dt}\left[\int_{-\infty}^{t}\int_{-\infty}^{t} \varphi(x,y)\ dx\ dy\right]$$

$$= \int_{-\infty}^{t}\varphi(x,t)\ dx + \int_{-\infty}^{t}\varphi(t,y)\ dy$$

$$= \frac{1}{\sigma_Y}\phi(\frac{t-\mu_Y}{\sigma_Y})\Phi\left[\frac{(\frac{t-\mu_X}{\sigma_X}) - \rho(\frac{t-\mu_Y}{\sigma_Y})}{(1-\rho^2)^{1/2}}\right]$$
$$+ \frac{1}{\sigma_X}\phi(\frac{t-\mu_X}{\sigma_X})\Phi\left[\frac{(\frac{t-\mu_Y}{\sigma_Y}) - \rho(\frac{t-\mu_X}{\sigma_X})}{(1-\rho^2)^{1/2}}\right]$$

Appendix II

We consider approximating Z with a Gaussian $Z_G \triangleq G(\mu_Z, \sigma_Z)$ and approximating $Z' \triangleq max(X', Y')$ with a Gaussian $Z'_G \triangleq G(\mu_{Z'}, \sigma_{Z'})$. $\varphi_Z(t)$ and $\varphi_{Z_G}(t)$ denote PDFs as defined earlier and $\varphi_{Z'}(t)$ and $\varphi_{Z'_G}(t)$ denote the PDFs of Z' and Z'_G respectively. From (9) and (10), we have the following.

$$\varphi_{Z'_G}(t) \triangleq \frac{1}{\sqrt{2\pi}\sigma_{Z'}}e^{-\frac{(t-\mu_{Z'})^2}{2\sigma_{Z'}{}^2}} \qquad (19)$$

$$\varphi_{Z'}(t) \triangleq \frac{1}{\sigma_{Y'}}\phi(\frac{t-\mu_{Y'}}{\sigma_{Y'}})\Phi\left[\frac{t - \rho(\frac{t-\mu_{Y'}}{\sigma_{Y'}})}{(1-\rho^2)^{1/2}}\right]$$

$$+ \phi(t)\Phi\left[\frac{(\frac{t-\mu_{Y'}}{\sigma_{Y'}}) - \rho t}{(1-\rho^2)^{1/2}}\right] \qquad (20)$$

Lemma 3 $\mu_Z = \mu_X + \sigma_X\mu_{Z'}\ and\ \sigma_Z = \sigma_X\sigma_{Z'}$

Proof: The results are trivially derived from (5)-(6). ☐

Lemma 4 $\varphi_{Z_G}(t) = \frac{1}{\sigma_X}\varphi_{Z'_G}(\frac{t-\mu_X}{\sigma_X})$

Proof: The result is immediate from a simple substitution in (19) and from Lemma 3. ☐

Lemma 5 $\varphi_Z(t) = \frac{1}{\sigma_X}\varphi_{Z'}(\frac{t-\mu_X}{\sigma_X})$

Proof: The result is immediate from a simple substitution in (20) which yields (10). ☐

We next prove that the error in approximating Z with Z_G is the same as the error in approximating Z with Z'_G.

Lemma 6 $\Xi_{(Z')(Z'_G)} = \Xi_{(Z)(Z_G)}$

Proof: From our error definition in (8), and from the results obtained above, we have the following.

$$\Xi_{(Z')(Z'_G)}$$
$$= \int_{-\infty}^{\infty}|\varphi_{Z'}(t) - \varphi_{Z'_G}(t)|\ dt$$
$$= \int_{-\infty}^{\infty}|\varphi_{Z'}(\frac{t-\mu_X}{\sigma_X}) - \varphi_{Z'_G}(\frac{t-\mu_X}{\sigma_X})|\ d(\frac{t-\mu_X}{\sigma_X})$$
$$= \sigma_X\int_{-\infty}^{\infty}|\varphi_Z(t) - \varphi_{Z_G}(t)|\ dt\ \frac{1}{\sigma_X}\ =\ \Xi_{(Z)(Z_G)}$$

☐

References

[1] C. E. Clark, "The greatest of a finite set of random variables," in *Operations Research, Vol. 9, No. 2 (Mar - Apr)*, 1961, pp. 145–162.

[2] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," in *Proc. Intl. Conf. on Computer-Aided Design*, 2003, pp. 607–614.

[3] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proc. Intl. Conf. on Computer-Aided Design*, 2003, pp. 900–907.

[4] S. Bhardwaj, S. Vrudhula, and D. Blaauw, "TAU: Timing analysis under uncertainty," in *Proc. Intl. Conf. on Computer-Aided Design*, 2003, pp. 615–620.

[5] D. Sinha and H. Zhou, "A unified framework for statistical timing analysis with coupling and multiple input switching," in *Proc. Intl. Conf. on Computer-Aided Design*, 2005, pp. 837–843.

[6] H. Chang and S. S. Sapatnekar, "Statistical timing analysis under spatial correlations," in *IEEE Transactions on Computer-Aided Design, accepted for publication*, 2004.

[7] J. Le, X. Li, and L. Pileggi, "STAC: Statistical timing with correlation," in *Proc. of the Design Automation Conf.*, 2004, pp. 343–348.

[8] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. of the Design Automation Conf.*, 2004, pp. 331–336.

[9] E. W. Weisstein, *Taylor series - Wolfram web resource*. http://mathworld.wolfram.com/TaylorSeries.html.

[10] D. E. Knuth, *Fundamental algorithms*, 2nd ed., ser. The art of computer programming. Addison-Wesley, 1973, vol. 1.

[11] A. Srivastava, S. Shah, K. Agarwal, D. Sylvester, D. Blaauw, and S. Director, "Accurate and efficient gate-level parametric yield estimation considering correlated variations in leakage power and performance," in *Proc. of the Design Automation Conf.*, 2005, pp. 535–540.