Correlation-Preserved Non-Gaussian Statistical Timing Analysis with Quadratic Timing Model

Lizheng Zhang, Weijen Chen, Yuhen Hu, John A. Gubner, Charlie Chung-Ping Chen
ECE Department,University of Wisconsin, Madison, WI53706-1691, USA
{lizhengz,weijen}@cae.wisc.edu,{hu,gubner,chen}@engr.wisc.edu

ABSTRACT

Recent study shows that the existing first order canonical timing model is not sufficient to represent the dependency of the gate delay on the variation sources when processing and operational variations become more and more significant. Due to the nonlinearity of the mapping from variation sources to the gate/wire delay, the distribution of the delay is no longer Gaussian even if the variation sources are normally distributed.

A novel *quadratic timing model* is proposed to capture the non-linearity of the dependency of gate/wire delays and arrival times on the variation sources. Systematic methodology is also developed to evaluate the correlation and distribution of the quadratic timing model. Based on these, a novel statistical timing analysis algorithm is propose which retains the complete correlation information during timing analysis and has the same computation complexity as the algorithm based on the canonical timing model.

Tested on the ISCAS circuits, the proposed algorithm shows 10× accuracy improvement over the existing first order algorithm while no significant extra runtime is needed.

Categories and Subject Descriptors

B7.2 Hardware [**INTEGRATED CIRCUITS**]: Design Aids—*Verification*

General Terms

Algorithms, Performance, Verification

1. INTRODUCTION

The timing performance of deep-submicron micro-architecture will be dominated by several factors. IC manufacturing process parameter variations will cause device and circuit parameters to deviate from their designed value. Low supply voltage for low-power applications will reduce noise margin, causing increased timing delay variations. Due to dense integration and non-ideal on-chip power dissipation, rising temperature of substrate may lead to hot spot, causing excessive timing variations. Classical worst case timing analysis produces timing predictions that are often too pessimistic and grossly conservative. On the other hand, statistical timing analysis (STA) that characterizes timing delays as statistical random variables offers a better approach for more accurate and realistic timing prediction.

Existing STA methods can be categorized into two distinct approaches: **path based STA** [1–4] and **block based STA** [5–11]. The path based approach seeks to estimate timing statistically on selected *critical paths*. However, the task of selecting a subset of paths whose time constraints are statistically critical has a worst-case complexity that grows exponentially with respect to the circuit size. Hence it is not easily scalable to handle realistic circuits. The block based approach, on the other hand, champions the notion of *progressive computation*. Specifically, each gate/wire is treated as a timing block and the timing analysis is performed block by block in the forward direction in the circuit timing graph without looking back to the path history. As such, the computation complexity would grow linearly with respect to the circuit size.

However, to realize the full benefit of block based STA, one must address a challenging issue that gate/wire delays in a circuit could be correlated since two delays might be affected by the same variation sources of *global variations* such as voltage supply uncertainties, gate channel length variations, wire geometry variations,...,etc. In [6, 7, 10] the delay D is explicitly related with these global variations G_i by the *canonical timing model*:

$$D = \mu + \alpha R + \sum_i \beta_i G_i \tag{1}$$

where R, called *local variation*, accounts the cumulative effect of variation sources other than considered global variations.

The canonical timing model (1) provides an elegant way to deal with the correlations([10]). Unfortunately, the nonlinear relationship between the gate/wire delay and the global variation sources can not be accurately approximated by the current *linear* canonical timing model. And even the global variations are often modeled as Gaussian random variables([12,13]) , the gate/wire delays, in general, will not be Gaussian distributed random variables. This yields unsatisfactory results for deep-sub-micron IC circuit where relative magnitudes of global variations are often larger, while more accurate STA is demanded.

To mitigate this deficiency, in this paper, we propose a novel *quadratic timing model* that augments the linear

DAC 2005, June 13–17, 2005, Anaheim, California, USA.

canonical timing model with second order terms:

$$D = m + \alpha R + \sum_i \beta_i G_i + \sum_{i,j} \Gamma_{ij} G_i G_j \qquad (2)$$

where Γ_{ij} are quadratic coefficients and m is a constant term which may be different from the mean value of the delay timing variable.

Preliminary work reported in [14] indicated that a quadratic timing model delivered 4× accuracy improvement over a first order canonical model. Nevertheless, [14] does not address the important question of *how to systematically develop a quadratic timing model to perform accurate STA of large scale circuits*. The main objective of this paper is to develop such a practical, efficient solution to this question. To this aim, we have made a number of tangible contributions:

(1) A novel quadratic timing model is formulated for both gate/wire delay and signal arrival time to represent the correlation between them. Systematic methodology is also developed to evaluate the correlations and to compute distributions for the quadratic timing model.

(2) A novel statistical timing algorithm is developed based on the quadratic timing model which successfully retains the complete correlation information among arrival times during timing analysis while has the same computation complexity as algorithms based on canonical timing model.

(3) A novel *conditional linear MAX approximation* method is proposed to deal with cases when MAX operator is significantly non-linear. By assuming inputs to be Gaussian, we are able to detect the linearity of the MAX operator by just checking the skewness of the output. Linear approximation of MAX is only used when MAX is decided to be linear. If MAX is non-linear, the evaluation is delayed within a format of *MAX tuple* until it becomes linear in the later timing steps.

The rest of the paper is organized as following: Section 2 presents the quadratic timing model for gate/wire delay and the arrival time; Section 3 introduces the mathematics tools used for correlation and distribution evaluation for quadratic timing model; Section 4 describes the statistical timing algorithm based on the quadratic timing model; Section 5 presents the C/C++ implementation and testing results; Section 6 gives the conclusions.

2. QUADRATIC MODEL OF TIMING VARIABLES

Since the time variables, either gate/wire delays or arrival times, are modeled as non-Gaussian random variables, the mean(μ) and std(σ), used for canonical delay cases, are not sufficient to characterize the distributions of the time random variables. For the interest of the timing analysis, we define a third parameter to assist the distribution characterization:

Definition 1. *For a random time variable X, its* **equivalent two sigma** *value, abbreviated as "e2σ", is defined as:*

$$P(X < e2\sigma) = 97.7\%$$

Except the special case of Gaussian random variable, $e2\sigma \neq \mu + 2\sigma$ in general. Knowing this, we hereafter pay more attention to $e2\sigma$ than the mean and std since it is $e2\sigma$ that really means performance of the considered circuit.

2.1 Quadratic Gate Delay Model

It is generally accepted that the gate delay D_g is a nonlinear function of the global variation variables. We formulate the quadratic gate delay model by taking the second order Taylor expansion of D_g with respect to the global variation variables (evaluated around the mean value of these global variations):

$$\begin{aligned} D_g &\approx m_g + \alpha R + \frac{\partial D_g}{\partial L} L + \frac{\partial D_g}{\partial V} V + ... \\ &+ \frac{1}{2}\frac{\partial^2 D_g}{\partial L^2} L^2 + \frac{\partial^2 D_g}{\partial L \partial V} LV + \frac{1}{2}\frac{\partial^2 D_g}{\partial V^2} V^2 + ... \end{aligned} \qquad (3)$$

In this equation, m_g is a constant and $L, V...$ are global variations. The coefficients in this Taylor expansion can be analytically extracted from the Spice model of the gate delay. Hence, in the following discussion, we will assume these parameters are known in advance.

Assume that there are p global variation variables, one may define a $p \times 1$ Gaussian *variation vector*

$$\boldsymbol{\delta}_g = [G_1, G_2, ..., G_p]^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_g)$$

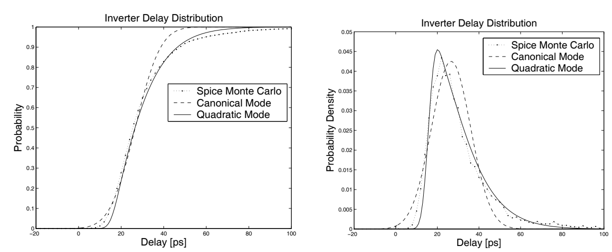
where "*" represents the transpose operation. $\mathbf{0}$ is a zero vector. The *correlation matrix* ($\boldsymbol{\Sigma}_g = E\{\boldsymbol{\delta}_g \boldsymbol{\delta}_g^*\}$) is a $p \times p$ matrix. Generally it is not a unit matrix $\boldsymbol{I}$ since these global variation random variables may be correlated among themselves.

Consolidate equations (2) and (3) into a compact quadratic form:

$$D_g = m_g + \alpha R + \boldsymbol{\beta}_g^* \boldsymbol{\delta}_g + \boldsymbol{\delta}_g^* \boldsymbol{\Gamma}_g \boldsymbol{\delta}_g \qquad (4)$$

where the vector $\boldsymbol{\beta}_g$ and matrix $\boldsymbol{\Gamma}_g$ are only vectorized representation of the Taylor expansion coefficients in equation (3):

$$\beta_g(i) = \frac{\partial D_g}{\partial G_i} \quad and \quad \Gamma_g(i,j) = \frac{1}{2}\frac{\partial^2 D_g}{\partial G_i \partial G_j} \qquad (5)$$



(a) *c.d.f.* of Inverter Delay (b) *p.d.f.* of Inverter Delay

Figure 1: Distributions of Inverter Delay

To demonstrate the advantage of the quadratic gate delay model over the first order canonical model, the probability distribution of an inverter delay is estimated using the Monte Carlo method where the timing delay of each trial is evaluated using SPICE circuit simulator. Using the parameters analytically extracted from the SPICE model, the delay distributions are also computed using both the quadratic timing model and the linear canonical timing model. Shown in figure 1(b), the "true" distribution from Monte Carlo simulation is significantly non-symmetric and non-Gaussian and can not be approximated by any canonical timing model.

2.2 Quadratic Wire Delay Model

As shown in Figure 2, the distributive wire delay model separates the long wire into N equal segments with length of l. Each wire segment i has width of W_i, thickness of T_i which are considered as global variations with identical Gaussian distributions.

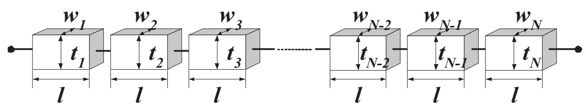


Figure 2: Distributed Wire Delay Model

Elmore's delay model states that the total wire delay will be

$$D_w = \sum_{i=1}^{N}\sum_{j=i}^{N} R_i C_j = \sum_{i=1}^{N}\sum_{j=i}^{N} \frac{r_s l^2 (c_s W_j + c_f T_j)}{W_i T_i} \qquad (6)$$

where R_i and C_j are the resistance and capacitance of the wire segment; r_s is the resistivity of the wire; c_s and c_f are the sheet and fringe unit capacity of the wire.

Applying the Taylor expansion to the Elmore's delay and truncating it until the second order, the quadratic wire delay model can be formulated similarly as that in the case of gate delay:

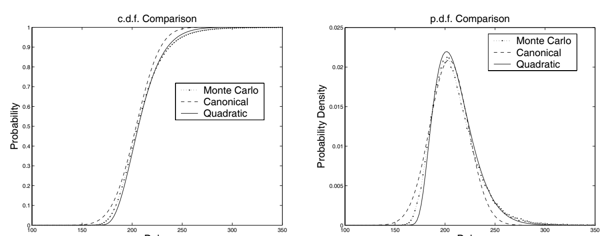
$$D_w \approx m_w + \alpha R + \boldsymbol{\beta}_w^* \boldsymbol{\delta}_w + \boldsymbol{\delta}_w^* \boldsymbol{\Gamma}_w \boldsymbol{\delta}_w \qquad (7)$$

where $\boldsymbol{\delta}_w$ is a $2N \times 1$ global variation vector:

$$\boldsymbol{\delta}_w = [W_1', W_2', ..., W_N', T_1', T_2', ..., T_N'] \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w)$$

with $W_i' = W_i - E\{W_i\}$ and $T_i' = T_i - E\{T_i\}$.

It is important to notice that width and thickness random variables are generally not statistically independent to each other since the wire usually spans a long distance and these random variables may be spatially correlated.



(a) *c.d.f.* from Three Methods (b) *p.d.f.* from Three Methods

Figure 3: Wire Delay Distribution Comparison from Three Approaches

Due to the non-linearity of the wire delay with respecting to the process variations of width and thickness shown in equation (6), the delay distribution of the wire will not be Gaussian even if the width and thickness are usually considered to be Gaussian.([12, 13]) This fact has been clearly shown in Figure 3.

2.3 Quadratic Arrival Timing Model

During timing analysis for a given circuit, the signal arrival timing at each net is the cumulative effect of all gate/wire delays at the input cone of the net. If all gate delays are expressed as the quadratic form, then the arrival time will be:

$$D_a \approx m_a + \sum_{i=1}^{q} \frac{\partial D_a}{\partial D_{g,i}} (D_{g,i} - m_{g,i}) \qquad (8)$$

It is important to comment here that the above form of linear combination is merely an approximation since there will be non-linear operations of MAX/MIN involved to compute the arrival time from the gate/wire delays in its input cone.

If there are q gate/wire delays in the input cone of the arrival time D_a, and there are p global variations involved in the q gate/wire delays, then the arrival time approximated as equation (8) will have the following quadratic form:

$$D_a = m_a + \boldsymbol{\alpha}_a^* \boldsymbol{r}_a + \boldsymbol{\beta}_a^* \boldsymbol{\delta}_a + \boldsymbol{\delta}_a^* \boldsymbol{\Gamma}_a \boldsymbol{\delta}_a \qquad (9)$$

where random variation vectors $\boldsymbol{r}_a = [R_1, R_2, ..., R_q]^* \sim N(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{\delta}_a = [G_1, G_2, ..., G_p]^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_a)$ are mutually independent.

3. CORRELATIONS AND DISTRIBUTIONS

Mathematically, the gate/wire delay quadratic equations (4) and (7) are only special cases of the arrival timing quadratic form (9), so it is safe to conclude that:

Theorem 1. *If every arrival time in a circuit is approximated as a linear combination of its input gate/wire delays, and all gate/wire delays have the quadratic delay form as equation (4) and (7), then all timing variables in the circuit, including gate/wire delays and arrival times, will have the* **quadratic timing model***:*

$$D \sim Q(m, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}) = m + \boldsymbol{\alpha}^* \boldsymbol{r} + \boldsymbol{\beta}^* \boldsymbol{\delta} + \boldsymbol{\delta}^* \boldsymbol{\Gamma} \boldsymbol{\delta} \qquad (10)$$

where $\boldsymbol{r} \sim N(\mathbf{0}, \boldsymbol{I})$ *and* $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ *are mutually independent local variations and global variations.*

Both linear and quadratic dependencies are involved in the quadratic timing model (10). In order to evaluate the correlations between timing variables with quadratic forms, three types of correlation are needed to be computed: (1) correlation between linear terms; (2) correlations between linear and quadratic terms; (3) correlations between quadratic terms.

Applying theorem 4 proved in the Appendix A, it is then easy to compute the correlations between quadratic timing variables as:

Theorem 2. *For quadratic timing variable* $D \sim Q(m, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma})$, *its mean* μ_D *and variance* σ_D^2 *are*

$$\mu_D = E\{D\} = m + tr\{\boldsymbol{\Sigma\Gamma}\} \qquad (11)$$

$$\sigma_D^2 = \boldsymbol{\alpha}^* \boldsymbol{\alpha} + \boldsymbol{\beta}^* \boldsymbol{\Sigma} \boldsymbol{\beta} + 2tr\{\boldsymbol{\Sigma}^2 \boldsymbol{\Gamma}^2\} \qquad (12)$$

and for random variables $D_1 \sim Q(m_1, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\Gamma}_1)$ *and* $D_2 \sim Q(m_2, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_2, \boldsymbol{\Gamma}_2)$, *the correlation between them is:*

$$cov(D_1, D_2) = \boldsymbol{\alpha}_1^* \boldsymbol{\alpha}_2 + \boldsymbol{\beta}_1^* \boldsymbol{\Sigma} \boldsymbol{\beta}_2 + 2tr\{\boldsymbol{\Sigma}^2 \boldsymbol{\Gamma}_1 \boldsymbol{\Gamma}_2\} \qquad (13)$$

where $tr\{\cdot\}$ *means "trace" and equals the sum of the diagonal elements in the matrix.*

To compute the distribution of the quadratic delay D defined in equation (10), we use a statistics technique called *characteristic function*:

Theorem 3. *If the random variable X has a characteristic function of $C_X(\xi)$, then the p.d.f. of the random variable X will be:*

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-j\xi x} C_X(\xi) d\xi \tag{14}$$

The formal proof of this theorem can be found in textbooks of probabilistic theory such as [15].

For the quadratic timing variable $D \sim Q(m, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ defined in equation (10), its *exact* characteristic function can be analytically derived as:

$$C_D(\xi) = |\boldsymbol{\Omega}|^{-\frac{1}{2}} exp\{j\xi m - \frac{1}{2}\xi^2(\boldsymbol{\alpha}^*\boldsymbol{\alpha} + \boldsymbol{\beta}^*\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\beta})\} \tag{15}$$

where $|\boldsymbol{\Omega}|$ is the determinant of matrix $\boldsymbol{\Omega} = \boldsymbol{I} - 2j\xi\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{\frac{1}{2}}$. So the *p.d.f.* of the quadratic time variable, $f_D(x)$, can then be computed from theorem 3.

4. STA WITH QUADRATIC TIMING MODEL

In block based STA, the arrival time random variable propagation involves two elemental operations:

(1)ADD: When an input arrival time X propagates through a gate delay Y, the output arrival time will be $Z = X + Y$;

(2)MAX: When two arrival times X and Y merge in a gate, a new arrival time of $Z = max(X, Y)$ will be formulated before the gate delay is added.

Before analysis, the quadratic parameters of individual gate/wires are extracted from their Spice models and a gate/wire library is then formed. This library, together with the circuit being analyzed, serves as the input of the STA algorithm. The overall data flow of the algorithm is summarized in figure 4.

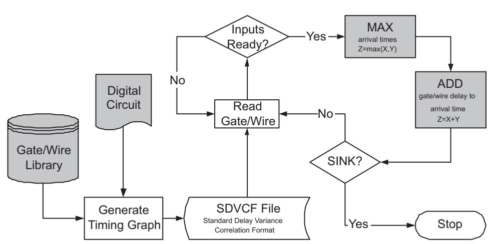


Figure 4: Block-Based STA Algorithm with Quadratic Timing Model

The timing graph in the STA is represented by a file called *standard delay variance correlation format(sdvcf)* where both quadratic gate/wire delays and the gate/wire connections are specified.

4.1 ADD Operation

If both X and Y are expressed in the quadratic form of (10) $X \sim Q(m_X, \boldsymbol{\alpha}_X, \boldsymbol{\beta}_X, \boldsymbol{\Gamma}_X)$ and $Y \sim Q(m_Y, \boldsymbol{\alpha}_Y, \boldsymbol{\beta}_Y, \boldsymbol{\Gamma}_Y)$, then the output of the ADD operator is very straightforward as:

$$Z = X + Y \sim Z(m_Z, \boldsymbol{\alpha}_Z, \boldsymbol{\beta}_Z, \boldsymbol{\Gamma}_Z)$$

where the quadratic parameters are computed as:

$$m_Z = m_X + m_Y \quad ; \quad \boldsymbol{\alpha}_Z = \boldsymbol{\alpha}_X + \boldsymbol{\alpha}_Y$$
$$\boldsymbol{\beta}_Z = \boldsymbol{\beta}_X + \boldsymbol{\beta}_Y \quad ; \quad \boldsymbol{\Gamma}_Z = \boldsymbol{\Gamma}_X + \boldsymbol{\Gamma}_Y$$

4.2 Conditional Linear MAX Operation

MAX operator, however, is more complicated since it is intrinsically a non-linear operator and error will happen if we approximate it with a linear operator. So there is a need to evaluate how linear the MAX operator is. To do this, we accept the following fact:

> *If the inputs of the MAX operator are Gaussian, then the output of the MAX operator will be Gaussian if the MAX operator is linear.*

So the linearity of the MAX operator can be well evaluated by the Gaussianity of the MAX output assuming the inputs are Gaussian. Skewness, which is a symmetry indicator of the distribution, can then be applied for the purpose of Gaussianity checking since Gaussian distribution will always be symmetric.

For our purpose to propagate the quadratic timing model through the MAX operator, we firstly assuming the MAX operation is done on two Gaussian inputs whose mean and variance match what are computed from quadratic timing model. Then we use the equation given in [16] to compute the output skewness. If the skewness is smaller than a threshold, then we know the MAX operator can be well approximated by a linear operator. Otherwise, we put both inputs into a *MAX tuple*(Mt) which is a collection of random variables waiting to be MAXed. We can delay the MAX operation since the ADD operation for a MAX tuple can be simply done as:

$$Mt\{X, Y\} + D = Mt\{X + D, Y + D\}$$

and the MAX operation between two MAX tuples is just to merge these two tuples together:

$$max(Mt\{X, Y\}, Mt\{U, V\}) = Mt\{X, Y, U, V\}$$

To maintain the size of the MAX tuple to be as small as possible, we constantly check the linearity of the MAX operation between any two members of the MAX tuple and MAX them out as long as their MAX output skewness is small enough. With such conditional linear MAX operation, we will then be able to control the error of the linear approximation for MAX operators within an acceptable range.

When two quadratic random variables $X \sim Q(m_X, \boldsymbol{\alpha}_X, \boldsymbol{\beta}_X, \boldsymbol{\Gamma}_X)$, and
$Y \sim Q(m_Y, \boldsymbol{\alpha}_Y, \boldsymbol{\beta}_Y, \boldsymbol{\Gamma}_Y)$, are decided to be MAXed out with linear approximation of $Z = aX + bY + c$, the approximation parameters are computed assuming X and Y are Gaussian and using equations in [16]:

$$a = \Phi \quad b = 1 - \Phi \quad c = \varphi\sigma_{X-Y} \tag{16}$$

where Φ and φ are *c.d.f.* and *p.d.f.* of standard Gaussian distribution evaluated at μ_{X-Y}/σ_{X-Y}. So the quadratic timing variable $Z =\sim Q(m_Z, \boldsymbol{\alpha}_Z, \boldsymbol{\beta}_Z, \boldsymbol{\Gamma}_Z)$ can be easily computed as

$$\boldsymbol{\alpha}_Z = a\boldsymbol{\alpha}_X + b\boldsymbol{\alpha}_Y \quad ; \quad m_Z = am_X + bm_Y + c$$
$$\boldsymbol{\beta}_Z = a\boldsymbol{\beta}_X + b\boldsymbol{\beta}_Y \quad ; \quad \boldsymbol{\Gamma}_Z = a\boldsymbol{\Gamma}_X + b\boldsymbol{\Gamma}_Y$$

4.3 Extra Computation Complexity

When compared with the STA method based on first order canonical timing model, the extra computation complexity of the STA methods based on quadratic timing model will come from updating the quadratic coefficient matrix $\boldsymbol{\Gamma}$ at every arrival time propagation step. The number of quadratic

	mean: μ ($\Delta\mu$)			std: σ ($\Delta\sigma$)			equivalent two sigma delay: $e2\sigma$ ($\Delta e2\sigma$)		
Circuit	M.C.	CanoStat	QuadStat	M.C.	CanoStat	QuadStat	M.C.	CanoStat	QuadStat
C432	1828	1653(9.5%)	1853(1.4%)	779	537 (31%)	734 (5.8%)	3490	2650(24%)	3450(1.2%)
C880	1843	1671(9.4%)	1867(1.3%)	753	448 (40%)	689 (8.5%)	3440	2500(27%)	3360(2.3%)
C1355	1811	1636(9.7%)	1828(0.9%)	746	485 (35%)	697 (6.6%)	3280	2530(23%)	3360(2.4%)
C1908	2437	2190(10%)	2432(0.2%)	914	695 (24%)	880 (3.8%)	4410	3490(20%)	4370(0.9%)
C2670	2666	2404(10%)	2738(2.7%)	1019	618 (39%)	860 (16%)	4840	3560(26%)	4620(4.6%)
C3540	3468	3136(10%)	3499(0.9%)	1344	936 (30%)	1309(2.6%)	6230	4870(22%)	6370(2.3%)
C6288	8798	7950(10%)	9393(6.8%)	3785	2661 (30%)	3535(6.6%)	16799	12919(23%)	16639(1.0%)
C7552	2440	2202(10%)	2489(2.0%)	981	599 (39%)	828 (15%)	4510	3310(27%)	4270(5.3%)
Average Error	-	9.7%	2%	-	34%	8.1%	-	24%	2.3%
Improvement	-	4.9×		-	4.2×		-	10×	

Table 1: Distribution Parameters for ISCAS Circuits with three Approaches: (1)Monte Carlo(M.C.); (2)Canonical Model(CanoStat); (3)Quadratic Model(QuadStat)

coefficients is limited by the number of considered global variations and is usually a constant. Updating matrix $\mathbf{\Gamma}$ will not increase the computation complexity since it only involves moment computation of quadratic timing variables whichis not dependent on the circuit size.

So briefly, the computation complexity of STA based on quadratic timing model will be the same as its canonical timing model correspondence.

4.4 Application in Path Based STA

Although we propose above a block based STA method because path based STA will have potential difficulty to select statistically critical paths in complex circuits, nothing prevents us applying the proposed quadratic timing model in path based STA.

As long as the statistically critical paths are correctly selected, the overall delay distribution of the circuit can be computed very straightforwardly. For the i^{th} critical path cpi, its path delay will be $D_{cpi} = \sum_{g \in cpi} D_g$. When all gate/wire delays are quadratically represented, the path delay will also have quadratic format as:

$$D_{cpi} \sim Q(\sum_{g \in cpi} m_g, \sum_{g \in cpi} \boldsymbol{\alpha}_g, \sum_{g \in cpi} \boldsymbol{\beta}_g, \sum_{g \in cpi} \mathbf{\Gamma}_g) \qquad (17)$$

So if there are n statistically critical paths, the overall delay distribution will be:

$$D_{all} = max(D_{cp1}, D_{cp2}, ..., D_{cpn}) \qquad (18)$$

where the MAX operation is token in the statistical sense.

5. SIMULATIONS AND DISCUSSIONS

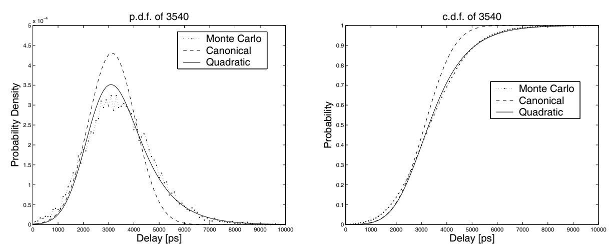
The proposed block based STA with quadratic timing model has been implemented in C/C++ with the name of *QuadStat* and tested on the ISCAS'85 benchmark circuits. For comparison, we also implement the STA based on first order canonical timing model, named *CanoStat*, is also implemented and tested. Monte Carlo simulation with 10,000 repetitions is used as a comparison standard.

All ISACS circuits are re-mapped into a simple standard gate library with gates of *not, nand2, nand3, nor2, nor3, xor/xnor*. All these standard gates are implemented with Cadence tools. Their quadratic timing model parameters are extracted from their Spice model with the variations specified by the technology file.

5.1 Accuracy Improvement

Timing results from both *QuadStat* and *CanoStat* are shown in Table 1 and compared with that from Monte Carlo simulation.

The estimation error is also shown in the table from which it is clear that there is a significant accuracy improvement just by switching the delay model from canonical to quadratic. Measured by the performance critical parameter of equivalent two sigma delay of the circuit, $e2\sigma$, average 10× accuracy improvement is achieved: the average error of CanoStat is 24% while that of QuadStat is small as 2.3%.



(a) *p.d.f.* of ISCAS c3540 (b) *c.d.f.* of ISCAS c3540

Figure 5: Distribution Comparison of ISCAS c3540 from Three Approaches: (1)Monte Carlo; (2)Canonical Model; (3)Quadratic Model

To graphically illustrate the accuracy improvement, the delay distributions of ISCAS circuit c3540 are show in figure 5. It is clear that the accuracy improvement of the QuadStat is mostly due to the high probability region of the distribution which is actually more critical for circuit performance. CanoStat will clearly underestimate the delay in the high probability region. This underestimation, in reality, will result in optimistic design and excessive chip failure. This example clearly shows the necessary to use quadratic timing model when variations become large in nowadays technology.

5.2 Performance Comparison

The CPU time of the three approaches is shown in Table 2 where it is clear that tremendous time is save from Monte Carlo simulation by using either QuadStat and CanoStat. And also, it is clear that there is no significant running time difference between QuadStat and CanoStat which demonstrates the conclusion we made in section 4.3: *NO extra computation is needed to switch from canonical STA methods to quadratic STA methods.*

Circuit	C432	C880	C1335	C1908
Gate Counts	280	641	717	1188
M.C.	84s	3543s	4675s	11530s
QuadStat	0.09s	0.37s	0.561s	1.162s
CanoStat	0.08s	0.36s	0.55s	1.152s
Circuit	C2670	C3540	C6288	C7552
Gate Counts	2004	2485	2704	5355
M.C.	29324s	45381s	54113s	242850s
QuadStat	2.985s	4.577s	5.739s	24.345s
CanoStat	2.954s	4.537s	5.438s	24.285s

Table 2: CPU Time of three Approaches in ISCAS Circuits (1)Monte Carlo(M.C.); (2)Canonical Model(CanoStat); (3)Quadratic Model(QuadStat)

6. CONCLUSIONS

A novel *quadratic timing model* is defined for time variables in statistical timing analysis and its advantages over the existing *canonical timing model* is demonstrated for both gate and wire delays.

Based on this quadratic timing model, the correlations and distribution between those non-Gaussian time variables can be elegantly evaluated. Furthermore, a novel block based statistical timing analysis algorithm is formulated using the quadratic timing model. Testing results on the benchmark circuits show that the new algorithm can significantly improve the timing accuracy without degrading performances. Finally, the advantage of using quadratic timing model in path based timing analysis is also demonstrated by a simple example.

7. ACKNOWLEDGEMENT

This work was partially funded by Intel, TSMC, UMC, Faraday, SpringSoft, National Science Foundation under grants CCR-0093309 & CCR-0204468 and National Science Council of Taiwan, R.O.C. under grant NSC 92-2218-E-002-030. Also great thanks to professor Barry D. Van Veen for the great discussions.

APPENDIX

A. MOMENTS OF QUADRATIC FORM

Theorem 4. *For random vector $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, sensitivity vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and quadratic coefficient matrices $\boldsymbol{\Gamma}$,$\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$*

$$E\{\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}\} = tr\{\boldsymbol{\Sigma}\boldsymbol{\Gamma}\} \tag{19}$$

$$cov(\boldsymbol{\alpha}^*\boldsymbol{\delta}, \boldsymbol{\beta}^*\boldsymbol{\delta}) = \boldsymbol{\alpha}^*\boldsymbol{\Sigma}\boldsymbol{\beta} \tag{20}$$

$$cov(\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}, \boldsymbol{\beta}^*\boldsymbol{\delta}) = 0 \tag{21}$$

$$cov(\boldsymbol{\delta}^*\boldsymbol{\Gamma}_1\boldsymbol{\delta}, \boldsymbol{\delta}^*\boldsymbol{\Gamma}_2\boldsymbol{\delta}) = 2tr\{\boldsymbol{\Sigma}^2\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2\} \tag{22}$$

where "$tr\{\cdot\}$" means "trace" and is the sum of the diagonal elements of the matrix.

PROOF. Equation (19) and (20):

$$E\{\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}\} = E\{tr\{\boldsymbol{\Gamma}\boldsymbol{\delta}\boldsymbol{\delta}^*\}\} = tr\{\boldsymbol{\Gamma}E\{\boldsymbol{\delta}\boldsymbol{\delta}^*\}\} = tr\{\boldsymbol{\Gamma}\boldsymbol{\Sigma}\}$$

$$cov(\boldsymbol{\alpha}^*\boldsymbol{\delta}, \boldsymbol{\beta}^*\boldsymbol{\delta}) = E\{\boldsymbol{\alpha}^*\boldsymbol{\delta}\boldsymbol{\delta}^*\boldsymbol{\beta}\} = \boldsymbol{\alpha}^*E\{\boldsymbol{\delta}\boldsymbol{\delta}^*\}\boldsymbol{\beta} = \boldsymbol{\alpha}^*\boldsymbol{\Sigma}\boldsymbol{\beta}$$

For equation (21), after the vector format is expanded, $E\{\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}\boldsymbol{\beta}^*\boldsymbol{\delta}\} = \sum_{i,j,k} c_{i,j,k}E\{G_iG_jG_k\} = 0$, since all summation terms will be moments with odd order and vanish given that $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Since $E\{\boldsymbol{\beta}^*\boldsymbol{\delta}\} = 0$, then

$$cov(\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}, \boldsymbol{\beta}^*\boldsymbol{\delta}) = E\{\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}\boldsymbol{\beta}^*\boldsymbol{\delta}\} - E\{\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta}\}E\{\boldsymbol{\beta}^*\boldsymbol{\delta}\} = 0$$

For equation (22), with eigenvalue decomposition, $\boldsymbol{\delta}^*\boldsymbol{\Gamma}\boldsymbol{\delta} = \sum_i \lambda_i X_i^2$, where $X_i \sim N(0,1)$ are identical independent Gaussian random variables. So the covariance $cov(\boldsymbol{\delta}^*\boldsymbol{\Gamma}_1\boldsymbol{\delta}, \boldsymbol{\delta}^*\boldsymbol{\Gamma}_2\boldsymbol{\delta})$ will be:

$$\sum_{i,j} \lambda_{1i}\lambda_{2j}\left(E\{X_i^2X_j^2\} - E\{X_i^2\}E\{X_j^2\}\right) = 2tr\{\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\} = 2tr\{\boldsymbol{\Sigma}^2\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2\}$$

where $E\{X_i^2X_j^2\} = E\{X_i^2\}E\{X_j^2\}$ when $i \neq j$. □

B. REFERENCES

[1] J.-J. Liou, A. Krstic, L.-C. Wang, and K.-T. Cheng, "False-path-aware statistical timing analysis and efficient path selection for delay testing and timing validation," *Design Automation Conference, 2002. Proceedings. 39th*, pp. 566 – 569, June 2002.

[2] M. Orshansky, "Fast computation of circuit delay probability distribution for timing graphs with arbitary node correlation," *TAU'04*, Feb 2004.

[3] M. Orshansky and K. Keutzer, "A general probabilistic framework for worst case timing analysis," *Design Automation Conference, 2002. Proceedings. 39th*, pp. 556 – 561, June 2002.

[4] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, and R. Panda, "Statistical delay computation considering spatial correlations," *Design Automation Conference, 2003. Proceedings of the ASP-DAC 2003. Asia and South Pacific*, pp. 271 – 276, Jan 2003.

[5] A. Agarwal, V. Zolotov, and D. Blaauw, "Statistical timing analysis using bounds and selective enumeration," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 9, pp. 1243 –1260, Sept 2003.

[6] C. Visweswariah, K. Ravindran, and K. Kalafala, "First-order parameterized block-based statistical timing analysis," *TAU'04*, Feb 2004.

[7] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," *Computer Aided Design, 2003 International Conference on. ICCAD-2003*, pp. 900 – 907, Nov 2003.

[8] S. Bhardwaj, S. B. Vrudhula, and D. Blaauw, "τau: Timing analysis under uncertainty," *ICCAD'03*, pp. 615–620, Nov 2003.

[9] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," *ICCAD'03*, pp. 607–614, Nov 2003.

[10] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single pert-like traversal," *ICCAD'03*, pp. 621–625, Nov 2003.

[11] S. Tsukiyama, M. Tanaka, and M. Fukui, "A statistical static timing analysis considering correlations between delays," *Proceedings of the 2001 conference on Asia South Pacific design automation*, January 2001.

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

[13] S. Nassif, "Within-chip variability analysis," *Electron Devices Meeting, 1998. IEDM '98 Technical Digest., International*, pp. 283 – 286, Dec 1998.

[14] X. Li, J. Le, P. Gopalakrishnan, and L. T. Pileggi, "Asymptotic probability extraction for non-normal distributions of circuit," *ICCAD'04*, pp. 2–9, Nov 2004.

[15] B. V. Gnedenko, *Theory of Probability*, 6th ed. Gordon and Breach Science Publishers, 1997, translated from Russian by Igor A. Ushakov.

[16] C. Clark, "The greatest of a finite set of random variables," *Operations Research*, pp. 145–162, March 1961.