

# Statistical Clock Tree Routing for Robustness to Process Variations

Uday Padmanabhan  
Dept. of Electrical and  
Computer Engineering  
University of Arizona at Tucson  
AZ 85712, U.S.A  
uday@email.arizona.edu

Janet M. Wang  
Dept. of Electrical and  
Computer Engineering  
University of Arizona at Tucson  
AZ 85712, U.S.A  
wml@ece.arizona.edu

Jiang Hu  
Dept. of Electrical Engineering  
Texas A&M University, College  
Station TX 77843, U.S.A  
jhu@ece.tamu.edu

## ABSTRACT

Advances in VLSI technology make clock skew more susceptible to process variations. Notwithstanding efficient zero skew routing algorithms, clock skew still limits post-manufacturing performance. Process-induced skew presents an ever-growing limitation for high speed, large area clock networks. To achieve multi-GHz operation for high-end designs, clock networks must be constructed to tolerate variations in various interconnect parameters. We propose a statistical centering based clock routing algorithm built upon DME that greatly improves skew tolerance to interconnect variations. The algorithm achieves the improvement by: i) choosing the best center measure which is dynamically based on the first three moments of the skew distribution, and ii) designing for all sink pairs in the subtrees simultaneously. In addition, a variation aware abstract topology generation algorithm is proposed in this paper. Experiments on benchmark circuits demonstrate the efficiency of the proposed method in reducing the number of skew violations by 12% – 37%.

**Categories and Subject Descriptors:** J.6 [COMPUTER-AIDED ENGINEERING]: Computer-aided design (CAD)

**General Terms:** Algorithms, Design, Reliability, Verification.

**Keywords:** Clock Tree, Routing, Robustness, Process Variations.

## 1. INTRODUCTION

Most digital chips are synchronous and the minimum achievable clock period dictates the overall speed of the system. Clock skew is one of the dominant factors that impose an upper bound on the period of the clock. Minimizing clock skew violations is therefore an important design objective.

Given a set of placed flip-flops (clock sinks), the goal of many traditional clock routing methods is to construct a minimal wirelength routing tree connecting these sinks such that certain skew constraint is satisfied. Perhaps the most common skew constraint is zero skew [1] which requires that the clock signal delay at every

sink is the same. Usually, the clock tree routing is a bottom-up recursive merging procedure starting from the sinks. An abstract tree tells which subtrees to be merged at each step. Layout embedding decides the location of each internal node of the abstract tree or the root (merging) node of the subtree resulted from each merging.

In [1], Tsay suggested a layout embedding method to achieve zero skew based on the Elmore delay model. Chao, et al. [2], developed the well known DME (Deferred Merging Embedding) technique to reduce wirelength for zero skew routing. During the bottom-up merging, DME embedding maintains a set of locations, which satisfy skew constraints and form the so-called *merging segment*, instead of fixing an exact location for each internal node. After the bottom-up merging, a top-down tree traversal is performed to choose a single location on each merging segment such that the total wirelength is minimized. Edahiro introduced a nearest neighbor based abstract tree method [3] and integrated it with DME embedding to achieve perhaps the best reported wirelength for zero skew routings. Another variant of skew constraint is useful skew [4] where the skew between a pair of clock sinks is required to be a certain non-zero value for the sake of circuit timing improvement. Cong, et al., proposed a bounded skew tree (BST) algorithm [5] where a global non-zero skew bound is enforced upon all pairs of sinks.

Due to imperfections in the very deep submicron process, the geometry parameters of the fabricated wire may differ from the nominal design values. A recent paper [6] by Liu, et al., stated that interconnect variations may cause as much as a 25% deviation in the clock skew. The presence of process variations makes the zero skew constraint unsatisfiable and existing zero skew algorithms cannot address process variations inherently.

As early as 1993, Pulella, et al., started to consider variations in clock routing [7]. Lu, et al., presented an approach [8] called MinSV (Minimal Skew Violations) to construct process variation tolerant clock trees that aligns the center of the skew range at the center of the permissible range for critical sink pairs. The worst-case skew due to process variations is used to guide the layout embedding to achieve this alignment. Improvements of this work were presented in [9]. An integration with the BST technique was reported in [10].

However, existing works deal with either deterministic or worst-case forms of skew optimization. The statistic behavior of the skew is not captured or utilized for design. The main contribution of this paper is to propose a probabilistic method that minimizes the number of skew violations given a set of skew permissible ranges for sink pairs. To illustrate, first, we propose a statistical-centering

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISPD'06, April 9 – 12, 2006, San Jose, California, USA.

Copyright 2006 ACM 1-59593-299-2/06/0004 ...\$5.00.

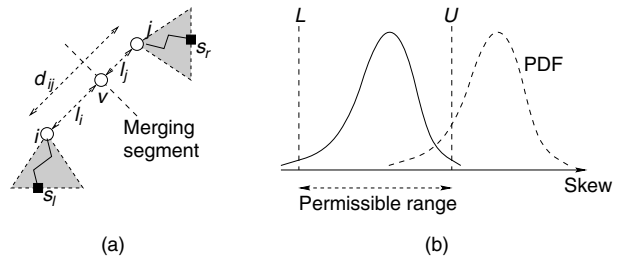
approach in contrast to the corner-point approach in [8]. The moments of the skew distribution are employed to choose the ideal center measure among the mean, median and mode. Our work includes a new delay computation mechanism that incorporates fitted Elmore delay model [11] with analysis of variance (ANOVA) [12] and principle component analysis (PCA) to take into consideration correlations among width and thickness geometric parameters, reduce the number of uncertainties and thus simplify the delay formulation. Second, to avoid a sub-optimal design and excessive violations due to one single critical pair comparison [8], we follow the all pair critical pair comparison idea presented in [9] [10] and impose a weight function based upon both path criticality and permissible range criticality that assigns appropriate importance to every sink pair. Finally, the current paper discusses a topology generation algorithm which takes the width and thickness variation trend into account. Two different variation situations are employed to demonstrate the efficacy of the methods. On all the benchmark circuits, a consistent improvement over existing approaches such as MinSV and nearest neighbor algorithm (NNA) [3] is observed in terms of skew violations and wirelength. The overall improvement is seen to be in the range of 12%-37% for both variation profiles.

## 2. MOTIVATION AND OVERVIEW

In this work, we will consider the effect of process variations during clock tree routing. In this scenario, the skew constraint requires that a skew  $q_{ij}$  between a pair of sinks  $i$  and  $j$  has to be within the *permissible range*  $[L_{ij}, U_{ij}]$ . In fact, this is the most fundamental timing constraint for synchronous circuits [13] and each of the other forms of skew constraints is a subset of it. For a circuit with  $n$  clock sinks, we only need to specify permissible range for  $n - 1$  pairs of sinks based on timing analysis. The permissible ranges of the other pairs can be derived from the  $n - 1$  basic permissible ranges [13]. Besides the traditional objective on wirelength, this work tries to minimize the chance of violating the skew permissible ranges when process variations are considered.

One focus of this work is to find statistical layout embedding techniques to minimize the chance of skew violations. Consider the example in Figures 1 where subtree  $T_i$  and  $T_j$  are merged at node  $v$ . Then, the location (or embedding) of  $v$  decides the skew between the sinks of the two subtrees. For skew  $q_{lr}$  between sink  $s_l$  and  $s_r$ , its permissible range is indicated by a pair of vertical dashed segments in Figures 1(b). Because of process variations, the skew  $q_{lr}$  is a random variable following certain probability distribution as illustrated by the PDF (probability density function) curves in Figures 1(b). When the location of  $v$  is changed, the PDF curve shifts along horizontal (skew) axis. Our effort is aimed to find the location (or merging segment) of an internal node (such as node  $v$ ) such that the overlapping area between the PDF curve and the permissible range is maximized. This is demonstrated by moving from the dashed PDF curve to the solid PDF curve in Figures 1(b).

The previous work of [8] estimates the worst case minimal skew  $\underline{q}_{lr}$  and maximal skew  $\bar{q}_{lr}$  and attempts to align the center  $(\underline{q}_{lr} + \bar{q}_{lr})/2$  of the skew range with the center of its permissible range. This is illustrated in Figure 2(a). If the PDF of the skew is asymmetric, which is by and large true, and the permissible range is small, the alignment of [8] may leave a fat tail of PDF out of the permissible range as indicated in Figure 2(b). In this case, the center of the permissible range is better to be aligned close to the peak probability, for example to the median as in Figure 2(c). If the permissible range is relatively large, however, such alignment may leave a long tail of PDF out of the permissible range on one side while there is abundant slack on the other side like in Figure 2(d). Therefore, a correct alignment depends on both the size of permis-



**Figure 1: (a) Merge subtree  $T_i$  and  $T_j$  at node  $v$  and consider skew between sink  $s_l$  and  $s_r$ . (b) PDF curves of skew  $q_{lr}$  when location of  $v$  is changed.**

sible range and the shape of PDF. A main contribution of this work is the statistical centering based layout embedding technique which can minimize the probability of skew violation for all of these scenarios which may occur in practice.

The details of this work will be described in Section 3, which focuses on delay and variation models, and Section 4 which introduces our statistical clock routing algorithm. In Section 3, basic delay model (Section 3.1) and process variation models (Section 3.2) will first be covered. Then, Section 3.3 will introduce concepts of center measure which is an important basis of our work. Section 3.4 will tell how to reduce the space of random variables so that the statistical clock routing can be performed at practical runtime. The algorithm part will start with a new method of handling multiple permissible ranges (Section 4.1). Our statistical layout embedding technique will be described in Section 4.2. A variation aware abstract tree method will be proposed in Section 4.3.

## 3. DELAY AND VARIATION MODELS

### 3.1 Delay Model

Most of existing clock routing works [8–10] are based on the Elmore delay model due to its simplicity. However, the Elmore delay is sometimes inaccurate and we will use the fitted Elmore delay (FED) [11] model which has better accuracy and still analytical expression. Since FED is derived from the Elmore delay model, we first describe the Elmore delay vs. process parameters. Considering the wire segment of length  $l$  with a driving resistance  $r_d$  (optional) at one end and a load capacitance  $c_L$  at the other end, the Elmore delay in terms of the wire width  $w$  and wire thickness  $t$  can be written as:

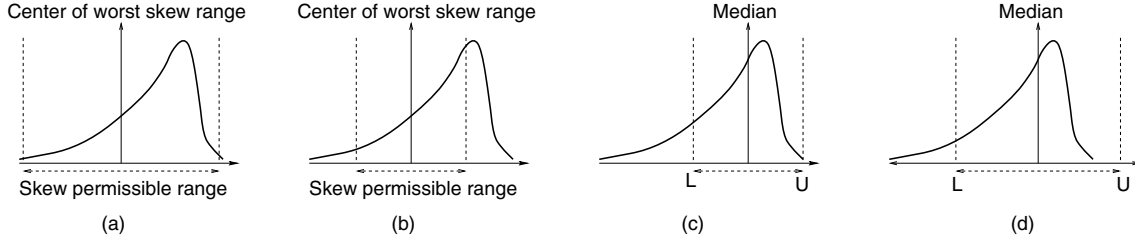
$$d_{ED} = r_d c_a l w + r_d c_f l + r_d c_L + \frac{\rho c_a l^2}{2t} + \frac{\rho c_f l^2}{2wt} + \frac{\rho l c_L}{tw} \quad (1)$$

where  $\rho$  is the resistivity, the sheet resistance  $r_s = \frac{\rho}{t}$ ,  $c_a$  the per unit area capacitance and  $c_f$  the per unit length fringing capacitance. In the rest of the paper,  $d_i$  represents the delay,  $l_i$  the length,  $w_i$  the width, and  $t_i$  the thickness of the edge connecting node  $i$  to its parent. Similarly,  $c_{Li}$  represents the load capacitance at node  $i$ .

Correspondingly, the fitted Elmore delay is defined as follows

$$d_{FED} = A r_d c_a l w + B r_d c_f l + C r_d c_L + D \frac{\rho c_a l^2}{2t} + E \frac{\rho c_f l^2}{2wt} + F \frac{\rho l c_L}{tw} \quad (2)$$

where the coefficients  $A, B, C, D, E$  and  $F$  are determined by a multiple linear regression to accurate delay data provided by simulators like HSPICE. Here, the fitted Elmore delay model also includes the wire thickness which is different from the original equations in [11]. For a general tree, let  $T$  be the set of all tree edges. Let  $E_i$

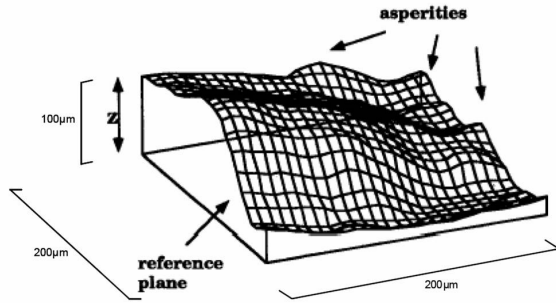


**Figure 2: Aligning center of skew range to the center of permissible range works fine when the permissible range is large as in (a), but not well when the permissible range is small as in (b). Aligning to the median works better when the permissible range is small as in (c), but not well when the permissible range is large as in (d).**

be the set of tree edges at the downstream of edge  $i$ . Let  $S$  be the set of all sinks. Let  $S_i$  be the set of sinks at the downstream of edge  $i$ . Let  $P_k$  be the set of tree edges along the path from the driver to node  $k$ . So, for any node  $k$ , the fitted Elmore delay is

$$\begin{aligned}
 d_{FEDk} = & Ar_d c_a \sum_{i \in T} l_i w_i + Br_d c_f \sum_{i \in T} l_i + Cr_d \sum_{j \in S} c_{Lj} \quad (3) \\
 & + D\rho c_a \sum_{i \in P_k} \frac{l_i}{w_i t_i} \left( \frac{l_i w_i}{2} + \sum_{j \in E_i} l_j w_j \right) \\
 & + E\rho c_f \sum_{i \in P_k} \frac{l_i}{w_i t_i} \left( \frac{l_i}{2} + \sum_{j \in E_i} l_j \right) \\
 & + F\rho \sum_{i \in P_k} \frac{l_i}{w_i t_i} \left( \sum_{j \in S_i} c_{Lj} \right)
 \end{aligned}$$

### 3.2 Process Variation Model



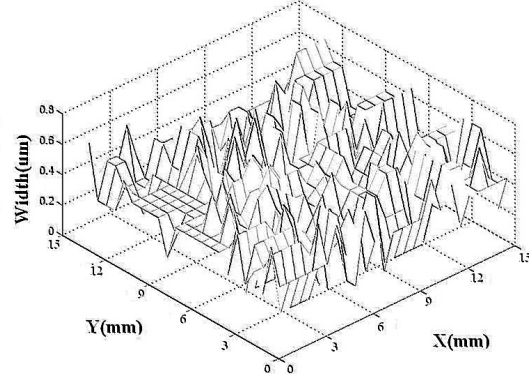
**Figure 3: Global variation of wire width.**

In this work, variations from sink capacitance, wire width and wire thickness are considered. The variation of ILD thickness can be included easily as it affects only wire capacitance. In general, wire variations are more difficult to model than those of sink capacitance because of the distributive nature of wires [6]. Therefore, the description of variation model is emphasized on wire width and thickness. It is stated in [6] that the random variations in the parameters can be decomposed into global and local variation terms where the global component of variation has a relatively low spatial frequency and simple models can be used to describe it. In general, the global spatial variations and the local random variations can be expressed together as,

$$w = w_o + f(x, y) + \delta_w \quad t = t_o + g(x, y) + \delta_t \quad (4)$$

where  $w_o$  and  $t_o$  represent the nominal values for width and thick-

ness, respectively.  $f(x, y)$  and  $g(x, y)$  model global spatial variations.  $\delta_w$  and  $\delta_t$  are random variables describing the local variations. Existing works [6] model the global spatial component by a slant plane. However, an analysis of [14] indicates that the spatial variations of wire across the chip are highly nonlinear. In this work,  $f(x, y)$  and  $g(x, y)$  are modeled as slow changing functions as in Figures 3 across the chip to reflect the observed non-linearity. They also reflect spatial correlations among the variations.  $\delta_w$  and  $\delta_t$  may be regarded as Gaussian distribution with mean 0 and standard deviation  $\sigma$ . Note that the proposed approach and algorithms in the current paper do not require the Gaussian distribution assumption. This point will be further explained in the later sections. In general, combining impacts from both global and local variations, wire width and thickness displays behavior as shown in Figure 4.



**Figure 4: Wire width across chip considering both global and local variations.**

### 3.3 Center Measures and Moments

Due to the nonlinearity of variation parameters, a number of researchers [6, 15, 16] have confirmed that the skew may be non-Gaussian distributed. That is, the skew distribution may not be symmetric and thus cannot be easily modeled by just mean and variance. In order to align the PDF of clock skew with its permissible range (Section 2), other center measures such as median and mode as well as higher order moments like skewness and kurtosis may be needed to better describe the shape of the skew distribution.

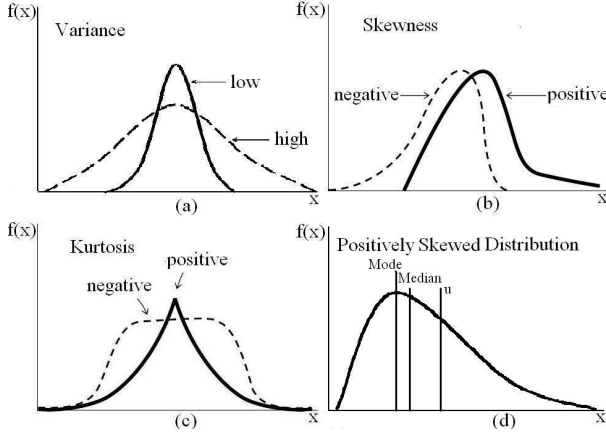
The mean, median and mode are the center measures of a distribution. The *mean*  $\mu$  is the most commonly employed measure of central tendency. It minimizes the  $L^2$  norm of the residuals. The

median  $M$ , on the other hand, minimizes the sum of absolute deviations, or equivalently, the  $L^1$  norm of the residuals. The *mode* is defined as the data value that has the maximum probability of occurrence. Figure 5(d) displays the Mode, Median  $M$  and Mean  $\mu$  definitions. The three center measures are equal to each other when  $f(x)$  is a Gaussian Distribution. Otherwise, Median is always between Mode and Mean. In the rest of the paper, the term **CM (center measure)** is used to refer collectively to the mean, median and mode.

The higher order moments provide information about the shape and spread of the distribution. The variance is the second central moment and it expresses the amount of spread of the distribution (Figure 5(a)). The third central moment, i.e. the skewness  $\varsigma$ , of a random variable  $x$  describes the *degree of symmetry* in the density function  $f_X(x)$  and is defined as,

$$\varsigma = \int \left( \frac{x - \mu_x}{\sigma_x} \right)^3 f_X(x) dx \quad (5)$$

A positive and negative skewness value indicate a distribution skewed to the right and left respectively as demonstrated in Figure 5(b). The kurtosis is the fourth central moment and describes the *relative peakedness/flatness* of a distribution. Figure 5(c) illustrates what kurtosis looks like for different cases.



**Figure 5: (a) Variance, (b) Skewness, (c) Kurtosis, and (d) Mean, Median, Mode on an Asymmetric Distribution**

The mean and variance of the skew in a subtree can be obtained by propagation of the means and variances of the delays in the subtrees. For a random variable  $y = g(x)$  (where  $x$  is a random variable), the first order estimates of the mean and variance can be obtained by Taylor series expansion as,

$$\mu_Y \simeq g + \frac{1}{2} \left( \frac{\partial^2 g}{\partial x^2} \right) \sigma_x^2 \quad \sigma_Y^2 \simeq \left( \frac{\partial g}{\partial x} \right)^2 \sigma_x^2 \quad (6)$$

respectively, where the function  $g$  and its derivatives are evaluated at  $x = \mu_x$  [17]. Since the delay  $d_{FED}$  is a function of the random variables  $w$ ,  $t$ ,  $r_d$  and  $c_L$ , the mean and variance of the delay are determined as,

$$\begin{aligned} \mu_T &\simeq d_{FED}|_{w=\mu_w, t=\mu_t, r_d=\mu_{r_d}, c_L=\mu_{c_L}} \\ &+ \frac{1}{2} \left[ \left( E \frac{\rho c_f l^2}{t} + 2F \frac{\rho l c_L}{t} \right) \frac{1}{\mu_w^3} \right. \\ &\left. + (D \rho c_a l^2 + E \frac{\rho c_f l^2}{w} + 2F \frac{\rho l c_L}{w}) \frac{1}{\mu_t^3} \right] \end{aligned}$$

$$\sigma_T^2 \simeq \left( \frac{\partial d_{FED}}{\partial r_d} \right)^2 \sigma_{r_d}^2 + \left( \frac{\partial d_{FED}}{\partial c_L} \right)^2 \sigma_{c_L}^2 + \left( \frac{\partial d_{FED}}{\partial w} \right)^2 \sigma_w^2 + \left( \frac{\partial d_{FED}}{\partial t} \right)^2 \sigma_t^2$$

Let  $\mu_{Til}$  represent the mean and  $\sigma_{Til}^2$  represent the variance of the delay along the path  $P_{il}$  from  $i$  to the sink  $s_l$  in its subtree. Treating the individual variational sources as independent random variables after OPCA (Orthogonal Principle Component Analysis) procedure, we have,

$$\mu_{Til} = \sum_{k \in P_{il}} \mu_{Tk} \quad \sigma_{Til}^2 = \sum_{k \in P_{il}} \sigma_{Tk}^2 \quad (7)$$

where  $\mu_{Tk}$  and  $\sigma_{Tk}^2$  represent the mean and variance of the delay for the edge connecting node  $k$  to its parent. Similar expressions can be written for the path  $P_{jr}$  from  $j$  to  $s_r$ . The mean and variance of the skew between paths  $P_{il}$  and  $P_{jr}$  is denoted by  $\mu_{S_{lr}}$  and  $\sigma_{S_{lr}}^2$ , respectively. Therefore, we have  $\mu_{S_{lr}} = \mu_{Til} - \mu_{Tjr}$  and  $\sigma_{S_{lr}}^2 = \sigma_{Til}^2 + \sigma_{Tjr}^2$ .

From Equation (2), the edge-delay is the part of  $d_{FED}$  with  $r_d = 0$  (not directly connected to driver). It is a monotonically decreasing function in  $w$  and  $t$  and monotonically increasing function in  $c_L$ . The median of  $d_{FED}$  ( $M_T$ ) can be determined in a straightforward way and is given by,

$$M_T = D \frac{\rho c_a l^2}{2M_t} + \frac{E \rho c_f l^2}{2M_t M_w} + \frac{F \rho M_{c_L}}{M_t M_w} \quad (8)$$

where  $M_w$  is the median of the width distribution,  $M_t$  median of the height distribution and  $M_{c_L}$  median of the load capacitance.

If  $M_{Til}$  represents the median of the delay from internal node  $i$  to the sink  $s_l$  in its subtree (and  $M_{Tjr}$  the median of the delay from  $j$  to  $s_r$ ),  $M_{Til} = \sum_{k \in P_{il}} M_{Tk}$  where  $P_{il}$  denotes the path from  $i$  to  $s_l$  and  $M_{Tk}$  represents the delay of the edges on  $P_{il}$ . The error in estimating the median of a sum of random variables by the sum of their medians is found to be a few percent at most. The reduction in computational complexity when using these algebraic computations justifies the minor loss in accuracy. The median of the skew between the sink pair  $(s_l, s_r)$  in the constructed subtree can be expressed as the difference in the path delay medians along  $P_{il}$  and  $P_{jr}$  i.e.  $M_{S_{lr}} = M_{Til} - M_{Tjr}$ .

Unlike the mean and the median, the mode of a function of a random variable cannot be expressed explicitly in terms of the function parameters. An estimate of the mode can be determined by approximating or fitting the delay distribution by a known distribution that provides an analytic expression for the mode. The Weibull distribution has this property. For our purposes, the Weibull distribution has the added advantage that it can be used to fit both positively and negatively skewed asymmetric distributions. The Weibull probability density function is,

$$f_{WB}(x) = \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad (9)$$

We present the approach used for determining the parameters  $\alpha$  and  $\beta$  of the Weibull distribution (from [18]). The mean and the variance of the Weibull distribution can be written as,

$$\begin{aligned} \mu_{WB} &= \beta \Gamma(1 + \theta) \\ \sigma_{WB}^2 &= \beta^2 \left( \Gamma(1 + 2\theta) - \Gamma^2(1 + \theta) \right) \end{aligned} \quad (10)$$

where  $\Gamma$  represents the Gamma function and  $\theta = 1/\alpha$  is used for convenience. After some derivation, we have,

$$\frac{\Gamma(1 + 2\theta)}{\Gamma^2(1 + \theta)} = \frac{\sigma_{S_{lr}}^2 + \mu_{S_{lr}}^2}{\mu_{S_{lr}}^2} \quad (11)$$

Let  $r = (\sigma_{S_{lr}}^2 + \mu_{S_{lr}}^2) / (2\mu_{S_{lr}}^2)$ . The parameter  $\theta$  (and therefore  $\alpha$ )

can be obtained by solving

$$r = \frac{1}{2} \frac{\Gamma(1+2\theta)}{\Gamma^2(1+\theta)} \quad (12)$$

It is observed that  $\theta$  exhibits an almost linear relationship with the logarithm of  $r$  and therefore a lookup table approach gives quite accurate results. Once  $\theta$  is known, the parameter  $\beta$  can be determined from Equation (10) as  $\beta = -m_1/\Gamma(1+\theta)$ . Given the mean and variance of the distribution, the parameters  $\alpha$  and  $\beta$  of the Weibull distribution can be calculated as in [18]. Then, the mode is given as,

$$Mode = \sqrt[\alpha]{\frac{\alpha-1}{\beta-\alpha}} \quad (13)$$

### 3.4 Random Variable Reduction Based on ANOVA and OPCA

For a clock tree with hundreds of edges, which is a very common size, the fitted Elmore delay may involve thousands of correlated random variables. We introduce analysis of variance (ANOVA) [12] and orthogonal principle component analysis (OPCA) to help create independent parameters and reduce the number of uncertainties.

Given a function of multiple random variables, analysis of variance (ANOVA) is a hierarchical statistical approach to find out each random variable's significance on the function by using variance. The variance ( $\sigma^2$ ) of the interconnect delay can be decomposed into parts that attributed to each random variable  $\sigma_{delay}^2 = \sum_i \sigma_i^2$ . Here the variance of the delay due to a particular random variable is denoted as  $\sigma_i^2$ .  $\sigma_i$  can be found by keeping that random variable constant and varying all other random variables. This is accomplished by integrating that random variable out of the model. As ANOVA is a hierarchical approach, we can define random variables  $\xi_i$  as a function or combination of variational parameters. For example, let  $\xi_i = r_d c_a w_i$  where  $r_d$ ,  $c_a$  and  $w_i$  are variational parameters. Because  $r_d$ ,  $c_a$  and  $w_i$  are random variables, so as to  $\xi_i$ . Thus we can use random variables to represent terms in the delay model. By reducing the number of random variables, we in fact reduce the number of terms in the delay model  $d_{FED}$ . To illustrate, assume that  $d_{FEDk}$  in Equation (3) has  $n$  terms. Let  $\{\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_n\}$  represent all the  $n$  random variables in Equation (3). The mean response that gives the main effect of each random variable  $\xi_i$  is defined as follows:

$$\hat{\mu}_i(\xi_i) \equiv \int \dots \int d_{FEDk}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 \dots d\xi_{i-1} d\xi_{i+1} \dots d\xi_n \quad (14)$$

Then the variance ( $\hat{\sigma}^2$ ) due to variable  $\xi_i$  is  $\hat{\sigma}_i^2 = \int [\hat{\mu}_i(\xi_i) - \mu]^2 d\xi_i$ . Here,  $\mu$  is the mean of the delay. The proportion of variance due to each random variable towards the total variance of the response is used as a statistical significance parameter of that particular random variable (or term) given by the ratio (refer as F-ratio) :

$$F_{\xi_i} = \frac{\int [\hat{\mu}_i(\xi_i) - \mu]^2 d\xi_i}{\sigma^2} \quad (15)$$

The higher F-ratio, the more importance the term. In the experimental section, we demonstrate that ANOVA can reduce on the average 40% terms in the fitted Elmore delay model.

The application of OPCA is similar as that in statistical timing analysis [19] and the description is omitted here due to page limit. With the large number of cross terms or correlations in the delay model, OPCA alone can only reduce 10% of the parameters as demonstrated in the experimental section. Thus, unlike the existing research where only OPCA is used, our work employs both ANOVA and OPCA at the same time. OPCA can also eliminate

cross terms like  $w_i t_i$  or correlations among the random variables. Without correlations, the central measures such as Mean, Median and Mode have simpler formulas.

## 4. ALGORITHM

### 4.1 All-pair Centering

When we try to align the center measure of skew distribution to the center of permissible range, there is a difficulty. That is, there are often multiple sink pairs between the two subtrees we are merging and it is not clear which sink pair to be aligned. Aligning for a certain pair may leave another sink pair with large probability of skew violation. The work of [8] solves this problem by aligning a critical pair which have tight permissible range and are relatively far apart from each other. However, this method cannot handle the case where there are several sink pairs with similar criticality.

We propose an all-pair centering scheme which can capture an overall skew and permissible range so that the weakness of previous work is overcome. Consider to merge subtree  $T_i$  and  $T_j$ , with sink set  $S_i$  and  $S_j$ , respectively. Before the merging, we first evaluate the existing skew between them. That is, the skew between delay from node  $i$  to sink  $l \in S_i$  and delay from  $j$  to sink  $r \in S_j$ . An overall center measure for the skew between all pairs of sinks in  $S_i$  and  $S_j$  is defined as:

$$CM_{sub} = \frac{\sum_{l \in S_i} \sum_{r \in S_j} \frac{CM_{slr}}{|PR_{lr}|}}{\sum_{l \in S_i} \sum_{r \in S_j} \frac{1}{|PR_{lr}|}} \quad (16)$$

where  $|PR_{lr}| = U_{lr} - L_{lr}$  is the size of the permissible range for skew between sink  $s_l$  and  $s_r$ . This is a weighted average of center measure of all pairs between  $S_i$  and  $S_j$ . The weight is inversely proportional to the size of permissible range, i.e., a pair with tight permissible range has a relatively large weight. Please note that the complexity of determining  $CM_{sub}$  is not greater than that of choosing a single critical pair [8] since both approaches require to go over/compare  $|S_i| * |S_j|$  pairs.

The overall skew center measure, which is  $CM_{sub} + q_{ij}$ , will be aligned to an overall center of permissible ranges for all pair of sinks which is defined as:

$$PR_{sub} = \frac{\sum_{l \in S_i} \sum_{r \in S_j} \frac{(U_{lr} + L_{lr})/2}{|PR_{lr}|}}{\sum_{l \in S_i} \sum_{r \in S_j} \frac{1}{|PR_{lr}|}} \quad (17)$$

**DEFINITION 1.** The sensitivity of  $CM_{Spq}$  (the center measure of the skew) between a sink pair  $(s_p, s_q)$  to the center measure of an edge delay  $CM_{Tk}$  is defined as  $Sn_{pqk} = \partial CM_{Spq} / \partial CM_{Tk}$ .

**LEMMA 1.** Given a sink pair  $(s_p, s_q)$ , the All-Pair Centering approach reduces  $Sn_{pqk}$ ,  $\forall k \in \{P_{ip}\} \cap \{T_i - P_{il}\}, \{P_{jq}\} \cap \{T_j - P_{jr}\}$  (where  $(s_l, s_r)$  is the critical pair) over the critical pair approach.

*Proof:* Omitted due to page limit. ■

### 4.2 Statistical Centering Based Layout Embedding

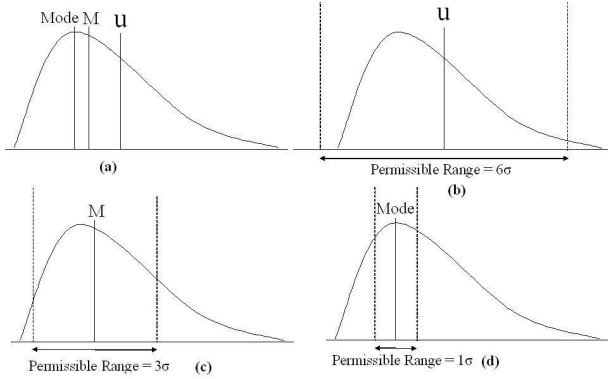
From Section 2, we know that the alignment between skew and corresponding permissible range depends on both the shape of skew PDF and the size of the permissible range. Here, we propose an alignment scheme for different cases. Figure 6 provides a graphical view of this scheme. In the following,  $\zeta$  represents the skewness,  $\sigma$  the standard deviation and  $|PR_{lr}| = U_{lr} - L_{lr}$ .  $\epsilon$  denotes the threshold value for the skewness. If  $|\zeta| > \epsilon$ , we consider the distribution asymmetric.

**CASE I.**  $|\zeta| \leq \epsilon, \forall |PR_{lr}|$ . The distribution is symmetric or almost symmetric. Therefore,  $CM_{Slr} = \mu_{Slr}$ .

CASE II.  $|\zeta| > \epsilon$ ,  $|PR_{lr}| > 5\sigma$ . The distribution can be easily fit within the permissible range by aligning the mean. Therefore,  $CM_{Slr} = \mu_{Slr}$ . Choosing the median or mode might lead to an excessive part of the distribution lying outside the permissible range.

CASE III.  $|\zeta| > \epsilon$ ,  $2\sigma < |PR_{lr}| < 5\sigma$ . The permissible range is not large enough to fit the entire distribution and is not extremely narrow either. As shown in Figure 6(c),  $CM_{Slr} = M_{Slr}$  is the best choice in this case.

CASE IV.  $|\zeta| > \epsilon$ ,  $|PR_{lr}| < 2\sigma$ . The permissible range represents a very stringent constraint. Based on the previous discussion, we choose  $CM_{Slr} = Mode_{Slr}$  to maximize the area within bounds.



**Figure 6: (a) Mean (' $\mu$ '), Median ('M') and Mode ('Mode') on an asymmetric distribution, (b) CASE II, (c) CASE III, (d) CASE IV**

Next, we describe statistical layout embedding based on the above alignment strategy. Consider merging subtree  $T_i$  and  $T_j$  at node  $v$ . Let the Manhattan distance between  $v$  and  $i$  ( $j$ ) be  $l_i$  ( $l_j$ ). An embedding needs to find the value of  $l_i$  and  $l_j$  such that the alignment between skew and permissible range is achieved subject to the constraint of  $l_i + l_j \geq d_{ij}$  where  $d_{ij}$  is the Manhattan distance between  $i$  and  $j$ . The case of  $l_i + l_j > d_{ij}$  corresponds to wire snaking [1].

According to Section 4.1, we need to align the overall center measure of skew ( $CM_{Sv}$ ) to the overall permissible ranges  $PR_{sub}$  between all pairs of sinks  $S_i$  and  $S_j$ . The skew after the merging at  $v$  includes two parts. One is the skew between delay from  $i$  to sinks in  $S_i$  and from  $j$  to sinks in  $S_j$ , which is  $CM_{sub}$  defined in Section 4.1. The other is the skew  $q_{ij}$  between  $i$  and  $j$  when the embedding of  $v$  is finished:

$$q_{ij} = A_s l_i^2 + B_s l_i - C_s \quad (18)$$

where

$$A_s = \frac{\rho}{2} [c_a (\frac{D_i}{t_i} - \frac{D_j}{t_j}) + c_f (\frac{E_i}{w_i t_i} - \frac{E_j}{w_j t_j})] l_i^2 \quad (19)$$

$$B_s = \rho [d_{ij} (\frac{D_j c_a}{t_j} + \frac{E_j c_f}{w_j t_j}) + \frac{F_i c_{L_i}}{w_i t_i} + \frac{F_j c_{L_j}}{w_j t_j}]$$

$$C_s = F_j \rho c_{L_j} d_{ij} \frac{1}{w_j t_j}$$

Either  $\mu_{Sij}$  or  $M_{Sij}$  is used for skew centering at node  $v$ . The terms mean-centric design and median-centric design respectively are used to refer to these approaches. Mode-centric design is not presented because analytic design equations for the mode of the skew are not available.

Since the lengths  $l_i$  and  $l_j$  are yet to be decided and are not known, the higher moments of  $q_{ij}$ , i.e. the variance and the skew-

ness, cannot be evaluated in order to choose between the two approaches. Therefore, mean-centric design is carried out and the moments of the resultant distribution are evaluated. If the skewness  $\zeta_{\mu S}$  of the distribution is high, median-centric design is carried out. Mean-centric design and median-centric design are discussed as follows.

#### 4.2.1 Mean-Centric Design

For mean-centric design,  $CM_{Sv}$  (the center measure of the skew at node  $v$ ) is the sum of  $CM_{sub}$  and the mean skew between from  $v$  to its children  $i$  and  $j$  ( $\mu_{Sij}$ ). In order to align  $CM_{Sv}$  the overall permissible range center  $PR_{sub}$ , we have

$$\mu_{Sij} = PR_{sub} - CM_{sub} \quad (20)$$

where the RHS can be computed from subtree  $T_i$  and  $T_j$  and the LHS is a function of  $l_i$  (or  $l_j$  if wire snaking is necessary). Therefore, the value of  $l_i$  can be decided by solving the above equation. Using Equation (6) and Equation (18), the mean of the skew is derived as,

$$\mu_{Sij} = A_\mu l_i^2 + B_\mu l_i - C_\mu \quad (21)$$

where

$$\begin{aligned} A_\mu &= A_{s\mu} + \frac{1}{2} (\frac{\partial^2 A_s}{\partial t_i^2} |_{t_i=\mu_{t_i}, w_i=\mu_{w_i}} \sigma_{t_i}^2) \\ &+ \frac{\partial^2 A_s}{\partial t_j^2} |_{t_j=\mu_{t_j}, w_j=\mu_{w_j}} \sigma_{t_j}^2 + \frac{\partial^2 A_s}{\partial w_i^2} |_{t_i=\mu_{t_i}, w_i=\mu_{w_i}} \sigma_{w_i}^2 \\ &+ \frac{\partial^2 A_s}{\partial w_j^2} |_{t_j=\mu_{t_j}, w_j=\mu_{w_j}} \sigma_{w_j}^2) \end{aligned} \quad (22)$$

with

$$A_{s\mu} = \frac{\rho c_a}{2} (\frac{D_i}{\mu_{t_i}} - \frac{D_j}{\mu_{t_j}}) + \frac{\rho c_f}{2} (\frac{E_i}{\mu_{w_i} \mu_{t_i}} - \frac{E_j}{\mu_{w_j} \mu_{t_j}}) \quad (23)$$

Likewise, we can derive  $B_\mu$  and  $C_\mu$ .

The solution to Equation (20) must lie in one of three ranges:

- (i)  $0 < l_i \leq d_{ij}$ . In this case,  $l_j = d_{ij} - l_i$ ;
- (ii)  $l_i > d_{ij}$ . Here,  $l_j = 0$ . The appropriate value of  $l_i$  is obtained by solving Equation (20).
- (iii)  $l_i \leq 0$ . In this case,  $l_i = 0$ .  $l_j$  is obtained by solving Equation (20).

#### 4.2.2 Median-Centric Design

The median of the skew from  $v$  to nodes  $i$  and  $j$  is used to guide the design of lengths  $l_i$  and  $l_j$ . Similar as the case of mean-centric design, the length  $l_i$  is then obtained by solving the equation

$$M_{Sij} = PR_{sub} - CM_{sub} \quad (24)$$

where  $M_{Sij} = A_{sM} l_i^2 + B_{sM} l_i - C_{sM}$ . The coefficients can be determined as

$$A_{sM} = \frac{\rho}{2} [c_a (\frac{D_i}{M_{t_i}} - \frac{D_j}{M_{t_j}}) + c_f (\frac{E_i}{M_{w_i} M_{t_i}} - \frac{E_j}{M_{w_j} M_{t_j}})] l_i^2 \quad (25)$$

$$B_{sM} = \rho [d_{ij} (\frac{D_j c_a}{M_{t_j}} + \frac{E_j c_f}{M_{w_j} M_{t_j}}) + \frac{F_i M_{c_{L_i}}}{M_{w_i} M_{t_i}} + \frac{F_j M_{c_{L_j}}}{M_{w_j} M_{t_j}}]$$

$$C_{sM} = F_j \rho M_{c_{L_j}} d_{ij} \frac{1}{M_{w_j} M_{t_j}}$$

### 4.3 Variation Aware Abstract Tree Topology

Previous topology generation algorithms such as the popular nearest neighbor algorithm (NNA) [3] ignore process variations when designing the topology and therefore prove less tolerant to variations. We propose an algorithm that takes the global variation

Number of segments	Mean delay (ps)			Delay Variation (ps)			Number of Spice Runs			Error% of ANOVA	
	MC	OPCA	ANOVA	MC	OPCA	ANOVA	MC	OPCA	ANOVA	$\mu$	$\sigma$
12	40.06	39.92	40.11	3.59	3.45	3.63	10000	45	38	0.12	1.11
24	163.67	163.94	164.04	9.24	9.37	9.42	15000	153	86	0.22	1.94
48	496.36	496.66	494.62	26.59	26.36	27.16	20000	561	261	0.35	2.15
196	1616.29	1617.15	1625.82	43.45	42.81	44.86	25000	2145	912	0.59	3.26

**Table 1: Comparison of Monte Carlo(MC), OPCA and ANOV+OPCA results.**

information into account, thereby increasing the tolerance to such variations. When choosing subtrees for merging, we consider the global (spatial) variations at the roots of subtrees in addition to the distance between the roots. The skew introduced between the pair of nodes  $(i, j)$  by the embedding of their parent node  $v$  is given by Equation (18). In Equation (18), only the first term contains the difference of the widths and thickness and can be used for judging pairs of nodes. Since  $l_i$  is unknown before we choose the node pair, the following observation enables the use of  $d_{ij}$  as an estimate of  $l_i$ .

**LEMMA 2.** *The length  $l$  being designed is positive dependent on the distance  $d_{ij}$  between the nodes.*

*Proof:* The observation results directly from an analysis of Equation (18). From the equation, it is easily derived that the dependence of  $l$  on  $d_{ij}$  is of the form,

$$\partial l = f(d_{ij})\partial d_{ij} \quad (26)$$

with  $f(\cdot)$  being positive for all values of  $d_{ij} > 0, l_i < d_{ij}$ , thereby proving the result. Note that the statement holds equally for both lengths  $l_i$  and  $l_j$ . ■

Based on LEMMA 2, we choose the subtree pair  $T_i$  and  $T_j$  with the minimal composite distance defined as

$$d'_{ij} = d_{ij}^2 \left( \lambda + \left| \frac{1}{w_i} - \frac{1}{w_j} \right| \right) \quad (27)$$

Notice that here we demonstrate only the width variation. The same strategy can be used for any other variation sources. This criterion will choose a pair with similar wire width besides close distance. The constant  $\lambda$  is virtually a weighting factor for the distance and its value is chosen empirically. At a step when there are  $|K|$  subtrees,  $|K|/2$  pairs with the minimal  $d'$  are selected to be merged [3].

#### 4.4 Algorithm Complexity

The worst-case time complexity of the abstract topology generation is  $O(n^3)$ , while that of embedding the topology through merging segment construction is  $O(n^2 \log_2 n)$ , where  $n$  is the number of clock sinks. The proof is omitted due to page limit.

## 5. EXPERIMENTAL RESULTS

The proposed algorithms were implemented in ANSI C on a Windows machine with 2.4GHz Pentium 4 processor and 256MB memory. The algorithms were tested on benchmarks (i) r1-r5 [IBM benchmarks], (ii) Primary1 and Primary2 [MCNC benchmarks], and (iii) s1423, s5378 and s15850 [ISCAS89 benchmarks]. All benchmarks were downloaded from the GSRC bookshelf (<http://vlsicad.ucsd.edu/GSRC/bookshelf/Slots/BST/>). The resistance of the wire per unit length is taken as  $0.0042/w\Omega$ , the per unit area capacitance and fringing capacitance are  $3.18e-6pF$ . The skewness threshold is chosen empirically as  $\epsilon = 0.1$ . The relative widths across the chip can be obtained from the variation profile, which is assumed to be known. The factor  $|1/w_i - 1/w_j|$  in Equation (27) lies in the range  $[0, 5]$  for typical values of widths  $w_i$  and  $w_j$ . A small value of  $\lambda$  results in excessive increase in wirelength

(and consequently, the number of violations). On the other hand, a high value of  $\lambda$  would result in swamping  $|1/w_i - 1/w_j|$ . A choice of  $\lambda = 10$  in the merging criterion is seen to produce good results. The permissible ranges for sink pairs are generated randomly such that  $L < 0 < U$  holds. The permissible ranges are not necessarily symmetric with respect to zero i.e.,  $L + U = 0$  is not always true though the values  $|L|$  and  $|U|$  are close to each other in most cases.

In the first part of the experiment, we tested the effectiveness of using ANOVA and OPCA for random variable reduction. The results are shown in Table 1. The comparison is done among Monte Carlo (MC) approach, OPCA reduction only approach (denoted as OPCA in the Table) and ANOVA together with OPCA reduction approach (represented as ANOVA in the Table). We compare the mean delay, delay variation, and reduction result (displayed as number of SPICE runs to fit the model). It is obvious that the more reduction we have, the less number of terms in the delay model we have and the less number of SPICE runs needed to fit the coefficients.

In the second part of the experiment, we compare our NEW statistical clock routing method with the MinSV [8] based on the abstract topology of nearest neighbor algorithm (NNA) [3]. We extend the NNA+MinSV method to incorporate the non-linear global wire variation across the chip. The comparison is made on two different global variation profiles. Profile 1 is the radial variation model described in [6]. Profile 2 is based on a real case which is depicted in Figure 3. On each of the benchmarks, 500-1000 Monte Carlo simulations were carried out. The average results are tabulated in Tables 2. One can see that our NEW method can reduce the number of skew violations by 12% – 37% compared to the NNA+MinSV approach. Figure 7 plots the probability densities of the number of violations for NNA+MinSV and our NEW method. From the histograms, it is seen that the upper bound on the number of violations for the proposed approach is less than the lower bound on the number of violations for NNA+MinSV. Though this may not always be the case, our algorithm outperforms NNA+MinSV consistently in the number the average violations. The CPU time reports in Table 3 verify the observations made about relative algorithm complexity in the previous section.

## 6. CONCLUSION AND FUTURE WORK

We have proposed new probabilistic algorithms for maximizing the tolerance of clock skew to process variations. Both the topology generation phase and the layout embedding phase have been addressed in the proposed algorithms. The number of skew violations for the tested benchmark circuits is lowered significantly in comparison to existing works, while a non-negligible wirelength decrease is also observed for most circuits. In addition, we have considered the fringing capacitance in the design of the clock tree and this is, to our knowledge, the first attempt to do so.

The insertion of buffers in the clock tree would lead to additional variation sources. In the future, we intend to extend the proposed method to incorporate the probabilistic effects of buffer-insertion

	Profile1						Profile2					
	wirelength( $\mu\text{m}$ )			#skew violations			wirelength( $\mu\text{m}$ )			#skew violations		
	NNA+MinSV	NEW	Imprv	NNA+MinSV	NEW	Imprv	NNA+MinSV	NEW	Imprv	NNA+MinSV	NEW	Imprv
r1	137848	127364	8%	74	61	17%	135012	131155	3%	53	42	21%
r2	292006	257338	12%	218	140	36%	286685	262892	8%	163	102	37%
r3	339958	322035	5%	246	194	21%	347469	329361	5%	173	138	20%
r4	695116	657460	5%	358	287	20%	691496	664143	4%	240	196	18%
r5	1034488	983639	5%	501	363	28%	1036414	986358	5%	351	272	22%
Prim1	134045	131013	2%	42	37	12%	135352	132238	2%	32	27	16%
Prim2	352054	308380	12%	155	97	37%	328610	312532	5%	118	94	20%
s1423	110625	104935	5%	26	20	23%	109588	108727	0%	19	16	16%
s5378	221017	169052	23%	75	56	25%	174745	170404	2%	50	44	12%
s15850	446656	431137	3%	226	199	12%	445104	437260	2%	174	145	17%

Table 2: Averaged results for global variation Profile 1 and 2.

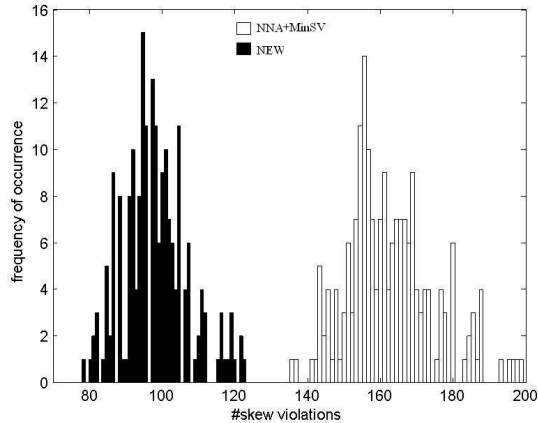


Figure 7: Histogram of the number of skew violations for Benchmark r2 (Monte Carlo Simulations Using Profile2)

and wire-sizing in the clock tree and thereby make the clock tree more robust to process variations. The use of a higher order delay model and consideration of practical issues like multi-layer clock routing handling the various inter-layer parasitics, will also be explored in our future research.

## 7. REFERENCES

- [1] R.-S. Tsay, "Exact Zero Skew", ICCAD, pp. 336 -339, 1991.
- [2] T.-H. Chao, Y.-C. Hsu, J.-M. Ho, K. D. Boese, and A. B. Kahng, "Zero Skew Clock Routing with Minimum Wirelength", IEEE Trans. on Circuits and Systems-II:

- Analog and Digital Signal Processing, Vol. CAS-39, No. 11, pp. 799-814, Nov. 1992.
- [3] M. Edahiro, "A Clustering-Based Optimization Algorithm in Zero-Skew Routings", DAC 1993, pp. 612-616, 1993.
- [4] J. G. Xi, and W. W.-M. Dai, "Useful-Skew Clock Routing with Gate Sizing for Low Power Design", DAC, pp. 383-388, 1996.
- [5] J. Cong, A. B. Kahng, C.-K. Koh, and C.-W. A. Tsao, "Bounded-Skew Clock and Steiner Routing under Elmore Delay", ICCAD, pp. 66-71, 1995.
- [6] Y. Liu, S. Nassif, L. Pileggi, and A. Strojwas, "Impact of Interconnect Variations on the Clock Skew of a Gigahertz Microprocessor", DAC, pp. 168-171, 2000.
- [7] S. Pallela, N. Menezes, and L. T. Pillage, "Reliable Non-Zero Skew Clock Trees Using Wire Width Optimization", DAC, pp. 165-170, 1993.
- [8] B. Lu, J. Hu, G. Ellis, and H. Su, "Process Variation Aware Clock Tree Routing", ISPD, pp. 174-181, 2003.
- [9] G. Venkataraman, C. N. Sze, and J. Hu, "Skew Scheduling and Clock Routing for Improved Tolerance to Process Variations", ASP-DAC, 594-599, 2005.
- [10] W. D. Lam and C.-K. Koh, "Process Variation Robust Clock Tree Routing," ASP-DAC, pp. 606-611, 2005.
- [11] A. I. Abou-Seido, B. Nowak, and C. Chu, "Fitted Elmore Delay: A Simple and Accurate Interconnect Delay Model", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 12, No. 7, July 2004.
- [12] D. C. Montgomery, Design and Analysis of Experiments, John Wiley and Sons, 1997.
- [13] I. S. Kourtev and E. G. Friedman, "Clock skew scheduling for improved reliability via quadratic programming," ICCAD, pp. 239-243, 1999.
- [14] V. Mehrotra, "Modeling the Effects of Systematic Process Variation on Circuit Performance", PhD Dissertation, MIT.
- [15] H. Chang, V. Zolotov, S. Narayan and C. Visweswariah, "Parameterized Block-Based Statistical Timing Analysis with Non-Gaussian Parameters, Nonlinear Delay Functions", DAC, pp. 71-76, 2005.
- [16] L. Zhang, W. Chen, Y. Hu, J. Gubner, and C. C-P. Chen, "Correlation-Preserved Non-Gaussian Statistical Timing Analysis with Quadratic Timing Model", DAC, pp. 83-88, 2005.
- [17] A. Papoulis and S. U. Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, Fourth Edition, 2002.
- [18] F. Liu, C. Kashyap, and C. J. Alpert, "A Delay Metric for RC Circuits Based on the Weibull Distribution", ICCAD, pp. 620-624, 2002.
- [19] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," ICCAD, pp. 621-625, 2003.

	Sinks	CPU(s)	
		NNA+MinSV	NEW
r1	267	1.5	1.5
r2	598	2.3	2.3
r3	862	3.3	3.4
r4	1903	10.1	10.8
r5	3101	24.5	26.1
Prim1	269	1.5	1.5
Prim2	603	2.2	2.3
s1423	74	1.3	1.3
s5378	179	1.4	1.4
s15850	597	2.3	2.3

Table 3: CPU Time Report Comparison