VGTA: Variation–Aware Gate Timing Analysis

Soroush Abbaspour, Hanif Fatemi, Massoud Pedram
Electrical Engineering Deapartment, University of Southern California
{sabbasp,fatemi,pedram}@ceng.usc.edu

Abstract

As technology scales down, timing verification of digital integrated circuits becomes an extremely difficult task due to gate and wire variability. Therefore, statistical timing analysis is inevitable. Most timing tools divide the analysis into two parts: 1) interconnect (wire) timing analysis and 2) gate timing analysis. Variational interconnect delay calculation for block-based σTA has been recently studied. However, variational gate delay calculation has remained unexplored. In this paper, we propose a new framework to handle the variation-aware gate timing analysis in block-based σTA. First, we present an approach to approximate variational RC-π load by using a canonical first-order model. Next, an efficient variation-aware effective capacitance calculation based on statistical input transition, statistical gate timing library, and statistical RC-π load is presented. In this step, we use a single-iteration C_{eff} calculation which is efficient and reasonably accurate. Finally we calculate the statistical gate delay and output slew based on the aforementioned model. Experimental results show an average error of 7% for gate delay and output slew with respect to the HSPICE Monte Carlo simulation while the runtime is about 145 times faster.

1. Introduction

Process technology and environment-induced variability of gates and wires in VLSI circuits makes timing analysis of such circuits a challenging task [1]. More precisely, advanced analysis tools must be developed that are capable of verifying the changes in the circuit timing that stem from various sources of variations. These sources may be broadly classified as follows: imperfect CMOS manufacturing processes (e.g., variations in L, T_{OX}, V_T or ILD thickness), environmental factors such as drops in V_{dd} (resistive drop and ground bounce), substrate temperature changes (due to movement of local hot spots over the chip area), and device fatigue phenomena (e.g., electro-migration, hot electron effects, and negative bias temperature instability) [5].

Static timing analysis (STA) is corner-based. As the number of sources of variation increases, the number of required STA runs rises exponentially. Since it is impossible to analyze all corners, some of the missing corners may result in failures after the chip is manufactured [5]. Notice that the identification of the corner-point is a complicated task which is dependent on the precise interconnect and gate structure [6]. Statistical timing analysis (denoted by σTA) provides an effective solution to this important problem [1][3][4][5].

σTA approaches can be classified into two major groups: 1) path-based σTA, 2) block-based σTA. Because of the shortcomings in path-based σTA, block-based σTA has been received a lot of attention. In block-based σTA, every timing quantity of interest (e.g., delay and slew, arrival time and required arrival time) is represented as a function of global sources of variation (denoted by X_i) and independent random sources of variation (denoted by S_i) in the canonical first-order (denoted by CFO) form[5]. The advantages of such a formulation are that (a) it can capture all correlations and (b) it can produce delay sensitivities due to changes in various environmental and process-related parameters. As a result, designers are able to precisely quantify the sensitivity of a timing parameter to different sources of variation and use this information for timing diagnostics.

Block-based σTA breaks its analysis into two parts: 1) variational interconnect timing analysis and 2) variational gate timing analysis. Variation-aware interconnect timing analysis is studied in [2]. The authors express the resistance and capacitance of a line as a linear function of random variables and then use these r.v.'s to compute circuit moments. These variability-aware moments are used in known closed-form delay metrics [8][9] to compute interconnect delay *PDF*'s. The authors in [3], propose a modeling technique for gate delay variability considering multiple input switching. In [4], a model for calculating statistical gate-delay variation caused by intra-chip and inter-chip variability is presented.

Unfortunately, block-based σTA is lacking in variation-aware gate timing analysis. Recent works do not provide an efficient means of analyzing the gate propagation delay and output slew as a function of variational input transition, variation-aware gate timing library, and variational gate load. In this paper a new framework is proposed for finding variational gate timing behavior. This is achieved by using VGTA (for Variation-Aware Gate Timing Analysis):

1) Given the variational resistive-capacitive load (where all resistances and capacitances are represented in the CFO form), an efficient and accurate algorithm is proposed to calculate variation-aware *RC*-π load. To perform the analysis, we calculate the variation-aware admittance moments (c.f. section 3), and as a result, the resistance and capacitances in the *RC*-π load can be written in CFO form.

2) Given the variational input transition, statistical gate timing library, and variational *RC*-π load, the objective is to find variational gate delay and output slew in the CFO form. In order to achieve the aforementioned goal, a "variation-aware effective capacitance" technique is proposed (c.f. section 4). This method is based on an efficient and reasonably accurate single-iteration C_{eff} approach.

The remainder of this paper is as follows. In section 2, we review the background of block-based σTA. The variation-aware *RC*-π modeling is presented in section 3. Section 4 explains the statistical gate timing analysis for the variational input rise time, variation-aware gate timing library, and variational *RC*-π load. Section 5 presents experimental results. Conclusions are discussed in section 6.

2. Background

2.1 Canonical first-order (CFO) model for timing and electrical parameters in block-based σTA

We employ a first-order variational model for all timing quantities such as the gate and wire delays, arrival times, required arrival times, slacks and slews, i.e., all timing quantities are expressed in CFO form as:

$$\begin{aligned} A &= a_{nom} + \sum_{i=1}^{m} a_i \Delta X_i + a_{m+1} \Delta S_a \\ &= a_{nom} \left(1 + \sum_{i=1}^{m} \frac{a_i}{a_{nom}} \Delta X_i + \frac{a_{m+1}}{a_{nom}} \Delta S_a \right) \end{aligned} \quad (1)$$

where a_{nom} is the the nominal value; ΔX_i's represent the variation of m global sources of variation, X_i, from their nominal values, a_i's are the sensitivities to each of the global sources of variation, ΔS_a is the variation of independent random variable S_a and a_{m+1} is the sensitivity of the timing quantity to S_a. By scaling the sensitivity coefficients, we can assume that ΔX_i and ΔS_a are unit normal distributions $N(0,1)$. Moreover, we define a_i/a_{nom} as the *normalized sensitivity coefficient* (denoted by NSC).

Variation in the physical dimensions of the wire causes change in its resistance and capacitance, thereby, making the gate delay and slew as well as wire delay and slew to vary accordingly. Therefore, we need to capture the effect of geometric variations on the electrical parameters. Classifying the sources of variation into global and independent random sources of variation, we represent electrical parameters of the wire (i.e., R and C) in the CFO form. For instance, R and C in the CFO form are calculated as follows [7]:

$$\begin{aligned} R &= R_{nom} + \sum_{i=1}^{m} r_i \Delta X_i + r_{m+1} \Delta S_r \\ C &= C_{nom} + \sum_{i=1}^{m} c_i \Delta X_i + c_{m+1} \Delta S_c \end{aligned} \quad (2)$$

where R_{nom} and C_{nom} represent nominal resistance and capacitance values, computed when the wire dimensions are at their nominal or typical values. ΔX_i's are the global sources of variations and ΔS_r and ΔS_c represent the independent random sources of variation for the resistance and capacitance, respectively. r_i and c_i are the sensitivity coefficients of resistance and capacitance with respect to the sources of variations, respectively. With appropriate scaling of the sensitivity coefficients, we can assume that ΔX_i, ΔS_r, and ΔS_c are unit normal distributions $N(0,1)$.

2.2 Converting into CFO form

As mentioned in sections 2.1 and 2.2, it is important to represent timing and electrical quantities in the CFO form. This in turn enables one to propagate first order sensitivities to different sources of variation through timing graph[5][7]. In addition, it makes statistical computation efficient and practical and provides timing diagnostics at a very small cost in run time. The remaining question is how to convert a quantity of interest (which itself is a function of different CFO variables) into the CFO form.

The following subsection presents a method to answer the above question. We use an example to show the procedure. The problem we address is how to convert gate propagation delay into the CFO form. However, this method can be easily applied to any quantity of interest.

2.2.1 Gate timing analysis for lumped capacitive load in block-based σTA

Problem Statement I: Given is a variational CMOS driver (due to process and environmental variations) where its input rise time, T_{in}, is in the CFO form and drives an output capacitive load, also, in the CFO form. The objective is to find the gate propagation delay and output slew in the CFO form.

The gate propagation delay is a function of the input transition time, the logic gate characteristics (e.g., the *W/L* ratio, threshold voltage of transistors, V_{dd}, and temperature), and the output load. In commercial ASIC cell libraries, it is possible to characterize various output transition times (e.g. 10%, 50%, and 90%) as a function of above variables, i.e.;

$$t_\alpha = f_\alpha \left(T_{in}, \left\{ \frac{W}{L}, V_T, V_{dd}, Temp, \ldots \right\}, C_L \right) \quad (3)$$

where α denotes the percentage of the output transition, t_α is the output delay with respect to 50% point of the input signal, and f_α is the corresponding delay function. The terms in the bracket capture the gate characteristics and environmental factors, T_{in} is the input transition time, and C_L is the output capacitive load. In block-based σTA, T_{in}, C_L, gate characteristics, and environmental factors are represented in CFO form as a function of global and independent random sources of variations. Hence, to represent t_α in CFO form, we substitute them with their corresponding CFO models. Differentiating with respect to the global and independent random sources of variation, t_α, in CFO form, as a function of m global sources of variation and p independent random sources of variation can be written as:

$$\begin{aligned} & t_\alpha = g_\alpha \left(\Delta X_i \big|_{i=1\ldots m}, \Delta S_j \big|_{j=1\ldots p} \right) \Rightarrow \\ & t_\alpha \cong g_\alpha \Big|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} + \sum_{i=1}^{m} \left. \frac{\partial g_\alpha}{\partial \Delta X_i} \right|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} \cdot \Delta X_i + \sum_{j=1}^{p} \left. \frac{\partial g_\alpha}{\partial \Delta S_i} \right|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} \cdot \Delta S_j \end{aligned} \quad (4)$$

Considering ΔS_js as independent unit normal sources of variations, Eqn. (4) in CFO form can be re-written as:

$$t_\alpha \cong g_\alpha \Big|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} + \sum_{i=1}^{m} \left. \frac{\partial g_\alpha}{\partial \Delta X_i} \right|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} \cdot \Delta X_i + \sqrt{\sum_{j=1}^{p} \left(\left. \frac{\partial g_\alpha}{\partial \Delta S_i} \right|_{\substack{\Delta X_i = 0 \\ \Delta S_j = 0}} \right)^2} \cdot \Delta S_{t_\alpha} \quad (5)$$

As an example, suppose A and B are two given CFO random variables as shown below:

$$A = a_0 + \sum_{i=1}^{m} a_i \Delta X_i + a_{m+1} \Delta S_a \qquad B = b_0 + \sum_{i=1}^{m} b_i \Delta X_i + b_{m+1} \Delta S_b$$

Therefore, for addition, subtraction, multiplication and division of a and b, we have;

a) Addition and subtraction:

$$C = A \pm B = (a_0 \pm b_0) + \sum_{i=1}^{m} (a_i \pm b_i)\Delta X_i + \sqrt{{a_{m+1}}^2 + {b_{m+1}}^2}\,\Delta S_c$$

b) Multiplication:

$$C = A \times B \cong a_0 b_0 + \sum_{i=1}^{m} (a_0 b_i + a_i b_0)\Delta X_i + \sqrt{(a_0 b_{m+1})^2 + (a_{m+1} b_0)^2}\,\Delta S_c$$

c) Division:

$$C = \frac{A}{B} \cong \frac{a_0}{b_0} + \sum_{i=1}^{m} \frac{a_i b_0 - a_0 b_i}{{b_0}^2}\Delta X_i + \sqrt{\left(\frac{a_{m+1}}{b_0}\right)^2 + \left(\frac{a_0 b_{m+1}}{{b_0}^2}\right)^2}\,\Delta S_c$$

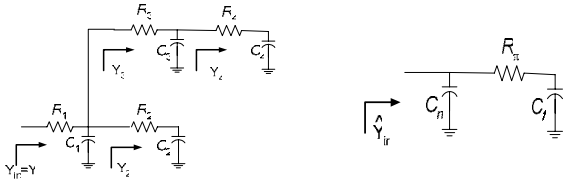
3. *RC*-π Load Calculation in CFO form

Previously the situation in which the load is purely capacitive was discussed. However, in VDSM technologies, one cannot neglect the effect of interconnect resistance of the load on the gate delay and output slew. In STA, a more accurate approximation for an n^{th} order load seen by the gate (i.e., a load with n distributed capacitances to ground) is to use a second order *RC*-π model (c.f. Figure 1(b)). Equating the first, second, and third moments of the admittance of the real load with the first, second, and third moments of the *RC*-π load, we can find C_n, R_π, and C_f as:

$$C_n = Y_{1,in} - \frac{{Y_{2,in}}^2}{Y_{3,in}} \qquad R_\pi = -\frac{{Y_{3,in}}^2}{{Y_{2,in}}^3} \qquad C_f = \frac{{Y_{2,in}}^2}{Y_{3,in}} \qquad (6)$$

where, $Y_{k,in}$ is the k^{th} moment of the admittance of the real load. In σTA, it is required to consider the effect of variability of the load on the gate timing analysis [10], Thus;

Problem Statement II: Given is an *RC* network representation in a design as shown in Figure 1(a), where each *R* and *C* is in the CFO form. The objective is to find an equivalent variational *RC*-π load (i.e., C_n, ,R_π, C_f of Figure 1(b) is in the CFO form), while its admittance matches the admittance of the real load in the frequency range of interest.



(a) (b)

Figure 1: (a) an *RC* network representation of a net in a design. (b) the equivalent *RC*-π model.

C_n, R_π, and C_f are functions of the admittance moments as in Eqn. (6). Hence, examining the variational admittance moments leads us to evaluate the CFO of *RC*-π load parameters. This can be done by differentiating the expressions in Eqn. (6) with respect to the sources of variation (c.f. section 2.3). However, as it is shown, since a recursive operation is utilized to calculate the variational admittance moments, we always represent the admittance moments in CFO form during the recursion, such that the complexity of presenting moments does not increase as recursive operation proceeds. As a result, we propose the following recursive approach to obtain the admittance moments in the CFO form.

Consider the *RCY* segment shown in Figure 2. Assume the admittance at nodes *i* and *j* are represented with infinite series by using the admittance moments as in Eqns. (7) and (8), respectively:

$$Y_i(s) = sY_{1,i} + s^2 Y_{2,i} + \ldots + s^k Y_{k,i} + \ldots \qquad (7)$$

$$Y_j(s) = sY_{1,j} + s^2 Y_{2,j} + \ldots + s^k Y_{k,j} + \ldots \qquad (8)$$

where $Y_{k,i}$ denotes the coefficients of s^k is the k^{th} moment of the admittance of the node *i*. Thus, the admittance at node *i* is computed recursively in terms of the admittance at node *j* as follows [11]:

$$\begin{aligned} &Y_{1,i} = Y_{1,j} + C_i \\ &Y_{k,i} = Y_{k,j} - R_i \sum_{l=1}^{k-1} Y_{l,i} Y_{k-l,j} - R_i C_i Y_{k-1,i} \quad \text{for } k \geq 2 \end{aligned} \qquad (9)$$

Assume the admittance moments of node *j* are written in the CFO form. Thus, by differentiating $Y_{k,i}$ with respect to the sources of variations, $Y_{k,i}$ moments can be also represented in the CFO form (c.f. section 2.3). This can help us not to increase the complexity of presenting the moments as the recursive function proceeds.

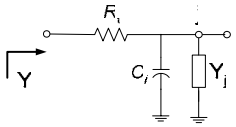


Figure 2: an *RCY* segment model for recursive admittance moment calculation.

As an example, consider the circuit shown in Figure 1. To find the admittance moments of Y_{in}=Y_1 in the CFO form, we need to start from the far end nodes of the *RC* tree (Y_2 and Y_4) and use recursive Eqn. (9). Therefore, we find the first three moments of Y_4 as in Eqn. (12):

$$\begin{aligned} &Y_{1,4} = C_4 \\ &Y_{2,4} = -R_4 C_4 Y_{1,4} = -R_4 {C_4}^2 \\ &Y_{3,4} = -R_4 C_4 Y_{2,4} = {R_4}^2 {C_4}^3 \end{aligned} \qquad (10)$$

Based on the problem statement assumption, C_4 is in CFO form, thereby, $Y_{1,4}$ is in the CFO form. However, since $Y_{2,4}$ and $Y_{3,4}$ are nonlinear functions of CFO variables, we use the technique described in section 2.3 to represent them in CFO form. Similarly, the first three admittance moments of Y_3 as a function of the moments of Y_4 are obtained:

$$\begin{aligned} &Y_{1,3} = Y_{1,4} + C_3 = C_4 + C_3 \\ &Y_{2,3} = Y_{2,4} - R_3 (Y_{1,3} Y_{1,4}) - R_3 C_3 Y_{1,3} \\ &Y_{3,3} = Y_{3,4} - R_3 (Y_{1,3} Y_{2,4} + Y_{2,3} Y_{1,4}) - R_3 C_3 Y_{2,3} \end{aligned} \qquad (11)$$

By using the above statistical recursive operations, we easily compute the moments of Y_{in}=Y_1 in the CFO form.

4. Gate Timing Analysis for the *RC*-π Load in block-based σTA

Problem statement III: Given is a variational CMOS driver, where its input rise time, T_{in}, is in CFO form and which drives a variational *RC*-π load while the resistance and capacitances of this load are also in CFO forms. The objective is to find the gate propagation delay and output slew in CFO form.

Section 2.3.1 solves the same problem where the gate drives a variational pure capacitive load in the CFO form. Therefore, if we substitute the *RC*-π load with its equivalent variational C_{eff}, then the solution to problem statement I is an acceptable solution to problem statement III.

To perform accurate gate delay and output slew calculation, an iterative calculation of C_{eff} is inevitable [12][13][14]. However, as the number of sources of variations increases, the number of required C_{eff} runs rises exponentially (it is proportional to number of corners), and thereby, it can be quite CPU-intensive. We propose an efficient technique to find C_{eff} in CFO form. Suppose the actual C_{eff} in the CFO from can be represented as:

$$\begin{aligned} C_{eff} &= c_{eff,nom} + \sum_{i=1}^{m} c_{eff,i}\Delta X_i + c_{eff,m+1}\Delta S_{c_{eff}} \\ &= c_{eff,nom}\left(1 + \sum_{i=1}^{m} \frac{c_{eff,i}}{c_{eff,nom}}\Delta X_i + \frac{c_{eff,m+1}}{c_{eff,nom}}\Delta S_{c_{eff}}\right) \end{aligned} \quad (12)$$

Since C_{eff} calculation is iterative, we define C_{eff}^{k} (in CFO form) as an approximate presentation for actual C_{eff} (in CFO form), which is resulted from the first k-iterations of the iterative C_{eff} algorithm as:

$$\begin{aligned} C_{eff}^{k} &= c_{eff,nom}^{k} + \sum_{i=1}^{m} c_{eff,i}^{k}\Delta X_i + c_{eff,m+1}^{k}\Delta S_{c_{eff}^{k}} \\ &= c_{eff,nom}^{k}\left(1 + \sum_{i=1}^{m} \frac{c_{eff,i}^{k}}{c_{eff,nom}^{k}}\Delta X_i + \frac{c_{eff,m+1}^{k}}{c_{eff,nom}^{k}}\Delta S_{c_{eff}^{k}}\right) \end{aligned} \quad (13)$$

C_{eff}^{0} means representing C_{eff} using total capacitance algorithm (i.e. C_n+C_f) and C_{eff}^{1} means the value of the effective capacitance by using single iteration and so on. We define $c_{eff,i}^{k}/c_{eff,nom}^{k}$ and $c_{eff,i}/c_{eff,nom}$ as iterative and actual *normalized sensitivity coefficients* (denoted by NSCs), respectively. The NSCs capture the effect of the load variation on the C_{eff} value. It can be shown that in each iteration, the iterative NSCs change slightly (for k≥1), and they converge to their actual NSC values. i.e.,

$$\frac{c_{eff,i}}{c_{eff,nom}} \simeq \frac{c_{eff,i}^{k}}{c_{eff,nom}^{k}} \quad \begin{matrix} 1 \le i \le m, \\ 1 \le k \end{matrix} \quad (14)$$

Using the above observation, problem statement III can be solved by the following steps:

1) Evaluate C_{eff}^{k} in the CFO form (sections 4.1 and 4.2) and therefore find $c_{eff,nom}^{k}$ and $c_{eff,i}^{k}$.for 1≤ i ≤m+1.
2) Find the actual $c_{eff,nom}$ by performing conventional static iterative C_{eff} algorithm for the nominal conditions of the circuit.
3) Using Eqn. (14) and the results of step 1 and 2 , we can find

$$c_{eff,i} = c_{eff,nom} \bullet \frac{c_{eff,i}^{k}}{c_{eff,nom}^{k}} \quad \forall i, 1 \le i \le m+1$$

4) By finding $c_{eff,nom}$ and $c_{eff,1}$, for 1≤ i ≤m+1, we can write C_{eff} in the CFO form. Using the method presented in section 2.3, we obtain the gate delay and output slew in the CFO form.

Step 2 is performed by using well-known STA-based (non-variational) C_{eff} algorithm [12][13][14]. Step 3 is a simple algebraic equation while step 4 is performed as per section 2.3. For step 1, the following sections show how to calculate the C_{eff}^{0} and C_{eff}^{1} in the CFO form.

4.1 Finding Variational C_{eff} using C_{eff}^{0}

As we mentioned before, C_{eff}^{0} approximates C_{eff} with the sum of the total capacitance (i.e., C_n+C_f). Therefore, the C_{eff}^{0} in the CFO form is equal to the sum of C_n in the CFO form and the C_f in the CFO form, i.e. if,

$$\begin{aligned} C_n &= c_{n,nom} + \sum_{i=1}^{m} c_{n,i}\Delta X_i + c_{n,m+1}\Delta S_{c_n} \\ C_f &= c_{f,nom} + \sum_{i=1}^{m} c_{f,i}\Delta X_i + c_{f,m+1}\Delta S_{c_f} \end{aligned} \quad (15)$$

Therefore,

$$\begin{aligned} C_{eff}^{0} &= \left(c_{n,nom} + c_{f,nom}\right) \times \\ &\left(1 + \sum_{i=1}^{m} \frac{\left(c_{n,i} + c_{f,i}\right)}{\left(c_{n,nom} + c_{f,nom}\right)}\Delta X_i + \frac{\sqrt{c_{n,m+1}^{2} + c_{f,m+1}^{2}}}{\left(c_{n,nom} + c_{f,nom}\right)}\Delta S_{c_{eff}}^{0}\right) \end{aligned} \quad (16)$$

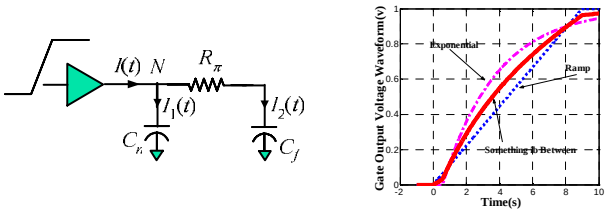
We must calculate C_{eff} for the nominal condition of the circuit (i.e., any quantity in the circuit is at its nominal value) to get $c_{neff,nom}$. Therefore, by using Eqns. (12), (14), and (16) the variational effective capacitance can be written as:

$$\begin{aligned} C_{eff} &= c_{eff,nom} + \sum_{i=1}^{m} \frac{\left(c_{n,i} + c_{f,i}\right)}{\left(c_{n,nom} + c_{f,nom}\right)} \bullet c_{eff,nom}\Delta X_i \\ &+ \frac{\sqrt{c_{n,m+1}^{2} + c_{f,m+1}^{2}}}{\left(c_{n,nom} + c_{f,nom}\right)} \bullet c_{eff,nom}\Delta S_{c_{eff}} \end{aligned} \quad (17)$$

Now, we can use the CFO form of C_{eff} in Eqn. (17) and the method presented in section 2.3 to generate the gate propagation delay and output slew in the CFO form. However, this approach may not capture the effect of the variations of the resistance in the *RC*-π load on the gate timing analysis. Therefore, the next approach, finds NSC's based on a reasonably accurate single-iteration C_{eff} calculation.

4.2 Finding Variational C_{eff} Using C_{eff}^{1}

In this section we find the nominal value of the effective capacitance by performing iterative C_{eff} calculation for the nominal conditions of the circuit. Next we find NSC's by applying a single-iteration effective capacitance method.



(a) (b)

Figure 3: (a) A gate, which drives an *RC*-π calculated load. (b) Gate output waveform is neither ramp nor exponential.

First, we present an efficient single-iteration technique for a reasonably accurate C_{eff} calculation in STA and we use it to further our discussion for calculating the NSCs. Based on its definition, the effective capacitance, C_{eff}, is a pure capacitance that can replace an *RC*-π load such that both *RC*-π and C_{eff} loads store the same amount of charge until a certain point of the output voltage transition (e.g., the 50% point of the output transition.)

To perform C_{eff} calculation, we need to assume a reasonable output waveform for the CMOS driver (c.f. Figure 3(a).) The actual output voltage waveform behaves as a combination of ramp and exponential waveforms as shown in Figure 3(b). We assume that the actual C_{eff} is calculated as a *simple average* of the C_{eff} obtained for the ramp output

waveform and the C_{eff} which is obtained for the exponential output waveform. Thus, it is required to find the C_{eff} for ramp and exponential waveforms of the gate output voltage.

We have shown that the iterative effective capacitance Eqn. for matching any θ% point of the gate output transition time can be written as (proof is omitted for brevity):

$$C_{eff}^{Exp}(\theta) = C_n + k_{Exp}(\theta) C_f \quad where$$
$$k_{Exp}(\theta) = \left[1 + \frac{y}{\theta}\left(e^{\ln(1-\theta)/y} - 1\right)\right] \quad and \quad y = \ln\left(\frac{1-\alpha}{1-\beta}\right)\frac{R_\pi C_f}{T_{R(\alpha-\beta)}} \qquad (18)$$

Furthermore, we have derived that if the output voltage of a gate is approximated with a ramp voltage waveform with α% to β% rise time of $T_{R(\alpha-\beta)}$, then the iterative C_{eff} equation for any θ% output transition point is written as (proof is omitted for brevity):

$$C_{eff}^{Ramp}(\theta) = C_n + k_{Ramp}(\theta) C_f \quad where$$
$$k_{Ramp}(\theta) = \left[1 - \frac{x}{\theta}\left(1 - e^{-\theta/x}\right)\right] \quad and \quad x = (\beta - \alpha)\frac{R_\pi C_f}{T_{R(\alpha-\beta)}} \qquad (19)$$

Thus, based on the simple average assumption, the iterative equation for actual C_{eff} calculation for any θ% point of the output transition time is:

$$\left.\begin{aligned} C_{eff}^{Exp}(\theta) &= C_n + k_{Exp}(\theta) C_f \\ C_{eff}^{Ramp}(\theta) &= C_n + k_{Ramp}(\theta) C_f \end{aligned}\right\} \Rightarrow$$
$$C_{eff}(\theta) = C_n + \left[\xi k_{Exp}(\theta) + (1-\xi) k_{Ramp}(\theta)\right] C_f \qquad (20)$$

where $0 \le \xi \le 1$ is the linear combination factor of the exponential and ramp waveforms. C_{eff}^{1} means using single-iteration of Eqn. (20) as the gate load. Thus, C_{eff}^{1} in the CFO form can be obtained by differentiating the variational Eqn. (20) with respect to the sources of variations (c.f. section 2.3).

Subsequently, using the same approach as in section 4.1, we can find the C_{eff} in the CFO form while the NSCs are calculated using the above single-iteration C_{eff} technique. Experimental results confirm that evaluating variational C_{eff} using the above approach shows an average error of 7% in the final delay and output slew calculation with respect to Monte Carlo simulation.

5. Experimental Results

Our experiments use 90nm CMOS process parameters to model gates and interconnect parasitics. We use standard CMOS gates of various sizes to determine the accuracy of our gate timing analysis. We assumed two different configurations for the experimental setup. The first one consists of two inverters connected in series whereas the second one is a CMOS inverter followed by a 2-input NAND gate. For both configurations, we apply a ramp input to the first inverter while its nominal value is chosen from the set $(t_{in})^{nom}$={10ps,80ps,150ps,220ps,300ps}. For the first configuration, size of the first inverter is fixed at W_p/W_n =30/15μm whereas size of the second inverter is chosen to be one of W_p/W_n={20/10, 50/25, 70/35, 100/50}μm. For the second configuration, size of the first inverter is again fixed at W_p/W_n =30/15μm whereas this time the size of the succeeding 2-input NAND gate is chosen to be one of W_p/W_n={40/40, 50/50, 100/100}μm.

To characterize the timing behavior of the gate, a k-factor equation based library is employed which represents the gate delay and output slew as a function of input rise time and output capacitive load, V_{dd}, and temperature.

We apply different loading scenarios for the second-stage gate as explained in the following subsections, i.e., pure capacitive load, and general *RC* load. We have also considered four different global sources of variation (V_{dd}, temperature, Metal layer 1 width, and ILD) and one independent random sources of variation for each electrical parameter (i.e., *r* and *c*) and timing parameter (for instance t_{in}) in the circuit. The sensitivity of each given data to the sources of variation is chosen randomly, while the total σ variation for each data is chosen to be 10% and 15% of their nominal value. Mean and variance of the effective capacitance, the gate 50% propagation delay, and 10%-90% output transition time (slew) are calculated using the approaches presented in this paper.

To compare the results, we ran HSPICE Monte Carlo simulation tool on each test scenario and derived mean and variance of effective capacitance, the gate 50% propagation delay, and 10%-90% output transition time. The average percentage errors for the mean and variance of effective capacitance, the gate 50% propagation delay, and 10%-90% output transition time between the obtained results from the HSPICE and the calculated results based on using VGTA algorithm are reported.

A. Capacitive Load:

The load in this section is considered to be purely capacitive. Its nominal value is chosen to be $(C)^{nom}$= {400, 500, 800, 1400}*fF*.

We performed our experiments on both circuit configurations explained above. The results for the first configuration (where the second gate is an inverter) are presented in Table 1. The results for the second configuration are provided in Table 2. Experimental results indicate an average error of about 3% for two different σ values, i.e. 10% and 15%. As we increase the σ value (i.e. the total σ variation for each data; e.g. σ variation of t_{in}, and c_l) from 10% to 15%, the error in calculated mean and variance of the delay and slew increase, but slightly. The sources of error can be mainly classified into two groups: 1) the inaccuracy of the gate library table lookup and 2) the linear first order approximation of the timing and electrical parameters with respect to the sources of variation. Note that, the runtime of the proposed algorithm in average is 165 times faster than the Monte Carlo based approach.

Table 1: Average error for the inverter driving pure capacitive load

	σ=10%		σ=15%	
Average error	**Delay**	**Slew**	**Delay**	**Slew**
Error in Mean	1.6%	1.8%	2.1%	2.2%
Error in Variance	1.4%	1.3%	1.9%	1.8%

Table 2: Average error for the 2-input NAND gate driving pure capacitive load

	σ=10%		σ=15%	
Average error	**Delay**	**Slew**	**Delay**	**Slew**
Error in Mean	3.1 %	3.2%	2.8%	3.0%
Error in Variance	3.0%	3.1%	2.9%	2.5%

B. General *RC* Load:

For this section, the load is considered to be an *RC* tree of varying topology. The nominal value of the total resistance of the load is chosen to be from the set $(R)^{nom}$= {150, 260, 300,

710, 1000}Ω and the nominal value of the total capacitance of the load is chosen to be from the set $(C)^{nom}$={400, 500, 800, 1400}*fF*.

Again, we performed the experiment on both circuit configurations as explained before. The results for the first configuration (where the second gate is an inverter) are presented in Table 3 (when the C_{total} is used for calculating the NSC) and Table 4 (when the single iteration C_{eff} is used for calculating the NSC). The results for the second configuration are also provided in Table 5 (when the C_{total} is used for calculating the NSC) and Table 6 (when the C_{total} is used for calculating the NSC). Experimental results indicate an average error of about 19% for different σ values when the C_{total} is used for calculating the NSC. It also shows an average error of about 7% for different σ values when the single iteration C_{eff} is used for calculating the NSC. As we increase the σ value (i.e. the total σ variation for each data; e.g. σ variation of t_{in}, c_n, r_π, and c_f) from 10% to 15%, the error in calculated mean and variance of C_{eff}, the gate delay, and output transition time increase, but slightly. The sources of error can be mainly classified into five groups: 1) the inaccuracy of the gate library table lookup, 2) the linear first order approximation of the timing and electrical parameters with respect to the sources of variation, 3) the error in calculating the variational *RC*-π load and 4) the error in the effective capacitance iterative equation. 5) the error in NSC approximation (Eqn. (14)). Note that, the runtime of the proposed algorithm is, in average, 145 times faster than the Monte Carlo based approach.

Table 3: Average error for the inverter driving general *RC* load when C_{total} is used for calculating *NSC*

	σ=10%		σ=15%	
Average error	**Delay**	**Slew**	**Delay**	**Slew**
Error in Mean	14.6%	15.8%	18.1%	18.3%
Error in Variance	15.4%	16.3%	16.9%	17.9%

Table 4: Average error for the inverter driving general *RC* load when single iteration C_{eff} is used for calculating *NSC*

	σ=10%			σ=15%		
Average error	**Ceff**	**Delay**	**Slew**	**Ceff**	**Delay**	**Slew**
Error in Mean	4.1%	6.5 %	6.7%	4.2%	6.4%	6.4%
Error in Variance	3.9%	5.6%	6.0%	4.3%	6.5%	6.3%

Table 5: Average error for the 2-input NAND gate driving general *RC* load when C_{total} is used for calculating *NSC*

	σ=10%		σ=15%	
Average error	**Delay**	**Slew**	**Delay**	**Slew**
Error in Mean	16.6%	16.8%	19.1%	18.2%
Error in Variance	16.4%	17.3%	17.9%	18.8%

Table 6: Average error for the 2-input NAND gate driving general *RC* load when single iteration C_{eff} is used for calculating *NSC*

	σ=10%			σ=15%		
Average error	**Ceff**	**Delay**	**Slew**	**Ceff**	**Delay**	**Slew**
Error in Mean	3.7%	5.6%	5.8%	4.6%	6.1%	6.2%
Error in Variance	4.1%	5.4%	5.3%	4.5%	5.9%	5.8%

6. Conclusion

In this paper we presented a framework to handle the variation-aware gate timing analysis in block-based σTA. First, we proposed an approach to calculate variational *RC*-π load, which can be utilized instead of the actual variational *RC* load for the gate timing analysis purposes. Following, we presented a reasonably accurate and efficient single-iteration technique for estimating the C_{eff}. We used this technique to calculate the statistical C_{eff} in canonical first-order (CFO) form, and thereby, calculated the gate delay and output slew in CFO form. Experimental results show an average error of 7% for gate delay and output slew with respect to the HSPICE Monte Carlo simulation while the runtime is about 145 times faster.

7. References

[1] R. Nassif, "Modeling and Analysis of Manufacturing Variations," *CICC*, pp. 223-228, 2001.

[2] K. Agarwal, D. Sylvester, D. Blaauw, F. Liu, S. Nassif, S. Vrudhula, "Variational delay metrics for interconnect timing analysis," Design Automation Conference, Proceedings. 41st, June 7-11, 2004 Pages:381 – 384, 2004.

[3] Agarwal, A.; Dartu, F.; Blaauw, D.;"Statistical gate delay model considering multiple input switching", Design Automation Conference, 2004. Proceedings. 41st , June 7-11, 2004 Pages:658 - 663

[4] Okada, K.; Yamaoka, K.; Onodera, H.; "A statistical gate-delay model considering intra-gate variability" Computer Aided Design, 2003. ICCAD-2003. International Conference on , 9-13 Nov. 2003 Pages:908 - 913

[5] Visweswariah, C.; Ravindran, K.; Kalafala, K.; Walker, S.G.; Narayan, S.;"First-order incremental block-based statistical timing analysis", Design Automation Conference, 2004. Proceedings. 41st , June 7-11, 2004, Pages:331 - 336

[6] Y. Liu, S. R. Nassif, L. T. Pileggi, and A. J. Strojwas, "Impact of Interconnect Variations on the Clock Skew of a Gigahertz Microprocessor," *DAC*, pp. 168-171, 2000.

[7] V. Mehrotra, S. Nassif, D. Boning, and J. Chung, "Modeling the Effects of Manufacturing Variation on High-Speed Microprocessor Interconnect Performance," *IEEE Electron Devices Meetings*, pp. 767-770, 1998.

[8] W. C. Elmore, "The transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers," Journal of Applied Physics, pp.55-63, 1948.

[9] C. Alpert, A.Devgan, and C. Kashyap,, "A two moment RC delay metric for performance optimization," Proc. of International Symposium on Physical Design, 2000.

[10] P.R. O'Brien and T. L. Savarino, "Modeling the Driving-Point Characteristics of Resistive Interconnect for Accurate Delay Estimation," Proc. of IEEE int'l Conf. on Computer Aided Design, pp.512-515, 1989

[11] A.B. Kahng, S. Muddu, "Improved effective capacitance computations for use in logic and layout optimization," Proc. of VLSI Design, pp.578 – 582, 1999.

[12] F. Dartu, N. Menezes, and L. Pillegi, "Performance Computation for Precharacterized Gates with RC Loads", IEEE Trans. On Computer Aided Design 15(5):544-533, 1996.

[13] S. Abbaspour, M. Pedram, "Calculating the Effective Capacitance for the RC Interconnect in VDSM Technologies," Proc. of Asia and South Pacific Design Automation Conference, 2003

[14] J. Qian, S. Pu;;ela, and L.Pillage, "Modeling the Effective Capacitance for the RC Interconnect of CMOS gates," IEEE Tran. On Computer Aided Design of VLSI Circuits and Systems, vol. 13 (1994),pp.1526-1535.