



ELSEVIER

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

journal homepage: www.elsevier.com/locate/vlsi

Fast timing analysis of clock networks considering environmental uncertainty[☆]

Hai Wang^a, Hao Yu^b, Sheldon X.-D. Tan^{a,*}

^a Department of Electrical Engineering, University of California at Riverside, Riverside, CA 92521, USA

^b School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639669, Singapore

ARTICLE INFO

Article history:

Received 3 July 2010

Received in revised form

21 December 2010

Accepted 1 March 2011

Available online 4 May 2011

Keywords:

Clock network analysis

Macromodeling

Environmental uncertainties

Incremental SVD

Adaptive sampling

ABSTRACT

Dynamic power management can significantly introduce environmental uncertainties such as non-uniform temperature gradients and supply voltage fluctuations. This can bring many challenges for the system-level timing verification such as for global clock networks. This paper presents a fast verification of clock-skew by an incremental-SVD-based compact modeling assisted with adaptive sampling. Firstly, an incremental-SVD is developed to perform an efficient update of environmental uncertainties avoiding a repeated full SVD. Secondly, an adaptive sampling is presented to build accurate models to sample clock and clock-skew for generating macromodels in a wide frequency range. Experiments on a number of clock networks show that when compared to the traditional fast TBR method, our macromodeling by incremental-SVD and adaptive sampling can significantly reduce the runtime with a similar accuracy. In addition, when compared to the Krylov-subspace-based method, our macromodeling further reduces the waveform error with a similar runtime.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Verifying global clock network, thus timing or clock-skew, under non-ideal environmental conditions and various design corners is becoming critical for robust system-level design at nanometer scale [1–9]. Since more devices are integrated into a single chip, power density has increased continuously, which leads to many hot spots and high temperature-gradients across a chip. Because a clock network is globally routed over the chip, the temperature variation can cause a significant clock-skew variation [7,8,10,11] as the interconnect resistance is linearly dependent on the temperature gradient. In addition, the high-current in power-delivery results in a considerable IR drop and $L(di/dt)$ noise, and the fluctuation in supply voltage can also lead to perturbations to the clock-skew through clock drivers. This situation becomes even worse when there exists dynamic power management, which can further bring non-uniform temperature-gradient and significant supply voltage fluctuation. We call these perturbations of the on-chip power-supply and temperature-

gradient from nominal conditions as *environmental uncertainties* in this paper.

To counter the skew variation from environmental uncertainties, the perturbation based deferred-merge-embedding (DME) method in Cho et al. [7] and Yu et al. [8] modifies merging points for the tree structure and minimizes the skew variation. However, when compared with the clock-tree structure, the non-tree structure is found to be more robust under skew variations [3–6]. Fig. 1 shows two typical non-tree clock networks: a spine-network and a mesh-network. The reason is that there are additional paths provided for a signal to reach a clock sink-node such that the clock-skew can be compensated. The hybrid-network in Su and Sapatnekar [5] applies a top (root) level mesh to drive multiple clock trees in the lower (leaf) level. Recently, Rajaram et al. [6] added cross-links to convert a clock-tree to a non-tree. Though non-tree clock network is robust under variations, an accurate clock-skew variation is still required to preventing failure. Unlike the tree-structured network, there could be more than one clock-driver at inputs for a non-tree-structured network. More importantly, the efficient iterative algorithm for tree-structures may not be used for the non-tree clock network such as clock meshes. The analysis of skew is therefore expensive for the non-tree clock network during a sign-off stage.

The difficulties of clock-skew verification at system-level are summarized as follows. Firstly, the post-layout model, containing massive amounts of extracted parasitics, would significantly increase the analysis complexity. Secondly, environmental

[☆]The work of Hai Wang and Sheldon X.-D. Tan is funded in part by NSF grant under no. CCF-0448534, CCF-1017090 and CCF-1116882 and in part by National Natural Science Foundation of China (NSFC) grant under no. 60828008. The work of Hao Yu was supported by Singapore MOE Tier-2 fund: MOE2010-T2-2-037.

* Corresponding author.

E-mail addresses: haoyu@ntu.edu.sg (H. Yu), stan@ee.ucr.edu (S.X. Tan).

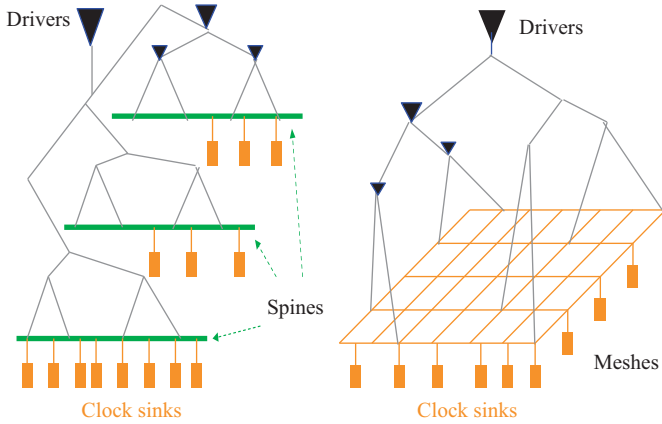


Fig. 1. The design of non-tree clock networks, such as the Intel's spine-network and IBM's mesh-network.

uncertainties perturb the clock network and further complicate the verification of clock-skew. It is thus imperative to find a fast skew analysis for the non-tree clock network considering environmental uncertainties.

Instead of using a full-*RLC* model for a non-tree clock network, macromodeling by model order reduction (MOR) [5,9,12–20] can be employed here to reduce the complexity from not only layout parasitics but also from environmental perturbations. MOR can identify dominant states but reduce unnecessary states. In Su and Sapatnekar [5], a hierarchical macromodeling by the Krylov-subspace-based MOR method is applied for compact representation of parasitics. The work in Ye et al. [9] improves the macromodeling by taking account of the input clock waveform information via a weighted reduction based on input harmonics. However, the Krylov-subspace-based macromodeling only has a localized accuracy (accurate around the expansion point due to moment-matching property). An accurate macromodeling to calculate clock-skew hence needs a better control of the global error.

As an alternative, the fast truncated-balanced-realization (TBR) method, PMTBR [18,17] applies a singular-value-decomposition (SVD) to identify dominant states (singular-vectors) from sampled frequency responses (waveforms) over a frequency range. Although such sampling-based reduction method does not have global error bounds as in standard TBR methods, they typically generate more globally accurate models than Krylov-subspace-based approach. This is required for building macromodels such as *RLC* clock networks that usually have a high operating frequency. The SVD-based macromodeling can be further employed to determine the dominant terminals (or called ports for input signals). The work in Feldmann and Liu [21] shows a SVD-based macromodeling by reducing both terminals and states in the framework of the Krylov-subspace-based approach. The recent work [20] ETBR extends PMTBR for terminal reductions. As such, one can employ ETBR to generate the compact models for networks with many ports like non-tree clock networks, while balancing the accuracy and efficiency requirements. However, how to perform an efficient yet accurate sampling for a wide frequency-band still remains an open problem. Furthermore, in order to consider environmental variabilities, parameterized MOR has been developed by Krylov-subspace-based approach [11,8] and TBR-based approach [17], respectively. Note that the cost of SVD in TBR becomes high when dealing with environmental variabilities, because it requires a full decomposition whenever the network or the input is perturbed.

In this paper, in order to provide a system-level timing verification for clock, we introduce a fast clock-skew analysis via a parameterized and incremental-SVD based fast TBR method, called

piTBR. To consider environmental perturbations in an efficient way, *piTBR* improves the previous TBR methods [17,18,20] by an incremental SVD algorithm that can partially update the macromodel under environmental perturbations. Moreover, to generate a wide-band macromodel of one *RLC* clock, *piTBR* also employs a new adaptive sampling scheme to identify the best sampling points. This can reduce the error of reduced models and improve the efficiency of samplings. The flow of *piTBR* is briefly reviewed as follows. We first parameterize the environmental perturbation into a structured-state matrix. After applying a TBR reduction with the structure-preserving, we can perform an incremental singular-value-decomposition (iSVD), which only analyzes the updated sensitivities every time when an environmental fluctuation exists. This can dramatically reduce the cost of SVD during the TBR reduction. Then we apply a new sampling method to adaptively adjust error and hence to produce a more accurate clock-skew estimation in a wide frequency range. In addition, since our method analyzes both waveforms and input signals, it can consider the variation and its spatial correlation, and is thereby more flexible for skew analysis while considering environmental variations simultaneously. Experiments show that compared to the previous TBR method for environmental variations, our *piTBR* reduces the computational cost by $100\times$ with a similar accuracy. Moreover, with the use of the adaptive re-sampling to control error, our approach is up to $40\times$ faster than the existing re-sampling based reduction method on our benchmarks and is up to $10\times$ more accurate than the simple linear sampling schemes. In addition, because of better global error control, our method further reduces the waveform error by $2\times$ with a similar runtime when compared to the Krylov-subspace-based method on the same benchmark.

The remainder of the paper is organized as follows. We first review background of the Krylov-subspace and truncated-balanced-realization (TBR) based MOR in Section 2. We present a parameterized TBR method in Section 3 with a parameterization procedure for environmental uncertainties, and the incremental SVD method to speedup the parameterized TBR in Section 4. In Section 5, we discuss an adaptive sampling to improve the accuracy of the TBR for calculating the *RLC* clock-skew in wide-band. We present the overall algorithm and the result in Section 6, and conclude the paper in Section 7.

2. Background

The mathematical notations of this paper are summarized in Table 1.

2.1. State-space description of clock network

A high-speed clock network is often modeled by a distributed *RLC* model, which is further described by a linear network in the modified-nodal-analysis (MNA)

$$\mathcal{G}x(t) + C \frac{dx(t)}{dt} = \mathcal{B}u(t), \quad y(t) = \mathcal{L}x(t) \quad (1)$$

or in frequency-domain (s) by

$$(\mathcal{G} + sC)x(s) = \mathcal{B}u(s), \quad y(s) = \mathcal{L}x(s). \quad (2)$$

Here, for \mathcal{N} state-variables $x(t)$ or $x(s)$ (nodal-voltage and inductive-branch-current), $\mathcal{G}(\in \mathcal{N} \times \mathcal{N})$ is the state matrix for conductive elements, $C(\in \mathcal{N} \times \mathcal{N})$ is the state matrix for inductive and capacitive elements, $\mathcal{B}(\in \mathcal{N} \times p_i)$ is the topological matrix to connect p_i inputs sources $u(\in p_i \times 1)$, and $\mathcal{L}(\in \mathcal{N} \times p_o)$ is the topological matrix to connect p_o outputs $y(\in \mathcal{N} \times p_o)$. Accordingly, a

Table 1
Mathematic notations.

$\mathcal{G}, \mathcal{C}, \mathcal{B}, \mathcal{L}$	Original MNA system matrices
$\hat{\mathcal{G}}, \hat{\mathcal{C}}, \hat{\mathcal{B}}, \hat{\mathcal{L}}$	Reduced MNA system matrices
u	Input sources for MNA system
y	Outputs for MNA system
x	Nominal state vector
δx_N	Perturbed state vector by temperature
δx_I	Perturbed state vector by voltage
$\delta \mathcal{G}$	Perturbation to the overall nominal \mathcal{G}
$H(s)$	System transfer function
P	Projection matrix
\mathcal{P}	Structured projection matrix
\mathcal{J}	Input current-vector
$\delta \mathcal{J}$	Perturbation to the overall normandy \mathcal{J}
$r_{unit}(x, y, t)$	Unit-length resistance, x and y are the coordinates here
ρ_0	Unit-length resistance at a reference temperature
β	Temperature coefficient
$T(x, y, t)$	Temperature, x and y are the coordinates here
X	Nominal waveform
dX_N	Waveform surtaxation by temperature
dX_I	Waveform perturbation by voltage
dX	Waveform perturbation by environmental fluctuations composed of dX_N and dX_I
U, V, S	SVD results for $[X; dX]$
U', V', S'	SVD results for X
U'', V'', S''	iSVD results for $[X; dX]$
$S^{(p)}$	Sample points at the previous optimization step
$S^{(a)}$	Sample points inserted
$S^{(c)}$	Sample points at the current optimization step
$S^{(e)}$	Error comparison points
$\hat{x}^{(c)}$	Waveform of the current optimization step reduced system
$\hat{x}^{(p)}$	Waveform of the previous optimization step reduced system

transfer function is defined by

$$H(s) = \mathcal{L}[\mathcal{G} + s_i \mathcal{C}]^{-1} \mathcal{B}. \quad (3)$$

2.2. Model order reduction by truncation balanced realization

The post-layout extraction usually results in a *RLC* network with a huge dimension in the scale of 10^6 variables. To find a compact model with dominant state-variables, model order reduction (MOR) is one solution. In general, MOR finds a small dimensioned matrix P ($\in \mathcal{N} \times q$) with an order $q < \mathcal{N}$, and applies a two-side projection to reduce the dimension of the original model from $\mathcal{N} \times \mathcal{N}$ to $q \times q$. The projection matrix also needs to preserve the accuracy after reduction. To this end, the works in Feldmann and Freund [12], Grimme [13], Kerns and Yang [14], and Odabasioglu et al. [15] show that the orthonormalized subspace can be used as a projection. One way to construct the subspace is based on a block-Krylov subspace, moments of the transfer function $H(s)$ expanded locally at one specified frequency s_0 . However, since it is a local expansion around s_0 , the error of Krylov-subspace based MORs cannot be globally controlled.

Another way to find the orthonormalized subspace is based on the spectrum vectors, i.e., the singular-value vectors or eigenvalue vectors, of the transfer function $H(s)$, sampled at a number of frequency points in a range $[s_{min}, s_{max}]$. As the projection matrix is constructed from globally sampled waveforms, the MOR has a better global error control than the Krylov-subspace-based approach. The fast TBR method [18,17], called the PMTBR method, constructs the projection matrix by n -sampled waveforms at outputs in the frequency-domain

$$Z = [z(s_1), z(s_2), \dots, z(s_n)] \in \mathcal{N} \times p \times n, \quad (4)$$

where each z is

$$z(s_i) = [\mathcal{G} + s_i \mathcal{C}]^{-1} \mathcal{B} \quad (5)$$

obtained from (2). The singular-value-decomposition (SVD) is then applied to Z and its first- K SVD-vectors are $U \in \mathcal{N} \times K$, $S \in K \times K$ and $V \in K \times p \times n$ from

$$Z = {}^{svd}U \cdot S \cdot V. \quad (6)$$

By applying orthonormalization to U , a projection matrix P ($\in \mathcal{N} \times K$) is obtained to reduce the original system by

$$\hat{\mathcal{G}} = P^T \mathcal{G} P, \quad \hat{\mathcal{C}} = P^T \mathcal{C} P, \quad \hat{\mathcal{B}} = P^T \mathcal{B}. \quad (7)$$

However, similar to Krylov subspace methods, the fast TBR methods still have low efficiency issues for circuits with large numbers of ports. Since the dimension of sampled waveform Z still depends on the number of ports p , its cost of SVD and the order of K are large when the number of ports p_i is large.

2.3. Extended reduction considering inputs

In order to reduce the reduction cost when there are a large number of input sources $u(s)$, the MOR can be further extended to consider right-hand-side (RHS) sources. As patterns of input waveforms are usually known during the sign-off stage for designs of the clock or power-delivery network, the EKS/IEKS method [16] uses an input-current-vector [14]

$$\mathcal{J}(s) = \mathcal{B}u(s) = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \\ \vdots \end{bmatrix} \in \mathcal{N} \times 1 \quad (8)$$

instead of the input-port matrix \mathcal{B} during the reduction. The port-dependence is eliminated as the dimension of $\mathcal{J}(s)$ is $\in \mathcal{N} \times 1$. In the extended truncated-balanced realization (ETBR) method [20], the input-current-vector based reduction constructs the projection matrix from the SVD of

$$X = [x(s_1), x(s_2), \dots, x(s_n)] \in \mathcal{N} \times n, \quad (9)$$

where each x is

$$x(s_i) = [\mathcal{G} + s_i \mathcal{C}]^{-1} \mathcal{J}(s_i) \quad (10)$$

obtained from (2) by substituting (8). As X is $\in \mathcal{N} \times n$, its SVD cost can be significantly reduced. Moreover, redundant input information such as the spatial correlation is still included in \mathcal{J} .

Different from the Krylov-subspace-based MORs, which reduce the transfer function, the MOR with right-hand-side (RHS) inputs requires a reduction every time inputs or networks change. Note that our objective is to calculate the clock-skew under perturbations from the environment. Both the network and the input can change slightly from the temperature or supply fluctuation, respectively. Therefore, to avoid repeatedly reducing the network with RHS by ETBR, we apply the ETBR reduction once to the nominal network \mathcal{G} and \mathcal{C} with the nominal input vector \mathcal{J} . Then, with a structured parameterization that considers environmental uncertainties, we further develop an incremental SVD method that only requires an efficient update instead of a full and expensive TBR.

3. Parameterized macromodel of non-tree clock network

Broadly speaking, the skew variation of a clock network is either from environmental uncertainty or process uncertainty. This paper deals with two types of environmental uncertainties. The first one comes from the temperature uncertainty due to the increased power dissipation over the chip. The second one comes from the supply voltage uncertainty due to the decreased voltage room. There are two approaches to deal environmental

uncertainties. One assumes that the fluctuation from the environment can be still deterministically characterized by the worst-case corner [2,6,7,11]. The other [8] assumes a stochastic characterization of the environmental fluctuation. Our method developed in this paper supports both of them. For the simplicity of presentation, we assume that the uncertainty from temperature or supply voltage is already characterized initially.

Notice that we do not consider the delay variations due to buffers in this paper. The delay variations due to buffers and general gates were considered in many existing literature in the context of statistical timing analysis. We assume that buffer variations are known and they can easily be accounted for in our formulation as part of the input current variations shown below.

3.1. Parameterization with environmental uncertainty

Note that the thermal perturbation is in the time-scale of millisecond, quite different from the supply voltage perturbation in the time-scale of nano-second. As such, the perturbed state vector x_N by temperature is naturally not correlated to the perturbed state vector x_I by supply voltage. Hence, we can separately apply the perturbation analysis to them as follows.

When a clock wire experiences a temperature-gradient, the unit-length resistance r_{unit} is [10]

$$r_{unit}(x,y,t) = \rho_0 \cdot [1 + \beta \cdot T(x,y,t)],$$

where ρ_0 is the unit-length resistance at a reference temperature, and β ($1/^\circ\text{C}$) is the temperature coefficient of the metal for clock. Systematically, we denote the perturbation to the overall nominal state matrix \mathcal{G} as $\delta\mathcal{G}$ ($\in \mathcal{N} \times \mathcal{N}$). Such a $\delta\mathcal{G}$ is parametrically dependent on the temperature profile, and hence can be a temperature-dependent distribution all over the non-tree network. In addition, one can evaluate $\delta\mathcal{G}$ once in a sampling-period T_s at the thermal-time-constant scale.

The perturbed MNA by the temperature-gradient becomes

$$(\mathcal{G} + \delta\mathcal{G} + s\mathcal{C}) \cdot (x + \delta x_N) = \mathcal{J}. \quad (11)$$

Similar to [22,11,8], by reorganizing the terms in both sides according to the perturbation order up to the first-order, (11) leads to

$$(\mathcal{G} + s\mathcal{C}) \cdot x = \mathcal{J},$$

$$(\mathcal{G} + s\mathcal{C}) \cdot \delta x_N + \delta\mathcal{G} \cdot x = 0. \quad (12)$$

Note that the clock-skew is the delay difference at different nodes [1–4]. As such, the expansion of the state-variable to the first-order is sufficient to calculate the delay and further the clock-skew.

Moreover, the fluctuation from the supply voltage can lead to small perturbation to clock drivers, i.e., the input waveforms to the clock network. We denote the perturbation to the overall nominal input vector \mathcal{J} as $\delta\mathcal{J}$ ($\in \mathcal{N} \times 1$). The perturbation to MNA by input vector is

$$(\mathcal{G} + s\mathcal{C}) \cdot (x + \delta x_I) = \mathcal{J} + \delta\mathcal{J}. \quad (13)$$

In addition, one can evaluate $\delta\mathcal{J}$ once in a sampling-period T_p at the electric-time-constant scale.

By reorganizing the terms in both sides according to the perturbation order up to the first-order, (13) leads to

$$(\mathcal{G} + s\mathcal{C}) \cdot x = \mathcal{J},$$

$$(\mathcal{G} + s\mathcal{C}) \cdot \delta x_I = \delta\mathcal{J}. \quad (14)$$

Note that the high-order perturbations to x from \mathcal{J} can be added by superposition. Similarly, as clock-skew is the delay difference, the expansion to the first-order perturbation is sufficient for calculating clock-skew.

As such, we can formulate a parameterized state matrix

$$\left(\begin{array}{ccc|ccc} \mathcal{G} & 0 & 0 & \mathcal{C} & 0 & 0 \\ 0 & \mathcal{G} & 0 & 0 & \mathcal{C} & 0 \\ \delta\mathcal{G} & 0 & \mathcal{G} & 0 & 0 & \mathcal{C} \end{array} \right) +_s \begin{pmatrix} x \\ \delta x_I \\ \delta x_N \end{pmatrix} = \begin{pmatrix} \mathcal{J} \\ \delta\mathcal{J} \\ 0 \end{pmatrix} \quad (15)$$

using a new state-variable composed of the nominal $x(s)$ and its sensitivity $\delta x_N(s)$ with respect to the network change due to temperature perturbations, and $\delta x_I(s)$ with respect to the input change due to supply voltage perturbations.

Because of the block-triangular-structure in the augmented state matrix, the nominal state-variable

$$x = (\mathcal{G} + s\mathcal{C})^{-1} \cdot \mathcal{J} \quad (16)$$

and its sensitivities

$$\delta x_I = (\mathcal{G} + s\mathcal{C})^{-1} \cdot \delta\mathcal{J},$$

$$\delta x_N = -(\mathcal{G} + s\mathcal{C})^{-1} \cdot \delta\mathcal{G} \cdot x \quad (17)$$

can be solved efficiently by a block-backward substitution, since there is only one LU-factor cost of $\mathcal{G} + s\mathcal{C}$. The overall response can be efficiently calculated by $x + \delta x_I + \delta x_N$. More importantly, as shown later on, since such a structured formulation separates sensitivities from the nominal response, it leads to an efficient incremental SVD update.

3.2. Input harmonic sampling

As the measurement of the clock-skew distribution is performed to the steady-state response, the output waveform can be measured under inputs of different harmonics [9]. Therefore, instead of applying an error-prone polynomial or rational polynomial fitting of the time-domain waveform as in Wang and Nguyen [16], the Fourier transformation can be applied by expanding the waveform with the fundamental harmonic as the clock frequency, i.e., $\omega_0 = 2\pi f_0$.

Assume the Fourier expansion of the input $\mathcal{J}(t)$ is

$$\mathcal{J}(t) = \sum_{k=0}^N \mathcal{J}(k\omega_0) e^{jk\omega_0 t}, \quad j = \sqrt{-1}, \quad (18)$$

where ω_0 is the fundamental clock frequency, $\mathcal{J}(k\omega_0)$ is the weight of the k th harmonic, and $k = k_1, \dots, k_n$ ($k_1 + k_n = n$) is a linear-scale-sampled harmonic, or $k = 10e^{-k_1}, \dots, 10e^{k_n}$ ($k_1 + k_n = n$) is a log-scale-sampled harmonic.

Define $s_k = j \cdot k \cdot \omega_0$ then the n -sampled frequency-domain waveforms are

$$\begin{bmatrix} x(s_1) & x(s_2) & \dots & x(s_n) \\ \delta x_I(s_1) & \delta x_I(s_2) & \dots & \delta x_I(s_n) \\ \delta x_N(s_1) & \delta x_N(s_2) & \dots & \delta x_N(s_n) \end{bmatrix} \quad (19)$$

with

$$x(s_k) = (\mathcal{G} + s_k\mathcal{C})^{-1} \cdot \mathcal{J}(s_k),$$

$$\delta x_I(s_k) = (\mathcal{G} + s_k\mathcal{C})^{-1} \cdot \delta\mathcal{J}(s_k),$$

$$\delta x_N(s_k) = -(\mathcal{G} + s_k\mathcal{C})^{-1} \cdot \delta\mathcal{G} \cdot x(s_k). \quad (20)$$

Note that similar to the PMTBR method [18], a selective sampling procedure is needed to decide the sampling points that ensures an optimized accuracy. This paper presents an adaptive sampling procedure for this purpose in Section 5.

3.3. Structured projection

As discussed in Section 2, for the TBR method Phillips and Silveira [18], Phillips [17], a flat projection matrix P ($\in \mathbb{R}^{3\mathcal{N} \times K}$) can be constructed from the first- K vectors \mathcal{U} obtained by

a singular-value-decomposition (SVD)

$$\begin{bmatrix} x(s_1) & x(s_2) & \dots & x(s_n) \\ \delta x_I(s_1) & \delta x_I(s_2) & \dots & \delta x_I(s_n) \\ \delta x_N(s_1) & \delta x_N(s_2) & \dots & \delta x_N(s_n) \end{bmatrix} =^{SVD} U \cdot S \cdot V, \quad (21)$$

where $U \in 3\mathcal{N} \times K$, $S \in K \times K$ and $V \in K \times n$.

A few limitations can be observed at this moments. Firstly, the cost to find P is high as the sampling matrix with size $\in 3\mathcal{N} \times n$ needs to be analyzed many times whenever there exists a new update of either δx_I or δx_N . Secondly, the direct projection of P would destroy the block structure and one cannot distinguish the reduced x , δx_I , δx_N after reduction. In the following we show how to resolve the second limitation and further show the solution for the first limitation in the next section.

Similar to Yu et al. [11,8], in order to separate x , δx_I , δx_N after the reduction, we partition the flat projection matrix P into P_0 , P_I and P_N

$$P = \begin{bmatrix} P_0 \\ P_I \\ P_N \end{bmatrix} \quad (22)$$

according to the row-size of x , x_I and x_N , respectively. By further orthonormalizing P_0 , P_I and P_N with each other, we construct a structured projection matrix \mathcal{P} :

$$\mathcal{P} = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & P_I & 0 \\ 0 & 0 & P_N \end{bmatrix}. \quad (23)$$

Such a structured projection matrix will preserve the block-triangular structure during the projection in (7).

Recall that the main cost of TBR comes from the SVD analysis in (21). Whenever there is a change in the network or input, a full SVD needs to be applied, which is computationally expensive. In Section 4, we introduce an incremental TBR method, which only analyzes the waveform of x once, and then partially updates the contribution from a changed δx_I or δx_N .

4. Incremental SVD

In order to efficiently deal with environmental uncertainties in a TBR method, in this section, we introduce an incremental TBR method with two incremental SVDs (iSVDs): an exact iSVD and a fast iSVD.

4.1. Exact iSVD

For an exact iSVD, no approximation is necessary and the singular-value-vector is updated partially without using a full SVD. In our structured formulation, the output waveforms can be

denoted by

$$\begin{bmatrix} X \\ dX \end{bmatrix} = \begin{bmatrix} x(s_1) & x(s_2) & \dots & x(s_n) \\ \delta x_I(s_1) & \delta x_I(s_2) & \dots & \delta x_I(s_n) \\ \delta x_N(s_1) & \delta x_N(s_2) & \dots & \delta x_N(s_n) \end{bmatrix}, \quad (24)$$

where $X (\in \mathcal{N} \times n)$ is the nominal waveform and $dX (\in 2\mathcal{N} \times n)$ is the perturbation by environmental fluctuations composed of δx_I and δx_N . Only dX changes accordingly with environmental fluctuations, and X is constant after the first SVD. Since fully decomposing $[X \ dX]^T$ by SVD every time is expensive, an incremental SVD is needed to only perform the decomposition of the newly updated data dX .

Many fast SVD methods are developed in the literature [23–25]. They avoid the full decomposition by estimating an appropriate low-rank approximation with a partial analysis of the data. However, none of them can be directly applied to the structured data in our problem. For example, the incremental SVD in Band [25] assumes that new sampled data can be updated only columns by columns, while the sampled environmental variations in our case needs to be updated rows by rows. As such, we have developed an exact incremental SVD (iSVD) method in this paper to consider the update rows by rows. One working diagram of iSVD is shown in Fig. 2, which illustrates the primary difference from the existing approach [25].

Assume the SVD of the nominal data X is

$$X =^{SVD} U \cdot S \cdot V$$

with $U \in \mathcal{N} \times K$, $S \in K \times K$ and $V \in K \times n$. Recall that U is composed of the singular-value spectrum and hence can be used as basis to span the original clock system state-space. According to (6), one can build a compact model and extract delay by projecting the original system with U . To further consider the update to the singular-vectors U , V and S from the updated data dX , we derive an incremental SVD scheme below by analyzing the update dX only.

Lemma 1. *The nominal sampled waveform X with update dX can be pre-decomposed by*

$$\begin{bmatrix} X \\ dX \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} S & 0 \\ L & Q \end{bmatrix} \begin{bmatrix} V \\ R \end{bmatrix}, \quad (25)$$

where $L = dX \cdot V^T$, Q and R are the QR-decomposition results of $dX - L \cdot V$, and I is the identity matrix.

Proof. Taking the SVD-vector $V^T (\in n \times K)$ as the subspace basis, the projected component of $dX (\in 2\mathcal{N} \times n)$ along V^T is

$$L = dX \cdot V^T \in 2\mathcal{N} \times K.$$

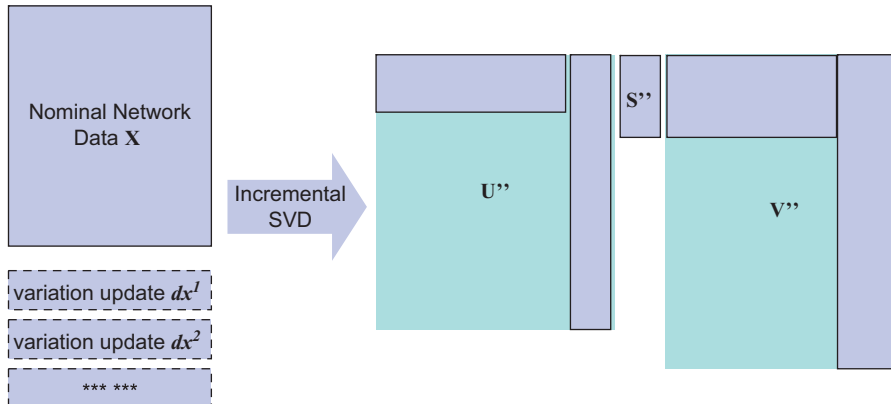


Fig. 2. The working diagram of how iSVD works in an incremental fashion.

Accordingly, one can further define L 's orthogonal component by $H = dX - L \cdot V \in 2\mathcal{N} \times n$.

Applying a QR decomposition to H , one can find two orthogonal matrices $Q (\in 2\mathcal{N} \times n)$ and $R (n \times n)$:

$$H = QR \cdot R.$$

As such, one can have

$$dX = Q \cdot R + L \cdot V.$$

Therefore, the following identity can be derived

$$\begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} S & 0 \\ L & Q \end{bmatrix} \begin{bmatrix} V \\ R \end{bmatrix} = \begin{bmatrix} U \cdot S \cdot V \\ L \cdot V + Q \cdot R \end{bmatrix} = \begin{bmatrix} X \\ dX \end{bmatrix} \quad \square \quad (26)$$

Based on Lemma 1, the incremental SVD can be performed only to the updated columns and rows in dX as follows.

Apply SVD to the middle matrix by

$$\begin{bmatrix} S & 0 \\ L & Q \end{bmatrix} =^{svd} U' \cdot S' \cdot V'.$$

Recall that S is $\in K \times K$, L is $\in 2\mathcal{N} \times K$, and Q is $\in 2\mathcal{N} \times n$. In general, the system dimension \mathcal{N} is much larger than the SVD-approximation order K and the number of samplings n . As such, the dimension of the middle matrix can be approximately $\in 2\mathcal{N} \times n$. Therefore, one only needs to analyze the newly updated data $dX (\in 2\mathcal{N} \times n)$ by an incremental SVD, instead of analyzing the whole data $[X \ dX]^T (\in 3\mathcal{N} \times n)$ by a full SVD.

Then, since

$$U'' \cdot S'' \cdot V'' = \begin{bmatrix} X \\ dX \end{bmatrix}$$

only following updates

$$U'' = \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix} U', \quad (27)$$

$$S'' = S', \quad (28)$$

and

$$V'' = V' \begin{bmatrix} V \\ R \end{bmatrix} \quad (29)$$

are needed to obtain the exact singular-value-vectors U'' , S'' and V'' .

As explained above, this algorithm avoids a fully decomposing the data $[X \ dX]^T (\in 3\mathcal{N} \times n)$. Therefore, one can update U by U'' by only analyzing $dX ((\in 2\mathcal{N} \times n))$ and reusing the result of $X ((\in \mathcal{N} \times n))$. One new projection matrix therefore can be constructed from U'' to reduce the perturbed system equation with one new updated perturbation. We call this method the *exact incremental SVD*.

4.2. Fast iSVD

Note that usually only U'' is the part we are interested in for the projection. This motivates us to find a *fast iSVD* method that not only performs a partial analysis of dX but also only updates U , which can further reduce the computational cost.

As shown by our experiments, given the updated dX from environmental fluctuations, its projection along the singular-vector V^T , i.e., $L (\in 2\mathcal{N} \times K)$ is found to be a good approximation as the update to $U (\mathcal{N} \times K)$

$$U'' = \begin{bmatrix} U \\ L \end{bmatrix}. \quad (30)$$

Note that such a fast iSVD is an approximated SVD method. As shown by our experiments, the fast iSVD is faster than the exact iSVD but with a similar accuracy. The reason is that when the system is perturbed incrementally, it will not largely modify the direction of the dominant state vector such as V^T . In other words, we can still approximately extract the contribution of dX to U by projecting dX along V^T with bounded error. In addition, though the new U'' is not orthonormalized from the above construction, it will be orthonormalized later on when constructing the structured projection.

Compared to the exact iSVD, the fast iSVD avoids the additional cost of one SVD and one QR decomposition, and only has additional cost in the matrix multiplication when building L . This is inexpensive when waveforms are sparsely compressed.

5. Adaptively sampling

To produce an accurate macromodel for clock-skew, we further discuss how to apply an adaptive sampling to improve accuracy. As fast TBR methods are mainly based on the spectrum analysis of the sampled waveform, the selection of sampling points is thus critical.

One observation we have is that if we partition a whole frequency range into a number of sub-bands, for example based on the harmonics of clock frequency, and obtain one reduced model just for one certain sub-band, each sub-band may require a different number of sampling points to reach a common accuracy requirement. To this end, we introduce an adaptive sampling scheme that adaptively adds new sampling points to a number of sub-bands until the errors at all sub-bands are reduced to a pre-defined error bound. In the following, we present a practical sampling scheme with a walk-through example.

As shown in Fig. 3, Step (a) assumes an already sampled system in the previous optimization step p with n sampled points $S^{(p)} = [s_1^{(p)}, s_2^{(p)}, \dots, s_n^{(p)}]$. The reduced model is then described by $\hat{\chi}^{(p)}$ from the previous iteration. In this example, $n=4$. In the initial step ($p=0$), we partition the given frequency band into $n-1$ sub-bands, for example in a log-scale.

Next, in Step (b), $S^{(a)} = [s_1^{(a)}, s_2^{(a)}, \dots, s_{n-1}^{(a)}]$ are inserted into $S^{(p)}$ as the middle points of $s_i^{(p)}$ ($i = 1, \dots, n-1$) in the log-scale, i.e.,

$$s_i^{(a)} = \exp(\frac{1}{2}(\log(s_i^{(p)}) + \log(s_{i+1}^{(p)}))), \quad (31)$$

where $i = 1, 2, \dots, n-1$. Now we have a new set of sampling points $S^{(c)} = [s_1^{(c)}, s_2^{(c)}, \dots, s_{2n-1}^{(c)}]$ for the current optimization step, where $s_{2i-1}^{(c)} = s_i^{(p)}$ and $s_{2i}^{(c)} = s_i^{(a)}$ ($i = 1, 2, \dots, n-1$). The reduced model is then described by $\hat{\chi}^{(c)}$ at current iteration.

Then, in Step (c), we compare the two reduced models $\hat{\chi}^{(p)}$ and $\hat{\chi}^{(c)}$ at a number of error comparison points $S^{(e)} = [s_1^{(e)}, s_2^{(e)}, \dots, s_{2n-2}^{(e)}]$, which are the middle points among $S^{(c)}$:

$$s_i^{(e)} = \exp(\frac{1}{2}(\log(s_i^{(c)}) + \log(s_{i+1}^{(c)}))), \quad (32)$$

where $i = 1, 2, \dots, 2n-2$.

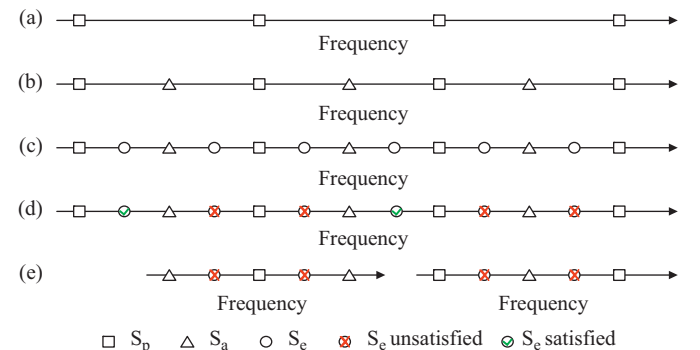


Fig. 3. An illustrative example for the adaptive sampling strategy.

As a result, in Step (d), we can compare the error $\hat{E}(s_i^{(e)})$ of the two reduced models at each point $s_i^{(e)}$

$$\hat{E}(s_i^{(e)}) = \|\hat{x}^{(p)}(s_i^{(e)}) - \hat{x}^{(c)}(s_i^{(e)})\|_2 / \|\hat{x}^{(p)}(s_i^{(e)})\|_2 \quad (33)$$

to represent the error for a sub-band $[s_i^{(c)} s_{i+1}^{(c)}]$ in the frequency-domain. The process continues adding new sampling points (Step (a)) until the errors at each sub-band are reduced to a pre-defined value. In Fig. 3, the unsatisfied points are shown as circles with crosses.

Finally, in Step (e), all the consecutive unsatisfied sub-bands are merged together to form a new unsatisfied frequency range $S^{(p)}$. Note that we may end up with more than one unsatisfied frequency-band such as the two shown in this example. After all the bands meet the accuracy requirement, the sampling responses from all the bands are merged into one matrix to build the final projection matrix.

Such an adaptive sampling scheme is driven by a dynamic error estimation. Experiments show that compared to the existing sampling schemes such as linear, logarithmic, and the recently proposed re-sampling methods [19], our approach can handle large-scale problems and is much faster with a similar accuracy.

6. Algorithm and results

6.1. Overall piTBR algorithm

Algorithm 1. Parameterized and incremental truncated-balanced realization method (piTBR)

Inputs: $\mathcal{G}, \mathcal{C}, \mathcal{J}(t)$, n and K

1. Compute $\delta\mathcal{G}$ and $\delta\mathcal{J}$ according to temperature-gradient and power-supply fluctuations
2. Convert all input signals $\mathcal{J}(t)$ into $\mathcal{J}(s)$ using FFT with n harmonics
3. Perform adaptive sampling of n waveforms: $x(s_k)$ and its sensitivities: $\delta x_I(s_k)$, and $\delta x_N(s_k)$ by (20)
4. Form matrices $X = [x(s_1), \dots, x(s_n)]$, $dX_I = [\delta x_I(s_1), \dots, \delta x_I(s_n)]$, and $dX_N = [\delta x_N(s_1), \dots, \delta x_N(s_n)]$ with a sparse compression
5. If first time, perform full SVD $[X; dX_I; dX_N] = \mathcal{U} \cdot \mathcal{S} \cdot \mathcal{V}$; else perform iSVD update $[X; dX_I; dX_N] = U'' \cdot S'' \cdot V''$
6. Goto 1 whenever there is new environmental fluctuations
7. Partition and orthonormalize U'' into structured \mathcal{P}
8. Project and reduce $\hat{\mathcal{G}} = \mathcal{P}^T \mathcal{G} \mathcal{P}$, $\hat{\mathcal{C}} = \mathcal{P}^T \mathcal{C} \mathcal{P}$, $\hat{\mathcal{J}} = \mathcal{P}^T \mathcal{J}$
9. Calculate skew from x and its sensitivities from δx_I and δx_N

Outputs: Skew and skew variation

In this subsection, we summarize the flow of piTBR method in Algorithm 1. Given the nominal state matrices (\mathcal{G} and \mathcal{C}) of the clock network, specified sampling number n and SVD order K , the algorithm first calculates the network perturbation $\delta\mathcal{G}$ and input perturbation $\delta\mathcal{J}$ from the provided temperature-gradient and power-supply fluctuations. Then, it applies FFT ($O(n \log n)$) to the input signals and calculates the weighted sampling at output waveforms and their sensitivities according to (20). As the augmented state matrix is lower-block-triangular, the cost primarily comes from the LU decomposition of nominal state matrices ($O(\mathcal{N}^{1.2})$). The waveforms and their sensitivities are folded together after a sparse compression. A full SVD with order- K ($O(\mathcal{N}nK^2)$) [24] is applied if it is the first time, and an incremental SVD (iSVD with $O(\mathcal{N}nK)$) [25] is applied afterwards. The procedure repeats when there exist successive updates of environmental fluctuations. After the reduction, the skew and its variation are calculated from the reduced macromodel.

6.2. Experimental results

The proposed piTBR with adaptive sampling is implemented in Matlab. The sparse SVD (SVDPACK) in Matlab is used to analyze the sparsely compressed waveform (by a threshold $1e-5$). All experimental data are simulated and measured on an Intel dual-core 2.0 GHZ PC with 2 GB memory. For the comparison of the piTBR, the fast-TBR-based PMTBR [18] and the Krylov-subspace-based EKS method [16] are implemented together with a variational analysis and a single input vector. In addition, the re-sampling-based method [19] is used for comparison with the new adaptive sampling.

The experimental settings are summarized in Table 2. A chip with size 5 cm^2 is divided into a uniform grid to place the clock network and sample its temperature variations. The temperature maps at nodes of the grid are obtained from a micro-architecture level thermal simulation. The resistance $\rho_0 = 0.03 \Omega/\mu\text{m}$ (at room temperature) and $\beta = 0.007$ (for copper) are used to calculate the perturbed network change, where β is the temperature coefficient of the resistance. Moreover, a clock signal with 1 V Vdd, 2 ns period, 1 ns pulse width and 0.1 ns falling/rising time is used. Up to 10%Vdd random voltage fluctuation is added in each clock period to the clock input. It represents the power-supply fluctuation. In addition, the macromodel is used to generate the transient voltage response and its sensitivities, and then to calculate the clock-skew and skew variation. The waveforms are measured at 50%Vdd for the skew.

6.2.1. Validation of incremental SVD

We first show the accuracy of our parameterized model (15). A clock network with 1377 nodes and 85 ports is used with 70°C temperature change, and 5%Vdd input perturbations. The result of the parameterized model is compared with the HSPICE simulation of the original RLC SPICE netlist. As shown in Fig. 4, the proposed model has negligible error.

Table 2
Experimental settings.

Chip size	5 cm ²
ρ_0	0.03 $\Omega/\mu\text{m}$
β	0.007
Clock Vdd	1 V
Clock pulse width	1 ns
Clock rise time	0.1 ns
Clock fall time	0.1 ns
Clock period	2 ns
Clock variation	Up to 10%

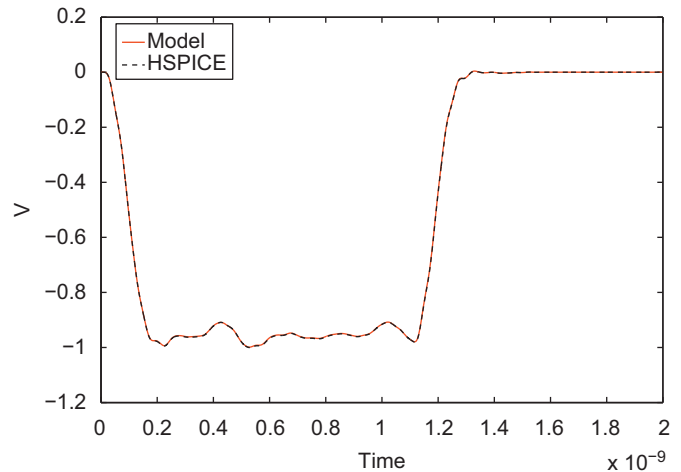


Fig. 4. Simulation results of the original parameterized model, compared with the HSPICE simulation results of the original RLC netlist.

Next, we show the accuracy comparison of waveforms for the full-RLC model and the macromodels by piTBR using exact iSVD and fast iSVD in the time-domain, respectively. The clock network and perturbation settings are the same as the previous experiment. One hundred sampling points are used to sample the waveform in the frequency-domain, and the SVD with an order of 60 is applied to analyze the waveform.

In Fig. 5, (a) compares the overall responses ($x_0 + dx_i + dx_N$), (b) compares the nominal responses (x_0), (c) compares the variations (dx_i) due to the input fluctuation caused by power-supply, and (d) compares the variations (dx_N) due to the network fluctuation caused by temperature-gradient. Compared to the full-RLC model without using macromodel, both the exact iSVD and the fast iSVD macromodels can accurately catch the delay/skew (50%Vdd) and about $61 \times$ (1.16 vs. 61.23 s) times faster. Moreover, with the provided nominal response x_0 and its sensitivities, i.e., the variations dx_i and dx_N , designers can easily modify their designs iteratively to reduce the skew variations.

For the same example above, Fig. 6 further compares the time-domain waveform dX_s due to input variation for the full-RLC

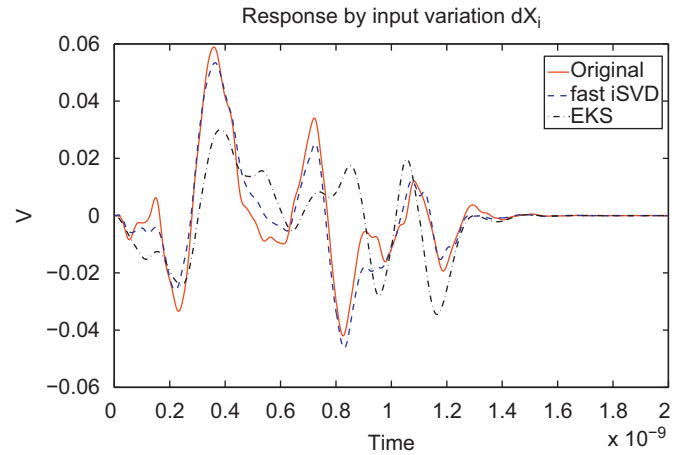


Fig. 6. Accuracy comparison by the original model and the macromodels by the Krylov-subspace-based EKS and the fast iSVD. The waveform is the first-order sensitivity due to the supply voltage perturbation.

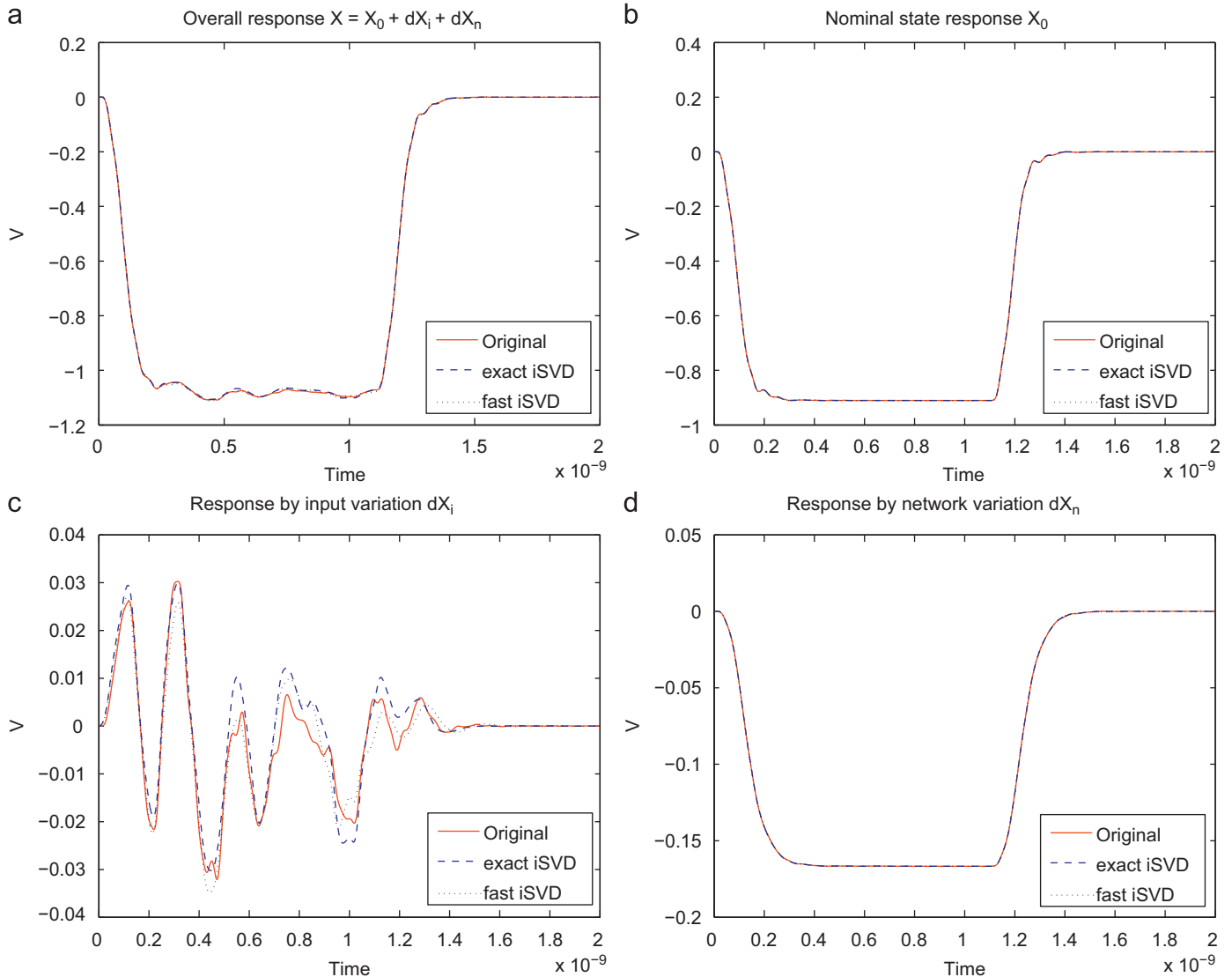


Fig. 5. Accuracy comparison by the original model and the macromodels by the exact iSVD and the fast iSVD. The waveforms are the overall response, nominal response, and first-order sensitivity. The first-order sensitivity dX has a different magnitude-scale from the nominal value X .

model, the macromodels (the same size) by the fast iSVD and the Krylov-subspace-based EKS method at one port. As the macromodel by TBR method is constructed by globally sampling the waveform, it has a better accuracy than the Krylov-subspace-based method.

6.2.2. Validation of adaptive sampling

In our implementation, the number of points in the pool is set to 20, the number of the reduced model used at each iteration is 10, the size of the reduced model is 20. We also used 20 search points, 1/3 of which will be replaced. One point in the pool will be substituted in every iteration. We also implemented speedup techniques such as efficient construction of projectors and heuristic search. We show the accuracy comparison of adaptive sampling, re-sampling, simple uniform linear and log-linear sampling methods on one RLC circuit that has 1724 nodes with eight inputs and outputs. For all the algorithms, we use the same number of sampling points to reach a fair comparison.

Fig. 7 shows the accuracy comparison among the four methods. From the figure we can see that adaptive sampling and re-sampling both achieve much better accuracy than the uniform linear and log-linear methods, where both of them successfully find the spiky frequency band and reach the similar global accuracy by adding more sampling points there. However, because re-sampling requires much more time to perform SVD and compute the variance as the error bound, it takes a much longer time than our adaptive sampling method. For this example with 1724 nodes, adaptive sampling converges with six iterations in 1.1 s and a maximum error 1.8 mV. But the re-sampling method converges with 53 iterations in 41.9 s and a maximum error 1.7 mV. Our adaptive sampling about $40\times$ faster. In addition, our adaptive sampling improves accuracy by $10\times$ compared to the linear-scale sampling and by $2\times$ compared to the log-scale sampling. Our adaptive sampling finds the set of optimal sampling points that provides an accurate TBR reduction to calculate the clock-skew.

6.2.3. Clock skew variation

We then study the skew variation under different types of perturbations: purely by temperature-gradient fluctuation; purely by power-supply fluctuation; and by the combination of temperature and supply. For the same circuit in Section 6.2.1, we generate the various time-domain waveforms using macromodels by piTBR with an order of 60.

Fig. 8(a) shows the clock-skew variation under the temperature-gradient with the following changes: 10, 70 and 140 °C. It shows about 6% clock-skew variation under the 140 °C temperature change. Moreover, Fig. 8(b) shows the skew variation under the supply fluctuation with the following changes: 1, 5 and 10%Vdd. It shows about 11% skew variation under 5%Vdd perturbation. In addition, Fig. 8(c) shows the clock-skew variation under both temperature and supply fluctuations. The combined environmental fluctuation causes up to 14% clock-skew variation.

6.2.4. Scalability study of piTBR

We further study the scalability of runtime and skew-error, by comparing the piTBR with incremental SVD (exact iSVD and fast iSVD), the PMTBR with the single input vector, and the Krylov-subspace-based EKS method. A number of RLC-meshes are used for modeling clock networks with increased sizes and port-numbers. The skew-error here is defined as the relative difference between the skew by the full-RLC model and the one by the macromodel. The runtime here includes the time of waveform sampling, the time of solving reduced macromodel, and the time of constructing the macromodel. Specifically, columns 2–4 in Table 3 summarize the experiments setups, column 5 is the sampling-time and column 6 is the analyzing-time. In addition as shown in Table 4, columns 2 and 3 show the time to construct macromodel, and the skew-error of the Krylov-subspace based EKS method. Columns 4 and 5 show the SVD-time to construct macromodel and the skew-error of piTBR using the fast iSVD, columns 6 and 7 show the same for piTBR using the exact iSVD, columns 8 and 9 show the same for PMTBR using the full-SVD.

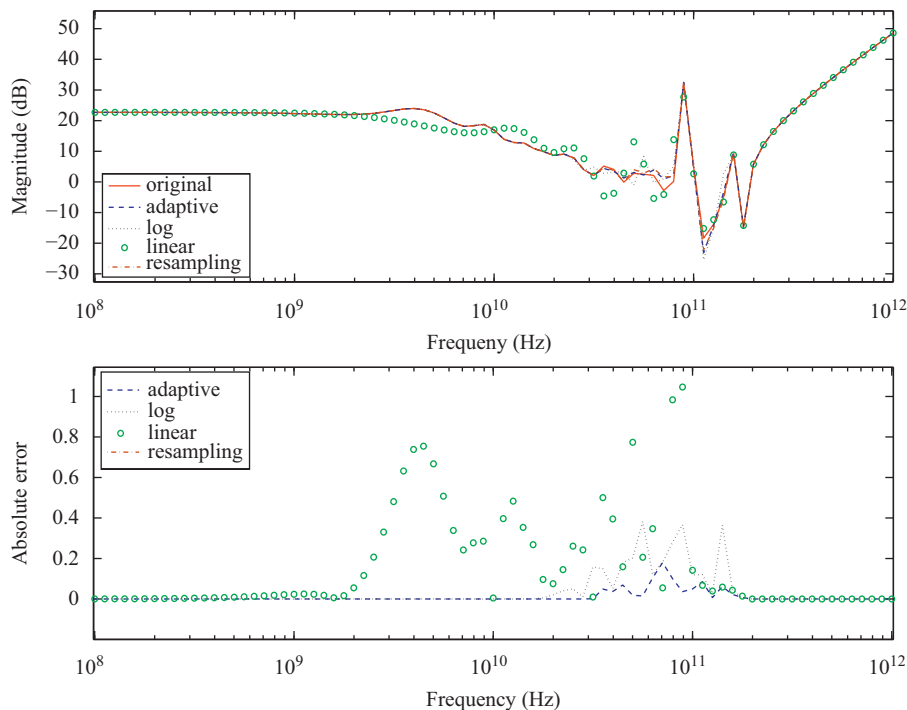


Fig. 7. Comparison of adaptive sampling, re-sampling, linear-scale sampling, and log-scale sampling, for a multiple-input RLC circuit with 1724 nodes using 53 samples.

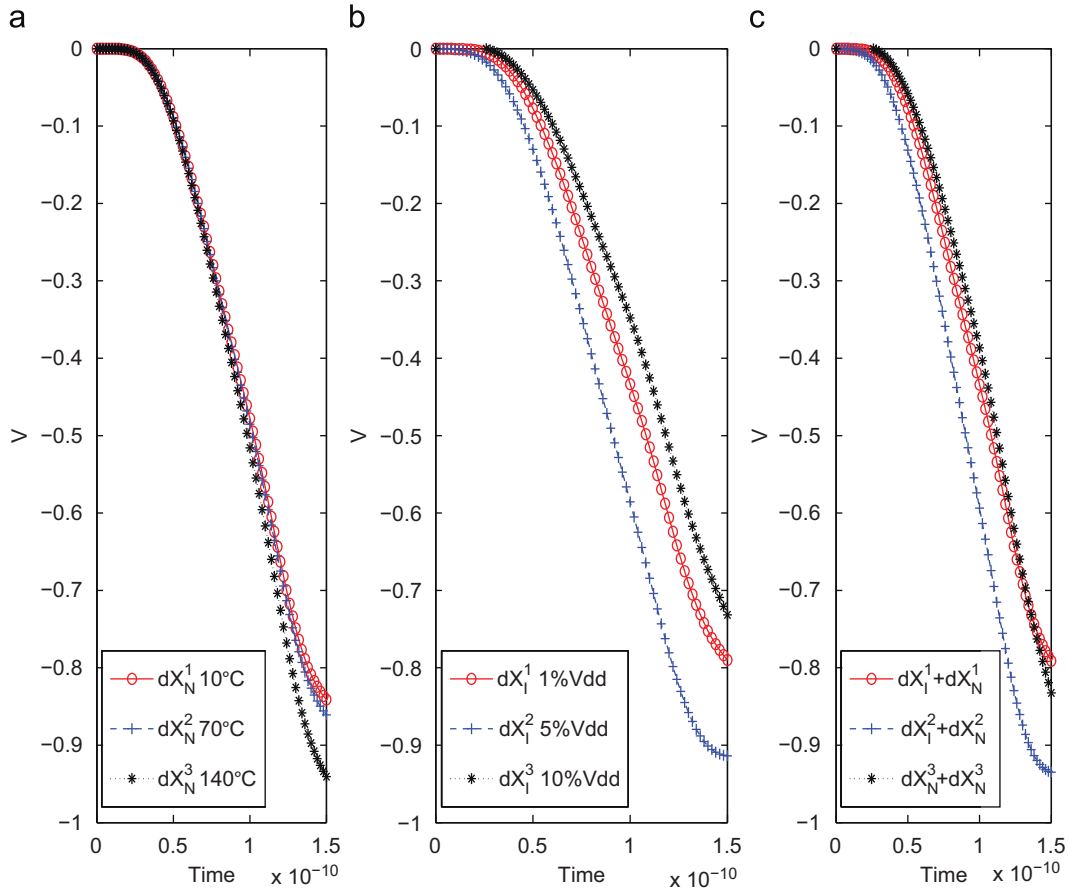


Fig. 8. Skew variation under temperature-caused network change dX_N and supply-caused input change dX_I . (a) The skew variation purely caused by the temperature-gradient; (b) the skew variation purely caused by the power-supply; (c) the skew variation caused by both.

Table 3
Setup of experiments and scalability study of macromodel sampling-time and analyzing-time.

1 Ckt	2 Node	3 Port	4 Order	5 Sample time (s)	6 Solve time (s)
rlc-mesh1	105	5	10	0.07	0.04
rlc-mesh2	369	21	40	0.10	0.41
rlc-mesh3	1377	85	60	0.29	1.56
rlc-mesh4	15939	341	100	1.95	7.49
rlc-mesh5	20865	1365	100	32.4	31.7
rlc-mesh6	69400	10000	100	119	138

As shown by the tables, the larger the size of the waveform, the larger the cost to reapply the full SVD under environmental fluctuations. As iSVD only analyzes the incrementally added rows, it has a better scalability for large sized circuits than the full SVD used in PMTBR. For a RLC circuit with 15 939 nodes and 341 ports, iSVD shows up to $126 \times$ (7.29 vs. 921 s) runtime reduction in SVD with similar waveform error (less than 1.5%). Note that the iSVD-based piTBR can complete simulations for all examples but the PMTBR cannot complete the largest sized example in a reasonable time. The error of all TBR methods are quite close to each other. Because the perturbation does not change much of the direction of the dominant SVD vector, the fast-iSVD-based piTBR can achieve a similar accuracy but is up to $38 \times$ faster than the exact-iSVD-based piTBR.

Moreover, compared to the EKS method, the TBR methods show better accuracy of skew-error due to the global waveform sampling. For the same RLC mesh above, the TBR method reduces the skew-error by $2 \times$. In addition, the computational cost of the TBR is dramatically reduced by the incremental SVD. The exact iSVD has a similar cost as the EKS method and the fast iSVD is even faster than the Krylov-subspace-based method.

7. Conclusions

For non-tree clock networks, we have presented a fast skew analysis that considers uncertainties caused by the temperature-gradient and the power-supply fluctuation. The new method, called piTBR, calculates the clock-skew based on a parameterized and incremental truncated-balanced realization method. Compared to existing macromodelings, our piTBR method is more flexible to be embedded within the skew verification for large-scale clock networks.

In our piTBR, the network changes perturbed by the temperature and the input-source changes perturbed by the supply voltage are parameterized into a structured-state matrix simultaneously. This formulation facilitates an incremental SVD algorithm to significantly reduce the computational cost of clock-skew evaluations. Our piTBR method reduces the runtime by $100 \times$ with a similar error when compared to the full SVD based TBR method. Moreover, as the projection matrix in sampled-based reduction method is obtained by globally sampling in the frequency domain, it leads to a more accurate macromodel than the one by the

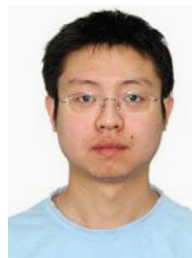
Table 4
Scalability comparison of macromodel building-time and skew-error for piTBR using exact iSVD and fast iSVD, PMTBR using full SVD and Krylov-subspace-based EKS.

1 Ckt	2		3		4		5		6		7		8		9	
	EKS		Fast-iSVD		Exact-iSVD		Full-SVD		Time (s)		Error (%)		Time (s)		Error (%)	
	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)
rlc-mesh1	0.05	0.84	0.01	0.65	0.02	0.60	0.06	0.61								
rlc-mesh2	0.61	1.35	0.04	0.84	0.46	0.81	0.73	0.82								
rlc-mesh3	3.90	1.40	0.15	1.19	3.87	1.16	14.3	1.16								
rlc-mesh4	28.4	5.32	1.15	2.75	7.29	2.04	921	2.02								
rlc-mesh5	142	NA	4.48	NA	73	NA	NA	NA								
rlc-mesh6	545	NA	19	NA	NA	NA	NA	NA								

Krylov-subspace-based method. Experiments in this paper show that our piTBR method reduces the waveform error by $2 \times$ with a similar runtime, when compared to the Krylov-subspace-based method. In addition, with the use of a novel adaptive sampling to reduce error, our piTBR method is up to $40 \times$ faster than the existing re-sampling method, and is up to $10 \times$ more accurate than the simple linear sampling method. Note that the proposed method provides a general framework of the parameterized macromodeling for other high-performance VLSI components and modules in addition to the clock-skew studied in this paper.

References

- [1] J. Cong, A.B. Kahng, C.-K. Koh, C.-W.A. Tsao, Bounded-skew clock and Steiner routing, *ACM Transactions on Design Automation of Electronics Systems* 3 (1997) 341–388.
- [2] Y. Liu, S. Nassif, L. Pileggi, A. Strojwas, Impact of interconnect variations on the clock skew of a gigahertz microprocessor, in: *Proceedings of the Design Automation Conference (DAC)*, 2000.
- [3] N. Kurd, et al., Multi-GHz clocking scheme for Intel(R) Pentium(R) 4 microprocessor, in: *IEEE International Solid-State Circuits Conference (ISSCC)*, 2001.
- [4] P. Restle, et al., A clock distribution network for microprocessors, *IEEE Journal of Solid-State Circuits* (2001) 792–799.
- [5] H. Su, S. Sapatnekar, Hybrid structured clock network construction, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2001.
- [6] A. Rajaram, J. Hu, R. Mahapatra, Reducing clock skew variability via cross links, in: *Proceedings of the Design Automation Conference (DAC)*, 2004.
- [7] M. Cho, S. Ahmed, D.Z. Pan, TACO: temperature aware clock-tree optimization, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2005.
- [8] H. Yu, Y. Hu, C. Liu, L. He, Minimal skew clock embedding considering time variant temperature variation, in: *Proceedings of International Symposium on Physical Design (ISPD)*, 2007.
- [9] X. Ye, P. Li, M. Zhao, R. Panda, J. Hu, Analysis of large clock meshes via harmonic-weighted model order reduction and port sliding, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2007.
- [10] A.H. Ajami, K. Banerjee, M. Pedram, Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 24 (2005) 849–861.
- [11] H. Yu, J. Ho, L. He, Simultaneous power and thermal integrity driven via stapling in 3D ICs, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2006.
- [12] P. Feldmann, R.W. Freund, Efficient linear circuit analysis by Pade approximation via the Lanczos process, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 14 (1995) 639–649.
- [13] E.J. Grimme, Krylov projection methods for model reduction, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1997.
- [14] K.J. Kerns, A.T. Yang, Stable and efficient reduction of large, multiport RC network by pole analysis via congruence transformations, *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems (TCAD)* 16 (1998) 734–744.
- [15] A. Odabasioglu, M. Celik, L. Pileggi, PRIMA: Passive reduced-order interconnect macro-modeling algorithm, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 17 (1998) 645–654.
- [16] J.M. Wang, T.V. Nguyen, Extended Krylov subspace method for reduced order analysis of linear circuit with multiple sources, in: *Proceedings of the Design Automation Conference (DAC)*, 2000.
- [17] J. Phillips, Variational interconnect analysis via PMTBR, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2004.
- [18] J. Phillips, L. Silveira, Poor man's TBR: a simple model reduction scheme, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 24 (2005) 43–55.
- [19] L. Silveira, J. Phillips, Resampling plans for sample point selection in multi-point model-order reduction, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 25 (2006) 2775–2783.
- [20] D. Li, S.X.-D. Tan, B. McGaughey, ETBR: extended truncated balanced realization method for on-chip power grid network analysis, in: *Proceedings of the European Design and Test Conference (DATE)*, 2008.
- [21] P. Feldmann, F. Liu, Sparse and efficient reduced order modeling of linear sub-circuits with large number of terminals, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2004.
- [22] X. Li, P. Li, L. Pileggi, Parameterized interconnect order reduction with explicit-and-implicit multi-parameter moment matching for inter/intra-die variations, in: *Proceedings of International Conference on Computer Aided Design (ICCAD)*, 2005.
- [23] G.H. Golub, C.F.V. Loan, *Matrix Computations*, third ed., The Johns Hopkins University Press, Baltimore, MD, 1989.
- [24] M. Berry, Large Scale Singular Value Computations, *International Journal of Supercomputer Applications* (1992) 13–49.
- [25] M. Band, Incremental singular value decomposition of uncertain data with missing values, in: *European Conference on Computer Vision*, 2002.



Hai Wang received his B.S. degree in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and the M.S. degree in Electrical Engineering from University of California, Riverside, in 2008. He is currently pursuing the Ph.D. degree at the University of California, Riverside. His research interests include circuit modeling and simulation, thermal/power-aware VLSI modeling and optimization.



Hao Yu (M'06) received his B.S. degree from Fudan University (Shanghai, China) in 1999, and both M.S./Ph.D degrees from Electrical Engineering Department at UCLA in 2007, in the field of the integrated circuit and embedded computing. He was a senior research staff at Berkeley Design Automation (BDA) till 2009, one of top-100 start-ups selected by Red-herrings at Silicon Valley. Since 2009, he is an assistant professor in Nanyang Technological University, Singapore. He has 43 refereed international publications, 5 book/chapters, 1 best paper award in ACM Transactions on Design Automation of Electronic Systems (TODAES), 2 best paper award nominations in design automation conference (DAC) and international conference of computer-aided-design (ICCAD), and 1 inventor award from Semiconductor research cooperation (SRC). His primary research interests include 3D cyber-physical computing system, and design exploration at nano-tera scale. He is in the editor board of several journals and serves as the technical program committee member and session chair of several conferences.



Sheldon X.-D. Tan received his B.S. and M.S. degrees in Electrical Engineering from Fudan University, Shanghai, China in 1992 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 1999. He is a Professor in the Department of Electrical Engineering, University of California, Riverside, CA, USA. He is the Associate Director of Compute Engineering Program (CEN) at Bourn College of Engineering at UC Riverside. He also is a cooperative faculty member in the Department of Computer Science and Engineering at UCR.

His research interests include statistical modeling, simulation and optimization of mixed-signal/RF/analog circuits, fast thermal analysis and modeling for microprocessors and platform systems, parallel circuit simulation techniques based on GPU and multicore

systems, and embedded system designs based on FPGA platforms. He also co-authored book “Symbolic Analysis and Reduction of VLSI Circuits” by Springer/Kluwer 2005 and “Advanced Model Order Reduction Techniques for VLSI Designs” by Cambridge University Press 2007. Now Dr. Tan is serving as an Associate Editor for three journals: ACM Transaction on Design Automation of Electronic Systems (TODAE), Integration, The VLSI Journal, and Journal of VLSI Design.

Dr. Tan received Outstanding Oversea Investigator Award from the National Natural Science Foundation of China (NSFC) in 2008. He received NSF CAREER Award in 2004. Dr. Tan received the Best Paper Award from 2007 IEEE International Conference on Computer Design (ICCD’07), two Best Paper Award Nomination from 2005 and 2009 IEEE/ACM Design Automation Conferences, the Best Paper Award from 1999 IEEE/ACM Design Automation Conference. He served as a technical program committee member for DAC, ICCAD, ASPDAC, ICCD, ISQED, BMAS, and ASICON.